

Open Research Online

The Open University's repository of research publications
and other research outputs

Making digital history: The impact of digitality on public participation and scholarly practices in historical research

Thesis

How to cite:

Ridge, Mia (2016). Making digital history: The impact of digitality on public participation and scholarly practices in historical research. PhD thesis The Open University.

For guidance on citations see [FAQs](#).

© 2015 The Author

Version: Redacted Version of Record

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

Making Digital History

The impact of digitality on public participation and scholarly practices in historical research

Mia Ridge, MSc.

Thesis submitted for the degree of Doctor of Philosophy

in the discipline of Digital Humanities

in the Department of History,

The Open University

June 2015

Abstract and keywords

This thesis investigates two key questions: firstly, how do two broad groups - academic, family and local historians, and the public - evaluate, use, and contribute to digital history resources? And consequently, what impact have digital technologies had on public participation and scholarly practices in historical research?

Analysing the impact of design on participant experiences and the reception of digital projects, this thesis makes a contribution to digital historiography by demonstrating the value of methods drawn from human-computer interaction, including heuristic evaluation, trace ethnography and semi-structured interviews. This thesis also investigates the relationship between heritage crowdsourcing projects (which ask the public to help with meaningful, inherently rewarding tasks that contribute to a shared, significant goal or research interest related to cultural heritage collections or knowledge) and the development of historical skills and interests. It situates crowdsourcing and citizen history within the broader field of participatory digital history, and then focuses on the impact of digitality on the research practices of faculty and community historians.

Chapter 1 provides an overview of over 400 digital history projects aimed at engaging the public or collecting, creating or enhancing records about historical materials for scholarly and general audiences. Chapter 2 discusses design factors that may influence the success of crowdsourcing projects. Following this, Chapter 3 explores the ways in which some crowdsourcing projects encourage deeper engagement with history or science, and the role of communities of practice in citizen history. Chapter 4 shifts our focus from public participation to scholarly practices in historical research, presenting the results of interviews conducted with 29 faculty and community historians. Finally, the Conclusion

draws together the threads that link public participation and scholarly practices, teasing out the ways in which the practices of discovering, gathering, creating and sharing historical materials and knowledge have been affected by digital methods, tools and resources.

Keywords

Digital history, historiography, crowdsourcing, citizen history, public participation, research practices, human-computer interaction.

Acknowledgements

Firstly, heartfelt thanks go to my supervisors, Dr Deborah Brunton and Dr Elton Barker, for their advice, guidance and patience as I found my way through the interdisciplinary maze of History and Digital Humanities. I am also grateful to the Open University for the studentship without which this research would not have been possible.

Throughout my doctoral thesis I benefitted from conversations with many people who discussed their projects and the wider field of digital history with me, including Alan Rose, Amanda E. Herbert, Andy Ellis, Asa Letourneau, Ben Brumfield, Chris Lintott, Claudine Chionh, Colin Greenstreet, David Klevan, Duncan Barrett, Elissa Frankle, Geoff Brown, Graham Tongs, Henriette Roued-Cunliffe, Jason Webber, Jim O'Donnell, Jon Voss, Kerri Young, Lesley Stringer, Lori Phillips, Melissa Magna, Melissa Terras, Michael Haley Goldman, Michael Holland, Michele Cohen, Rachel Murphy, Rhiannon Looseley, Ruth E. Richardson, Sarah Hyams, Sharon Howard, Simon Sadler, Stuart Dunn, Susan Pares, Tehmina Goskar, Thomas Bolton, Tim Hitchcock, Victoria Van Hyning and William B. Hafford. I also greatly enjoyed teaching workshops on crowdsourcing at the HILT Summer School 2014, Digital Humanities 2013 conference, and on many occasions at the British Library. I would like to thank the organisers and participants in those and other presentations both for the opportunity to share my work, and for their feedback which helped confirm and shape the discussion of my findings. Similarly, I would like to thank the authors who contributed chapters to my edited volume, *Crowdsourcing our Cultural Heritage*; their reflections have helped inform several arguments presented in this thesis.

During my PhD, I was privileged to take up short Fellowships and Residencies that each helped shape my thinking and created new connections. My thanks to the National

Endowment for the Humanities for two summer institute Fellowships; the first to attend the Polis Center Institute on 'Spatial Narrative and Deep Maps: Explorations in the Spatial Humanities', and the second for 'One Week One Tool' at the Center for History and New Media (where we made Serendip-o-matic). My thanks also the organisers and fellow participants for making the experiences so rewarding. I would also like to thank the CENDARI project for granting me a Visiting Research Fellowship at Trinity College Dublin for my exploration of 'Bridging collections with a participatory Commons: a pilot with World War One archives'. Two week-long museum residencies also helped my project, for which I would like to thank the Powerhouse Museum, Sydney, and the Cooper Hewitt, Smithsonian Design Museum.

I am indebted to, and incredibly appreciative of the interview participants who so generously gave up their time and shared their thoughts to help another researcher. I would also like to thank those to who responded to my survey, answered random questions for information, and helped put me in touch with potential research participants.

My thanks also to the friends and fellow travellers who variously kept me entertained, gave me perspective, provided encouragement, told me about projects and provided music mixes, particularly friends and acquaintances on social media, Min and Josie, and the Blue. I am eternally grateful to my family, Imelda, Ian, Brendan and Niall Ridge, for their support (and for reading a chapter each!), and finally, my thanks to Hannah Williams, who warned me not to do a PhD then encouraged, supported and entertained me throughout.

Table of Contents

Abstract and keywords.....	3
Acknowledgements	5
Introduction: Making digital history	9
Key concepts and definitions	12
Research questions.....	19
Research design and data collection methods.....	20
Thesis structure	27
Chapter 1: An overview of participatory digital history projects	31
Types of participatory projects discussed.....	40
Conclusion	80
Chapter 2: History with the public: crowdsourcing	85
The motivations of crowdsourcing participants.....	93
Design factors in the success of heritage crowdsourcing projects	105
Conclusion	141
Chapter 3: History with the public: from microtasker to historian?.....	149
Definitions and context	151
Examples of citizen history projects	166
How do crowdsourcing projects provide spaces for learning historical skills?.....	173
Can crowdsourcing projects help participants become historians?.....	195
Conclusion	201
Chapter 4: Historians' working practices and digital tools, resources and methods	205
About the interviewees	215
Historians' current practices	217
Discovering	223
Evaluating	228
Gathering	237
Creating.....	242
Sharing	245
Digitality in practice: historians using place, maps and geospatial tools	259
Conclusion	266
Conclusion: The impact of digitality on public participation and scholarly practices in history	272
The impact of digitality on history	275
The gap between the potential and reality of digital history.....	292
Reconsidering 'the historian'	295
Conclusion	301
Bibliography of works cited	305
Appendix A: Websites and projects reviewed	339
Appendix B: Interview participants.....	355
Appendix C: Interview and survey questions.....	357
Interview questions (faculty, family and local historians)	357
Questionnaire for project stakeholders.....	361
Online survey: Crowdsourcing and public participation in digital history	363

Introduction: Making digital history

How do the public, and faculty and community historians make history through creating and using digital history resources?¹ And what impact has digitality - the embedding of networked digital technologies in various aspects of personal and professional life - had on public participation and scholarly practices in historical research?² These are the questions explored in this thesis.

There is a growing body of literature on crowdsourcing in specific fields related to history. Works include Rose Holley's influential publications on crowdsourcing in libraries,³ Johan Oomen and Lora Aroyo's overview of crowdsourcing in cultural heritage,⁴ Stuart Dunn and Mark Hedges' work on humanities crowdsourcing,⁵ the work of Alexandra Eveleigh and

¹ In this thesis, 'faculty historians' refers to historians employed in academic teaching or research roles while 'community historians' refers to family, local and other historians voluntarily undertaking historical research. These definitions are discussed further later.

² 'Digitality' was inspired by Nicholas Negroponte's description of a digital age in which information has shifted from atoms to bits and is decentralised, globalised, harmonised and empowered. The predicted state of 'being digital' has arrived to the extent that computer-based work does not require a central location and can even be performed on devices that fit in our pockets, geographic proximity places few limits on our friendships or collaborations, the delivery of information is rarely tied to our presence at specific times or locations, and computers negotiate our access to networks and services. Nicholas Negroponte, *Being Digital* (London: Hodder & Stoughton, 1995).

³ Rose Holley, 'How Good Can It Get? Analysing and Improving OCR Accuracy in Large Scale Historic Newspaper Digitisation Programs', *D-Lib Magazine* 15, no. 3/4 (2009), doi:10.1045/march2009-holley. Rose Holley, 'Crowdsourcing: How and Why Should Libraries Do It?', *D-Lib Magazine* 16, no. 3/4 (April 2010), doi:10.1045/march2010-holley.

⁴ Johan Oomen and Lora Aroyo, 'Crowdsourcing in the Cultural Heritage Domain: Opportunities and Challenges', in *Proceedings of the 5th International Conference on Communities and Technologies*, 2011, 138-149, <http://www.cs.vu.nl/~marieke/OomenAroyoCT2011.pdf>.

⁵ Stuart Dunn and Mark Hedges, 'Crowd-Sourcing Scoping Study: Engaging the Crowd with Humanities Research' (London, U.K.: King's College, 2012), <http://www.humanitiescrowds.org>.

others on archives,⁶ and the contributions of various authors to my own edited volume.⁷

Libraries and academic projects have studied the research practices of historians and other scholars.⁸ Many research group or project-specific research papers are discussed throughout the thesis, including significant publications on citizen science projects,⁹ and many researchers have explored the motivations of participants in various forms of non-commercial crowdsourcing and peer production.¹⁰ However, there is little work that studies

⁶ Julia Noordegraaf, Angela Bartholomew, and Alexandra Eveleigh, 'Modeling Crowdsourcing for Cultural Heritage', in *Museums and the Web 2014*, ed. Nancy Proctor and Rich Cherry (MW2014: Museums and the Web 2014, Museums and the Web, 2014),

<http://mw2014.museumsandtheweb.com/paper/modeling-crowdsourcing-for-cultural-heritage/>.

⁷ This includes case studies discussing the design and reception of projects such as *What's on the Menu*, *Transcribe Bentham*, *Old Weather*, *Papers of the War Department*, *Waisda?*, and *Your Paintings Tagger*. Mia Ridge, ed., *Crowdsourcing Our Cultural Heritage*, Digital Research in the Arts and Humanities (Farnham, Surrey, UK: Ashgate, 2014),

<http://www.ashgate.com/isbn/9781472410221>.

⁸ Agiatis Benardou et al., 'A Conceptual Model for Scholarly Research Activity', in *ICConference 2010 Proceedings*, ed. John Unsworth, Howard Rosenbaum, and Karen E. Fisher (iConference 2010, Urbana-Champaign, IL: University of Illinois, 2010), 26–32, <http://hdl.handle.net/2142/14945>.

Andrew W University of Minnesota Libraries, 'A Multi-Dimensional Framework for Academic Support: A Final Report', June 2006, <http://purl.umn.edu/5540>. Carole L. Palmer, Lauren C. Tefreau, and Carrie M. Pirmann, 'Scholarly Information Practices in the Online Environment: Themes from the Literature and Implications for Library Service Development' (OCLC Research, 2009),

www.oclc.org/programs/publications/reports/2009-02.pdf. Jennifer Rutner and Roger C Schonfeld,

'Supporting the Changing Research Practices of Historians' (Ithaka S+R, 7 December 2012),

<http://www.sr.ithaka.org/research-publications/supporting-changing-research-practices-historians>.

⁹ For example, M. Jordan Raddick et al., 'Galaxy Zoo: Motivations of Citizen Scientists', *Astronomy Education Review* 12, no. 1 (December 2013), doi:10.3847/AER2011021. Gabriel Mugar et al., 'Planet Hunters and Seafloor Explorers: Legitimate Peripheral Participation Through Practice Proxies in Online Citizen Science', accessed 3 June 2014,

http://crowston.syr.edu/sites/crowston.syr.edu/files/paper_revised%20copy%20to%20post.pdf.

Alexandra Eveleigh et al., 'I Want to Be a Captain! I Want to Be a Captain!: Gamification in the Old Weather Citizen Science Project', in *Proceedings of the First International Conference on Gameful Design, Research, and Applications* (ACM, 2013), 79–82, <http://dl.acm.org/citation.cfm?id=2583019>.

¹⁰ Sultana Alam and John Campbell, 'Crowdsourcing Motivations in a Not-for-Profit GLAM Context:

participation in crowdsourcing alongside other forms of participatory digital history, or links the development of historical skills through crowdsourcing with the literature on historical thinking and education. This thesis links the creation of historical resources through crowdsourcing to other frameworks for creating such resources, including the 'grassroots' projects organised by local history societies, collaboration on family history research, and the personal research collections compiled by individual scholars. It seeks to understand the impact of these participatory projects on users, and the impact of the emergence of digital tools, resources and methods on the practices of historians and more broadly on the discipline of history. The decision to look at faculty and community historians together provides an opportunity to challenge pre-conceptions about their research practices and the attitudes of each group to using digital tools, methods and resources.

This thesis makes a contribution to digital historiography, the 'interdisciplinary study of the interaction of digital technology with historical practice' that seeks to understand the

The Australian Newspapers Digitisation Program', in *ACIS 2012: Location, Location, Location: Proceedings of the 23rd Australasian Conference on Information Systems 2012*, 2012, 1–11, <http://dro.deakin.edu.au/eserv/DU:30049107/alam-crowdsourcingmotivations-2012.pdf>. Max Arends et al., 'Analysing User Motivation in an Art Folksonomy', in *Proceedings of the 12th International Conference on Knowledge Management and Knowledge Technologies (i-KNOW '12)* (New York: ACM, 2012), doi:10.1145/2362456.2362473. Jason Reed et al., 'An Exploratory Factor Analysis of Motivations for Participating in Zooniverse, a Collection of Virtual Citizen Science Projects', *2013 46th Hawaii International Conference on System Sciences*, January 2013, 610–619, doi:10.1109/HICSS.2013.85. Oded Nov, Ofer Arazy, and David Anderson, 'Technology-Mediated Citizen Science Participation: A Motivational Model', in *Proceedings of the AAAI International Conference on Weblogs and Social Media* (Barcelona, Spain, 2011). Dana Rotman et al., 'Dynamic Changes in Motivation in Collaborative Citizen-Science Projects', in *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work* (Seattle, 2012), 217–226, doi:10.1145/2145204.2145238.

'construction, use, and evaluation of digital historical representations',¹¹ by applying research methods from a field dedicated to understanding the interactions between people and computing systems. It synthesises a broad, multi-disciplinary range of literature to understand the impact of design features on participants in crowdsourcing, and to relate participant communities to work on situated learning and historical thinking. Drawing on methods and theoretical lenses from human-computer interaction and user experience research to evaluate participatory history websites, to structure and analyse interviews with historians, and to analyse communications between projects and participants, this thesis enhances our understanding of the reception, as well as the production, of digital history resources.

The thesis is itself an interdisciplinary production, being written by the recipient of a Digital Humanities studentship in a department of History. As such, it is written for two audiences - the more traditional historian, and the digital historiographer or digital humanist - each with their different conventions and expectations. It also draws on personal experience in museum technology, and previous academic training in Computer Science and Human-Computer Interaction. The combination of History and Human-Computer Interaction, in particular, seems particularly suited to digital historiography.

Key concepts and definitions

To understand how digital history is made and received, this thesis moves from a broad outline of participatory digital history projects, to focus on crowdsourcing projects in *Chapter 2* then citizen history projects in *Chapter 3*; together these chapters consider the relationships between 'the public' and projects intending to create digital resources. The

¹¹ Joshua Sternfeld, 'Archival Theory and Digital Historiography: Selection, Search, and Metadata as Archival Processes for Assessing Historical Contextualization', *American Archivist* 74, no. 2 (2011): 544-575, <http://archivists.metapress.com/index/644851P6GMG432Ho.pdf>.

fourth chapter shifts our focus from projects to people, discussing historians' research practices and their use of digital tools, resources and methods. The *Conclusion* builds on the findings from previous chapters, examining the impact of the digital projects, platforms and paradigms discussed throughout the thesis on the ways in which members of the public, faculty and community historians engage in historical research.

As many of the terms and concepts used in the disciplines this thesis draws upon have become nebulous over time, for the sake of clarity I have briefly defined key terms used in this thesis. The term 'historical materials' (or 'historical sources') encompasses any document, map, image, object, sound or video recording that could be used in historical research. This includes material held by private individuals, families and organisations, and official museum, library, archive and academic collections. 'Images' includes physical images such as paintings or photographs, and may include images created from interpreted data (for example, radio signals converted to light waves to represent objects in space), or digital reproductions of physical images. In this thesis, metadata is information (or data) created to 'arrange, describe, track, and otherwise enhance access' to digital or physical objects.¹² It may include information about digitisation processes, the provenance of an object, and classifications within a cataloguing system. 'Digitisation' is defined as making 'a digital copy or digital recording of analogue information' through processes including 'data-entry and transcription, digital imaging, photography [and] sound and video recording'.¹³ This digitisation ranges from photos taken on mobile phones to extremely high-resolution representations made with specialist equipment. Several

¹² Anne J. Gilliland, 'Setting the Stage', in *Introduction to Metadata*, ed. Murtha Baca, 2nd ed., The Getty Research Institute Publications Program (Los Angeles, CA: Getty Publications, 2008), http://www.getty.edu/research/publications/electronic_publications/intrometadata/index.html.

¹³ DigitalNZ, 'Selecting for Digitisation', *Make It Digital*, 3 June 2009, <http://www.digitalnz.org/make-it-digital/selecting-for-digitisation>.

digitised forms of the same physical object may exist, including digital images, transcribed text, descriptions, and metadata. Optical character recognition (OCR) is a method for producing digital transcriptions of typewritten text. Historical material is often collected and digitised by 'GLAMs' - galleries, libraries, archives and museums, also sometimes called 'memory institutions'.

Many of the projects discussed are participatory. Within this thesis, 'participants' are people who actively contribute to a project or website by undertaking tasks such as commenting on, transcribing or adding other content, while 'visitors' on sites or projects read or view material. Visitors may copy or modify material for their own use but they do not undertake activities on the site. I have used 'forums' as shorthand for all forms of internet messageboard or social network in which users of an online site or service can read and post messages, sometimes including private (or 'direct') messages. In contrast, in-person discussions occur between people located in the same physical space at the same time.

The term 'crowdsourcing' was coined in 2006 by Jeff Howe and Mark Robinson to describe the act of taking work once performed within an organisation and outsourcing it to the general public through an open call for participants.¹⁴ For the purposes of my research, I have used the following definition of crowdsourcing: crowdsourcing in cultural heritage asks the public to help with meaningful tasks that contribute to a shared, significant goal or research interest related to cultural heritage collections or knowledge. As a voluntary

¹⁴ Jeff Howe, 'Crowdsourcing: A Definition', 2 June 2006, http://crowdsourcing.typepad.com/cs/2006/06/crowdsourcing_a.html. Howe later claimed that coining a portmanteau term was a joke at Silicon Valley's expense. William Safire, 'On Language - Fat Tail', *The Times Magazine*, 5 February 2009, <http://www.nytimes.com/2009/02/08/magazine/08wwln-safire-t.html>.

activity, the tasks and/or goals are designed to be inherently rewarding.¹⁵ Crowdsourcing projects often use, 'microtasks', small, self-contained actions that can be completed quickly. This thesis also discusses 'citizen science', a model in which dispersed volunteers contribute to professional research projects 'using methodologies that have been developed by or in collaboration with professional researchers'.¹⁶ After my analysis of the field, I have defined 'citizen history' projects as those requiring or teaching some historical skills beyond the technical skills of transcribing text or classifying images. That is, in distinction to crowdsourcing projects, they support an engagement with scholarly disciplines beyond the technical skills of palaeography or image classification. However, in common usage, 'citizen science' or 'citizen history' may also be used by crowdsourcing projects that aspire to go beyond microtasks, or as a marketing term to encourage participation or attract funding.

Turning to terms from human-computer interaction (HCI), 'user experience' encompasses the 'emotions, beliefs, preferences, perceptions, physical and psychological responses, behaviours and accomplishments' related to using a particular system.¹⁷ The 'user interface' includes any components of a system that provide 'information and controls' for the user.¹⁸ 'Mental models' are the knowledge that people develop about how to interact with a

¹⁵ An earlier version of this definition was published in Mia Ridge, 'Crowdsourcing Our Cultural Heritage: Introduction', in *Crowdsourcing Our Cultural Heritage*, ed. Mia Ridge (Farnham, Surrey, UK: Ashgate, 2014), <http://www.ashgate.com/isbn/9781472410221>.

¹⁶ Caren B. Cooper et al., 'Ecology and Society: Citizen Science as a Tool for Conservation in Residential Ecosystems', *Ecology and Society* 12, no. 2 (2007), <http://www.ecologyandsociety.org/vol12/iss2/art11/>.

¹⁷ International Organization for Standardization, 'ISO 9241-210:2010(En). Ergonomics of Human-System Interaction — Part 210: Human-Centred Design for Interactive Systems' (International Organization for Standardization, 2010), <https://www.iso.org/obp/ui/#iso:std:iso:9241:-210:ed-1:vi:en>.

¹⁸ International Organization for Standardization, 'Human-Centred Design for Interactive Systems'.

system; these models sometimes include an idea of how the system works.¹⁹ A key concept from HCI is perceived 'affordances', aspects of a device or interface that provide clues about how to use it, and therefore what it can do.²⁰ The screen-based affordances seen on websites and mobile applications may be based on cultural conventions and previous experiences.²¹ Bespoke software or systems have been custom-made. Some 'off-the-shelf' software systems are extensively configurable and can be tailored to meet specific requirements, but they will still retain some core functions or interaction models with other products made with the same software.

In order to assess whether crowdsourcing projects can provide spaces in which people can become historians, I sought to define what being or becoming 'a historian' actually means. The term is nebulously defined, and surprisingly slippery.²² In her examination of the discipline, *History in Practice*, historian Ludmilla Jordanova describes the discipline of history as 'a set of practices' rather than beliefs or theories: 'there is no essence of the discipline'.²³ The American Historical Association (AHA's) Standards of Professional Conduct emphasise the way historians honour the integrity of the historical record, use and document sources, and link evidence with arguments to build 'fair-minded, nuanced,

¹⁹ Helen Sharp, Yvonne Rogers, and Jenny Preece, *Interaction Design: Beyond Human-Computer Interaction*, vol. Second (Chichester: Wiley, 2007). pp. 116-118.

²⁰ Donald A. Norman, 'Affordance, Conventions, and Design', *Interactions* 6, no. 3 (1999): 38-43, <http://dl.acm.org/citation.cfm?id=301168>. Yvonne Rogers, 'New Theoretical Approaches for Interaction', *Human-Computer Interaction*, 2002, 87-143.

²¹ Rogers, 'New Theoretical Approaches for Interaction'.

²² This may be because historians see the discipline as 'largely atheoretical' and 'a craft to be mastered as much as a set of techniques to be learned', according to Becher's study of the 'linguistic features' of disciplines as represented in journals and scholarly discourse. Tony Becher, 'Disciplinary Discourse', *Studies in Higher Education* 12, no. 3 (1987): 261-74, doi:10.1080/03075078712331378052.

²³ Ludmilla Jordanova, *History in Practice*, Second edition (London, U.K.: Bloomsbury Academic, 2006). p. 14.

and responsible interpretations of the past'.²⁴

'Historical thinking' is an important concept in the teaching of history.²⁵ Sam Wineburg's research into 'historical cognition' found that 'mature historical thought' depends on the ability to 'navigate the uneven landscape of history', to move between 'the poles of familiarity and distance from the past'²⁶ and to find a balance between 'naïve historicism' and 'a rigid sense of disconnection'.²⁷ Historian John Tosh found that historical thinking in academic scholarship makes 'repeated use of the antithesis between continuity and change' and is 'particularly attentive to claims of novelty'.²⁸ Historians Thomas Andrews and Flannery Burke developed the 'five C's of historical thinking' to help teachers in the classroom. In their view, the concepts of 'change over time, causality, context, complexity, and contingency' describe the 'shared foundations' of the discipline of history.²⁹

Educational researchers Bill Tally and Lauren Goldenberg defined historical thinking as observation (noticing details while 'scanning and parsing' documents), sourcing (asking

²⁴ American Historical Association, 'Statement on Standards of Professional Conduct', *American Historical Association*, January 2011, <http://www.historians.org/about-aha-and-membership/governance/policies-and-documents/statement-on-standards-of-professional-conduct>.

²⁵ See for example T. Mills Kelly, 'Thinking: How Students Learn About the Past', in *Teaching History in the Digital Age* (University of Michigan Press, 2013), <http://hdl.handle.net/2027/spo.12146032.0001.001>.

²⁶ Sam Wineburg, *Historical Thinking And Other Unnatural Acts: Charting The Future Of Teaching The Past* (Philadelphia: Temple University Press, 2001). p. 5.

²⁷ Wineburg, *Historical Thinking And Other Unnatural Acts: Charting The Future Of Teaching The Past*. p. 12.

²⁸ John Tosh, 'Public History, Civic Engagement and the Historical Profession in Britain', *History* 99, no. 335 (2014): 191–212, <http://onlinelibrary.wiley.com/doi/10.1111/1468-229X.12053/full>.

²⁹ Thomas Andrews and Flannery Burke, 'What Does It Mean to Think Historically?', *Perspectives on History* 45, no. 1 (January 2007), <https://www.historians.org/publications-and-directories/perspectives-on-history/january-2007/what-does-it-mean-to-think-historically>.

who made the document and why), drawing inferences, citing evidence for arguments, posing questions ('cultivating puzzlement' and keeping track of questions), and corroboration (comparing what is found to other documents and prior knowledge).³⁰ I have used these definitions as cues when looking for signs of historical thinking in participant communities.

In the scope of this thesis, 'faculty' historians are historians employed in teaching or research roles in academia. 'Academic' (or 'academically-trained') historians have formal training as historians and/or produce scholarly work within a history-related academic body. Academic historians might work as public historians, in GLAMs, or in 'alt-ac'³¹ (alternative academic, or 'non-professoriate') jobs within universities. They might be looking for work as a faculty historian or have a career unrelated to their academic training as historians.³² 'Professional historian' includes any historian paid for their historical research, teaching or writing.

The term 'amateur' historian, when applied to the non-professional historians I interviewed and many like them, is problematic.³³ For example, according to the AHA,³⁴

³⁰ Bill Tally and Lauren B. Goldenberg, 'Fostering Historical Thinking With Digitized Primary Sources', *Journal of Research on Technology in Education* 38, no. 1 (2005): 1–21.

³¹ Katina Rogers, '#alt-Ac in Context', *Media Commons*, accessed 3 June 2015, <http://mediacommons.futureofthebook.org/alt-ac/alt-ac-context>.

³² I would like to thank the correspondent who alerted me to the usage of the term 'academic historians' to mean 'historians with academic qualifications' when I posted my call for interview participants to an international discussion list.

³³ Melissa Terras points out that 'amateur' can be used pejoratively, and does not take into account 'the range of expertise and the amount of knowledge' possessed by resource creators. Melissa Terras, 'Digital Curiosities: Resource Creation via Amateur Digitization', *Literary and Linguistic Computing* 25, no. 4 (2010): 425–438, doi:10.1093/llc/fqq019.

³⁴ I have used an American organisation because they have articulated their definition of and

professional historians are defined by their identification with a community of historians who share core values including 'conducting and assessing research, developing and evaluating interpretations, communicating new knowledge, navigating ethical dilemmas, and, not least, telling stories about the past' and who are engaged in the 'disciplined learned practice' of 'investigating and interpreting the past'.³⁵ Although this definition sets apart the 'professional' and the 'amateur' historian, many of the 'amateur' historians I encountered through my research have skills and attitudes similar to those possessed by 'professional historians'.³⁶ This accords with Charles Leadbeater and Paul Miller's coining of the term 'pro-ams' to describe 'innovative, committed and networked amateurs working to professional standards'.³⁷ Accordingly, within this thesis I have used 'community' historians to refer to family, local and other historians who undertake historical research as a voluntary activity.³⁸ This term does not describe their training as a historian, which may include academic training or professional historical experience.

Research questions

The main questions that shaped this research project are, firstly: how do faculty, community historians and the public evaluate, use, and contribute to digital history

standards for professional historians. The Royal Historical Society accepts as members those with 'a professional involvement, teaching role or engagement with historical work'. Royal Historical Society, 'Membership - Notes for Applicants', *Royal Historical Society*, accessed 21 May 2015, <http://royalhistsoc.org/membership/members/membership-notes-for-applicants/>.

³⁵ American Historical Association, 'Statement on Standards of Professional Conduct'.

³⁶ As Raphael Samuel said, 'if history was thought of as an activity rather than a profession, then the number of its practitioners would be legion'. Raphael Samuel, *Theatres of Memory. Volume 1, Past and Present in Contemporary Culture* (London: Verso, 1994). p. 17.

³⁷ Charles Leadbeater and Paul Miller, 'The Pro-Am Revolution: How Enthusiasts Are Changing Our Economy and Society' (Demos, 2004), <http://www.demos.co.uk/publications/proameconomy>.

³⁸ Other definitions include 'history research and writing' outside university history departments. Alison Twells, 'Community History', *Making History*, 2008, http://www.history.ac.uk/makinghistory/resources/articles/community_history.html.

resources? And consequently, what impact has digitality had on public participation and scholarly practices in historical research?

In order to answer these questions, sub-questions developed over the course of the research include:

- What form do current participatory history projects take?
- What can be learnt from more and less successful heritage crowdsourcing projects?
- How might participation in heritage crowdsourcing encourage the development of historical interests and skills?
- How are digital technologies changing (or not) the research practices of different types of historians?
- Do historians evaluate online resources differently from physical resources?
- When and how do historians currently share data or collaborate?
- Have faculty and community historians responded differently to the potential of digital history, particularly the availability of digitally-inflected models for sharing data and contributing to historical websites?
- How transformative are digital resources, tools and methods?

Research design and data collection methods

My main methods are an extensive, interdisciplinary literature review, heuristic evaluations of participatory history websites, trace ethnography of online communications between projects and participants, and semi-structured interviews with historians. I have described the methods briefly here and provided more detail at relevant points in later chapters. My literature review included topics such as cognitive surplus,³⁹ collective

³⁹ The constructive use of free time, for example, editing *Wikipedia* articles instead of watching television. Clay Shirky, *Cognitive Surplus: Creativity and Generosity in a Connected Age*. (London, *Introduction: Making digital history*

intelligence,⁴⁰ human computation,⁴¹ educational psychology, computer-supported cooperative work, and user experience design and human-computer interaction. As crowdsourcing is a fast-moving field (generally moving faster than academic publishing timelines), and social media is one of the communication channels commonly used by digital projects, I have also included informal publications such as blog posts in my review. The historians' interviews, and a supplementary interview piloted with project stakeholders,⁴² were approved by The Open University Human Research Ethics Committee (references HREC/2012/1152/1, HREC/2014/1152/Ridge/1).

Within human-computer interaction, heuristics can be thought of as 'rules of thumb' derived from formal reviews of known usability problems.⁴³ In heuristic evaluations, evaluators examine ('inspect') an interface and assess how well it complies with recognised usability principles, codified as heuristics.⁴⁴ I used heuristic evaluations when conducting the overview of participatory digital history sites, and conducted more detailed inspections

U.K.: Penguin, 2011).

⁴⁰ Systems in which human participation is directed by computational systems or processes. Alexander J. Quinn and Benjamin B. Bederson, 'Human Computation: A Survey and Taxonomy of a Growing Field', in *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems*, 2011, 1403–1412, <http://dl.acm.org/citation.cfm?id=1979148>.

⁴¹ Human computation is described by Law and von Ahn as 'using human effort to perform tasks that computers cannot yet perform, usually in an enjoyable manner'. Edith Law and Luis von Ahn, 'Input-Agreement: A New Mechanism for Collecting Data Using Human Computation Games', in *Proceedings of the 27th International Conference on Human Factors in Computing Systems* (Computer Human Interaction - CHI, 2009, Boston, MA, USA, 2009), 1197–1206.

⁴² The 'Questionnaire for project stakeholders' is included in *Appendix C*.

⁴³ Nielsen's heuristics were based on a review of nearly 250 usability problems. Jakob Nielsen, '10 Usability Heuristics for User Interface Design', 1995, <http://www.nngroup.com/articles/ten-usability-heuristics/>.

⁴⁴ Jakob Nielsen, 'Finding Usability Problems through Heuristic Evaluation', in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (ACM, 1992), 373–380, <http://dl.acm.org/citation.cfm?id=142834>.

on projects analysed for the chapters on crowdsourcing and citizen history. I also analysed projects in terms of their visible affordances to help understand the actions they appeared to encourage, allow or disable. To help prioritise the attributes I looked for in reviews, I conducted four detailed 'think aloud'-style⁴⁵ sessions in which I noted my initial impressions of participatory sites from the moment of landing on the site homepage to completing its core task; the recording sheet I created has additional prompts to return to and note specific aspects of the site design.⁴⁶ Finally, my evaluations were also informed by informal observations of the reception of different projects by participants in workshops where I presented a range of crowdsourcing projects.

Trace ethnography is the analysis of 'documents and documentary traces'⁴⁷ that result from computer-mediated communication, yielding granular and 'rich qualitative insight' into user interactions. I applied trace ethnography to sources including project documentation, website interfaces, forums and social media posts. I chose this method because participant forums are a rich data source, often providing access to (almost) all posts made since the launch of a project.⁴⁸ At various points I was looking for social norms,⁴⁹ evidence for historical thinking, attitudes to decisions made by project

⁴⁵ Jakob Nielsen, 'Guerrilla HCI: Using Discount Usability Engineering to Penetrate the Intimidation Barrier', *Nielsen Norman Group*, 1994, http://www.useit.com/papers/guerrilla_hci.html.

⁴⁶ The sites reviewed with the full heuristic prompt sheet were the New York Public Library's *Building Inspector*, *Cymru1900Wales*, *Old Bailey Online* and *Operation War Diary*.

⁴⁷ R. Stuart Geiger and David Ribes, 'Trace Ethnography: Following Coordination through Documentary Practices', in *System Sciences (HICSS)*, 2011 44th Hawaii International Conference On (IEEE, 2011), 1-10, http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5718606.

⁴⁸ It should be noted that some forum archives are incomplete, and participants may be able to delete or edit posts after posting.

⁴⁹ The self-enforced rules and standards of a community. Abhay Sukumaran et al., 'Normative Influences on Thoughtful Online Participation', in *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems (CHI 2011, Vancouver, BC, Canada: ACM, 2011)*, 3401-3410, *Introduction: Making digital history*

stakeholders and the forms of help sought and received by participants. The results of this process also informed my interview design and heuristic evaluations of project sites.

I have used semi-structured interviews to investigate how and when faculty and community historians evaluate, use and contribute to collaborative digital history resources. Semi-structured interviews contain some pre-determined closed questions (including demographic data), and a core script with open questions. This combination helps ensure consistency, but also allows for in-depth or exploratory discussions that can generate rich data.⁵⁰ This combination of flexibility and consistency meant that I could adapt the interviews for different interviewees as necessary. By noting the tools mentioned in the interviews, I was able to link the methods, tools and software services mentioned to different research activities; this in turn fed into my review of related projects and informed my thinking about the potential for different types of digital history. I imported the interview transcripts into the *NVivo* software tool and analysed them using thematic analysis, a method for 'identifying, analysing and reporting patterns (themes) within data'.⁵¹ I had considered grounded theory, but 'pure' grounded theory methodology does not encourage the integration of external theories,⁵² and therefore thematic analysis is better suited for this HCI-informed research project. Muller and Kogan's discussion of the need for work in human-computer interaction and computer-supported cooperative work (CSCW) to consider formal theories at an early stage influenced this decision.⁵³

<http://dl.acm.org/citation.cfm?id=1979450>.

⁵⁰ Sharp, Rogers, and Preece, *Interaction Design: Beyond Human-Computer Interaction*. p. 298.

⁵¹ Virginia Braun and Victoria Clarke, 'Using Thematic Analysis in Psychology', *Qualitative Research in Psychology* 3, no. 2 (2006): 77-101, doi:10.1191/1478088706qp0630a.

⁵² Michael J Muller and Sandra Kogan, 'Grounded Theory Method in HCI and CSCW' (Cambridge, MA: IBM Center for Social Software, 2010).

⁵³ Muller and Kogan, 'Grounded Theory Method in HCI and CSCW'.

I had planned to use contextual inquiry, which combines a semi-structured interview with an observation (usually recorded through photos or video) of a participant performing the task relevant to the research topic in their usual place of work. In theory, contextual inquiry combines the strengths of both interviews and observations, as observational data is more reliable than self-reported behaviours, and (unlike traditional observation), it allows discussion of the participant's actions. However, in practice, I found that contextual inquiry was impractical. In the first instance, many participants did their research in their own homes, and documenting these spaces with images was more intrusive than documenting semi-public spaces like campus offices. The requirement for participants to have representative tasks to hand for the observation phase also added considerable logistical difficulties to the process of organising interview sessions. Finally, the rich dataset generated did not necessarily yield any additional insight.⁵⁴ After deciding not to continue with contextual inquiry, I discovered that others had also found it problematic, though under different circumstances,⁵⁵ and for that reason it seems appropriate to report my own experiences with it here.

I conducted formal interviews or informal discussions with stakeholders from projects and organisations including *Founders and Survivors*, the *Ur Crowdsourcing* project, *Old Bailey Online*, *London Lives*, *Zooniverse*, *HistoryPin*, the Public Records Office of Victoria, *Marine Lives*, *Children of the Lodz Ghetto*, *Your Paintings Tagger* and *London Street Views 1840*. Some projects also shared internal evaluation reports and participant survey data. These have been cited where they have provided more than background information. Following

⁵⁴ Observations, in particular, generate a lot of data that is 'difficult to analyse unless a structured framework is adopted'. Sharp, Rogers, and Preece, *Interaction Design: Beyond Human-Computer Interaction*. p. 358.

⁵⁵ Jim Ross, 'Why Are Contextual Inquiries So Difficult?', *UXmatters*, 4 June 2012, <http://www.uxmatters.com/mt/archives/2012/06/why-are-contextual-inquiries-so-difficult.php>.

the suggestion of an interviewee, I also undertook some family history and some local history research (as a form of action research) to better understand the experiences of community historians using digital sources and tools. I also undertook a small-scale survey, which received 13 responses.⁵⁶

I have used purposive sampling when reviewing participatory, digital history projects. Purposive sampling involves selecting items for study based on attributes of interest for the research question.⁵⁷ I have located relevant projects through keyword searches in online journal repositories, conference proceedings, social media and web searches. I sought to identify a comprehensive range of projects rather than exhaustive list of (for example) every organisation with a collection on a particular platform, looking for typical as well as unusual or unique projects. My focus is on projects relevant to research on early

⁵⁶ The survey was designed to obtain references for crowdsourcing projects my sampling process may have missed (particularly projects in which participants learnt new skills) and collect information about the challenges of managing participatory history projects. The survey was open for less than a fortnight and promoted via the Museums Computer Group and Museums Computer Network mailing lists (<https://www.mail-archive.com/mcn-l@mcn.edu/msg07395.html>, <https://www.jiscmail.ac.uk/cgi-bin/webadmin?A2=mcg;6146fc50.1406>, both last accessed 10 June 2015) and my personal LinkedIn and academia.edu accounts.

⁵⁷ C. Teddlie and F. Yu, 'Mixed Methods Sampling: A Typology With Examples', *Journal of Mixed Methods Research* 1, no. 1 (1 January 2007): 77–100, doi:10.1177/2345678906292430. Here I am following citizen science researchers Wiggins and Crowston in using a purposive rather than a probabilistic sampling method. Andrea Wiggins and Kevin Crowston, 'From Conservation to Crowdsourcing: A Typology of Citizen Science', in *Proceedings of the 44th Annual Hawaii International Conference on System Sciences* (2011 44th Hawaii International Conference on System Sciences, Koloa, Hawaii, 2011), doi:10.1109/HICSS.2011.207. The limitations of this sampling method means that statistical statements about the projects reviewed would not be valid, and therefore I have used qualitative rather than quantitative statements when discussing projects. For example, without knowing how many participatory history sites currently exist, it is difficult to say what proportion of them has a certain characteristic. This point is also made in Terras, 'Digital Curiosities'.

modern England but the review also includes relevant projects from other regions and periods that have helped shape the field of participatory digital history. Some significant scholarly projects from other disciplines that might provide models for digital history projects have also been included in this study.

I have used a combination of methods for collecting data about websites reviewed. I have collected the title and URL for each website reviewed (the URLs for each site mentioned in the body of the thesis are listed in *Appendix A, Websites reviewed*). I have taken screenshots of website front pages, and in many cases also recorded content item pages and pages related to their core participatory tasks.⁵⁸ I have saved project descriptions from published literature, marketing material, blog posts and social media as *Zotero* snapshots or as web pages. I have tested many participatory sites by undertaking their core tasks, and explored repositories by browsing and using search functions to look for a specific type of item appropriate to the stated scope of the site.⁵⁹ My analysis is fundamentally informed by the idea that any website or software application is embedded in social contexts and histories that affect how it is received and used.⁶⁰ Therefore, in some instances I have also

⁵⁸ I use the Firefox browser extension *Aviary* to capture the full length of a web page (as most screenshots only capture the area shown on-screen). The company no longer supports the extension so where necessary I also use a plugin call *qSnap*. These screenshots are available on request.

⁵⁹ As genres of sites began to emerge and I became familiar with the various underlying platforms, I simplified my recording system.

⁶⁰ For related discussion, see Kling's work on 'web models' and more recent work on 'assemblages' by Prestopnik and Crowston, and Reed, Rodriguez and Rickhoff. Kling's web models later became 'Socio-Technical Interaction Network' (STIN), as described by Meyer. Rob Kling, 'Computerization and Social Transformations', *Science, Technology, & Human Values* 16, no. 3 (1991): 342–67, <http://www.jstor.org/stable/689919>. Eric T. Meyer, 'Socio-Technical Interaction Networks: A Discussion of the Strengths, Weaknesses and Future of Kling's STIN Model', in *Social Informatics: An Information Society for All? In Remembrance of Rob Kling*, ed. J. Berleur, M.I. Numinen, and J. Impagliazzo, vol. 223, IFIP International Federation for Information Processing (Boston: Springer,

sampled participant forum discussions and searched for project reviews on the web and social media.

I reviewed the HTML source of web pages to determine the platforms on which sites had been built. The website builtwith.com⁶¹ has also been useful for detecting common content management or blogging systems such as *WordPress* (however it is not able to detect specialised platforms like *Scripto* or *Scribe*). I have checked links provided in *Appendix A* manually in May 2015, and via the *WordPress* plugin Broken Link Checker in June 2015.⁶²

Thesis structure

Following a broad overview of participatory digital history, this thesis moves from a review of public history projects focused on crowdsourcing, to a smaller sub-set of projects related to citizen history, and then focuses on the research practices of faculty and community historians. The *Conclusion* draws together the various models for 'making history' discussed, and teases out the ways in which they have been affected by digital methods, tools and resources.

2006), 37–48. Nathan R. Prestopnik and Kevin Crowston, 'Citizen Science System Assemblages: Understanding the Technologies That Support Crowdsourced Science' (iConference 2012, Toronto, Canada, 2012), 7–10. Jason Reed, Whitney Rodriguez, and Angelique Rickhoff, 'A Framework for Defining and Describing Key Design Features of Virtual Citizen Science Projects', in *Proceedings of the 2012 IConference* (ACM, 2012), 623–625, <http://dl.acm.org/citation.cfm?id=2132314>. Theoretical frameworks like cultural-historical activity theory (CHAT) also provide useful prompts to consider how socio-technical systems are embedded in the wider contexts of communities, rules and divisions of labour, but I did not formally apply CHAT in this thesis. David Allen, Stan Karanasios, and Mira Slavova, 'Working With Activity Theory: Context, Technology, and Information Behavior', *Journal of the American Society for Information Science* 62, no. 4 (2011): 776–788, doi:10.1002/asi.

⁶¹ <http://builtwith.com>, last accessed 11 June 2015

⁶² Created by Janis Elsts and available from <https://wordpress.org/plugins/broken-link-checker/>. Last accessed 8 June 2015.

Supporting material is provided in three Appendices. *Appendix A, Websites reviewed*, lists the titles and URLs for the websites I reviewed for my research. *Appendix B, Interview participants*, contains a summary of interview participants, including relevant demographic information. *Appendix C, Interview and survey questions*, provides the interview scripts used in the semi-structured interviews with family, local and faculty historians, a script written for interviews with project stakeholders, and survey text.⁶³

Chapter 1: An overview of participatory digital history projects presents an overview of over 400 digital history projects that aim to engage the public and/or collect, create or enhance records about historical materials for scholarly and general audiences. This overview provides context for the analysis of projects examined in greater detail in later chapters, and gives an indication of emerging norms or patterns of presenting and interacting with historical materials. Projects are grouped by their main tangible outputs, supporting a more nuanced understanding of the impact of different design choices.

Chapter 2: History with the public: crowdsourcing discusses several design factors that may influence the success of crowdsourcing projects, including task design, the potential participant's initial experience of a site, and the role of project marketing and communications in connecting to potential motivations for participation. It also analyses the role of participant forums and the provision of more complex tasks in keeping participants interested in a project.

Following this, *Chapter 3: History with the public: from microtasker to historian?* explores how some crowdsourcing projects encourage deeper engagement with history or science.

⁶³ As Appendices are included in the word count, the interview transcripts could not be included.

It investigates the project attributes that may help people learn historical skills, or even begin to 'become historians', through their participation in historical crowdsourcing projects. It suggests that there are three types of citizen history projects: crowdsourcing projects that accidentally support citizen history; crowdsourcing projects that hope for citizen history but are not built around it; and citizen history projects that can succeed only if participants are able to learn or bring some disciplinary skills to the more complex tasks that contribute to the projects' goals. Projects that support citizen history do so by providing: opportunities for participants to actively engage with historical materials while undertaking meaningful tasks; access to both the historical materials and project data; access to a community similarly engaged with the meaningful goals of the wider project; and a visible and accessible expert presence.

Chapter 4: Historians' working practices and digital tools, resources and methods shifts our attention from public participation to scholarly practices in historical research. It presents the results of 29 interviews conducted in 2012 with faculty and community historians. It contributes empirical data on how faculty, family and local historians evaluate, use and contribute to 'traditional' and participative digital resources. It finds that community historians are generally more likely than faculty historians to engage in sharing data, but they are also still likely to be selective about the information they share publically and privately. These interviews show the impact that digital tools, resources and methods have had on the processes of discovering, evaluating, gathering, creating, and sharing information for historical research.

Finally, in *Conclusion: The impact of digitality on public participation and scholarly practices in history*, I consider the impact of the digital projects, platforms and paradigms discussed in previous chapters on public participation and scholarly practices in historical

research. Through an examination of the results of the interviews and analyses of participatory history projects, I argue that digitality has already enhanced many historical practices and has increased the number of those engaged in making history.

Chapter 1: An overview of participatory digital history projects

To understand the range of approaches to participatory and scholarly digital history currently available, this chapter presents an overview of over 400 digital history projects, from *Ancient Lives* to *Zotero Commons*. It contributes to our knowledge of how the public, and faculty and community historians make history by creating and using digital history resources. The projects included in this study aim to engage the public and/or collect, create or enhance records about historical materials for scholarly and general audiences. This overview provides context for the analysis of projects examined in later chapters, and an indication of emerging norms for presenting and interacting with historical materials. The process of visiting the sites, trying out the tasks offered, reading communications from and between project stakeholders and participants, and comparing the user experience design with similar projects also provided invaluable background knowledge and yielded insights that have informed the thesis that follows.

I have grouped together different projects with similar types of final output. This allows comparisons to be made between projects with similar outputs but different source materials, task types, participant interests, motivational frameworks or management structures. These comparisons support a more nuanced understanding of the intersection of different design choices, participant motivations and goals. Conversely, organising projects by task can mask the huge variations in intent, skill and information produced through a task like 'annotation' or 'metadata creation'. For example, the *Galaxy Zoo* project

uses a decision-tree¹ process to categorise the visual characteristics of items, while a transcription project like *Transcribe Bentham* asks participants to use a specific markup format to label passages within a text. While both produce forms of metadata related to the item, the tasks that create the metadata and the form of the data differ widely. Similarly, the uses to which the results are put will depend on the project. The classification data collected by *Galaxy Zoo* may be processed by statistical software, while the *Transcribe Bentham* data might be used to supplement archival finding aids; both may lead to wider public engagement as well as traditional academic publications. By collating different approaches to achieving the same outputs, this typology reduces the risk of conflating the output with the activities required to achieve it. It also provides a guide to the various combinations of tasks and skills available for projects wishing to achieve particular types of outputs.

The chapter begins with a brief methods section that discusses sampling, data recording, and the design attributes considered in the review process. The projects most relevant to the question of public participation and scholarly practices around collaborative digital resources are then grouped by their main tangible outputs: partial- and full-text transcription, text correction and proofreading, collections, information, repositories, and other outputs.

¹ Decision trees classify items by asking a series of questions with pre-determined answers; each answer adds detail to a classification and moves it down a particular path. A simple analogy is the way questions in a game of 'Animal, Vegetable, Mineral?' or '20 Questions' cumulatively function to more precisely identify an object. A more scholarly explanation is available in Kyle W. Willett et al., 'Galaxy Zoo 2: Detailed Morphological Classifications for 304 122 Galaxies from the Sloan Digital Sky Survey', *Monthly Notices of the Royal Astronomical Society*, 2013, stt1458, <http://mnras.oxfordjournals.org/content/early/2013/09/15/mnras.stt1458.short>.

Methods

As discussed in the *Introduction*, I undertook purposive sampling, seeking to identify a comprehensive range of projects and platforms.² The title and URL for each website are listed in *Appendix A, Websites reviewed*. The specific project attributes presented in this thesis are the result of iteratively developed lists of potential heuristics, informed by traits present in the sites themselves and by the existing literature on public participation in scientific research,³ crowdsourcing in the humanities,⁴ cultural heritage⁵ and amateur digitisation.⁶ As my research questions became more focused, some groups of attributes grew in importance while others that provided fewer insights were dropped. The final attributes reviewed relate to project management, source material, communication and marketing, graphic and interaction design, and aspects specific to encouraging initial and on-going engagement with participatory projects.

Attributes related to project management and set-up include the projects' stated aim, start date, institutional stakeholders and their relationship to the project, stage when reviewed (e.g. proposal, beta or mature project), any significant changes in project direction over time, and target and actual audiences for a project. Broad classifications within these attributes include whether projects were open to contributions of content and/or work;

² As used by similar research projects such as Wiggins and Crowston, 'From Conservation to Crowdsourcing: A Typology of Citizen Science'. While the range of projects aims to be comprehensive, the review does not attempt to include every single instance of a type of project. Similarly, I have not attempted to present a complete list of potentially informative attributes.

³ Rick Bonney et al., 'Public Participation in Scientific Research: Defining the Field and Assessing Its Potential for Informal Science Education. A CAISE Inquiry Group Report' (Washington D.C.: Center for Advancement of Informal Science Education (CAISE), July 2009), <http://caise.insci.org/uploads/docs/PPSR%20report%20FINAL.pdf>.

⁴ Dunn and Hedges, 'Crowd-Sourcing Scoping Study'.

⁵ Oomen and Aroyo, 'Crowdsourcing in the Cultural Heritage Domain'.

⁶ Terras, 'Digital Curiosities'.

the type of work;⁷ their geographic and temporal scope; whether the chief aim was public engagement or outreach, crowdsourcing, or scholarly or community history use; and the types of source material. Where possible I collected indications of the projects' success such as evidence for progress towards stated goals.

Data about source material include the size of the source dataset or repository, the size and format of items (e.g. from entire books to single images or lines of text), and how dependent parts of an item were on the whole.⁸ The combination of input materials, output types and participant tasks (the sequences of actions taken to produce the output) is also noted.

Aspects of project communication and marketing include project titles and straplines,⁹ the use of social media, and the extent and impact of traditional mass media coverage. This information was gathered from sources including site links, peer-reviewed publications, web searches and participant discussions about how they discovered a project. Statements demonstrating the impact of participation and updates of progress towards stated goals are considered part of project marketing or outreach, as are other motivational messages related to participatory tasks.¹⁰ Marketing and behavioural change literature on motivation, ability (including financial resources, time budget, physical capacities and mental resources),¹¹ and opportunity (or the absence of barriers to action)¹² are also

⁷ For example, transcription, geolocation, metadata, collections, quality control, research or analysis tasks.

⁸ For example, tabular data in self-contained documents, narrative texts ordered by date or place, or single images or maps.

⁹ A secondary heading used with a website title; also known as a 'tagline'.

¹⁰ For example, displays of the percentage of sources processed or posts with stories of discoveries during the task.

¹¹ Mia Stokmans, 'MAO-Model of Audience Development: Some Theoretical Elaborations and

considered.

Interface elements include the types of search and browse functions available (particularly for larger sites), the types of data visualisations (generally maps or timelines) available for exploring content on the site, the visual metaphors underlying the interface and the overall quality of the user experience.¹³ Usability (or 'ease-of-use')¹⁴ is particularly important for reducing the barriers to using a website, while positive aspects of design include features that encourage participation and use. Negative features noted during the review include delays between completing a task and the results being available in the project; the requirement to register an account before being able to try the activities on a website; broken links or spam; and the absence of information that suggests recent activity by others. Positive design features for participatory sites include the use of instructions and guided tasks to help novice users learn how to do specific activities, the timing and quality of feedback on actions, and evidence of quality control methods.¹⁵

Practical Consequences', in *Proceedings on CD-Rom of the 8th International Conference on Arts and Cultural Management* (8th International Conference on Arts and Cultural Management, Montréal, Québec, Canada, 2005), 1–11, http://neumann.hec.ca/aimac2005/PDF_Text/Stockmans_Mia.pdf.

¹² Jennifer International Journal, 'Motivation, Ability and Opportunity to Participate: A Reconceptualization of the RAND Model of Audience Development', *International Journal of Arts Management* 7, no. 1 (2004).

¹³ Here I drew particularly on previous study and experience in human-computer interaction and usability testing, as well as references from citizen science and participatory design.

¹⁴ Reed, Rodriguez, and Rickhoff, 'A Framework for Defining and Describing Key Design Features of Virtual Citizen Science Projects'.

¹⁵ For example, quality control methods outlined by Quinn and Bederson include output agreement, input agreement, economic models, defensive task design, and ground truth seeding. Quinn and Bederson, 'Human Computation'. Also discussed by Yochai Benkler, 'Coase's Penguin, or, Linux and "The Nature of the Firm"', *The Yale Law Journal* 112, no. 3 (December 2002): 369–446, <http://www.jstor.org/stable/1562247>.

Attributes specific to participatory projects include presence or absence of text and design elements that connect with likely participant motivations and reward structures for participants.¹⁶ The presence of a community around the project, the integration of any community functions into a project interface and the relationship between project stakeholders and communities (where possible) is noted, as is any evidence for in-person events and pre-existing relationships with relevant communities. Information about the effectiveness of a website such as the participation rate,¹⁷ repeat participants or the number of tasks per participant is recorded, when available. The size and complexity of participatory tasks is relevant, and following Quinn and Bederson's overview of 'human computation',¹⁸ I consider the 'human skills' required (such as visual recognition, language understanding) for participatory tasks. The work of theorists of the 'commons'¹⁹ and collaborative repositories²⁰ provide further background for my review. Jessop's discussion of early 'data provider' projects²¹ is a reminder that some digitisation projects provide better support than others for scholarly needs (such as analysing or exporting the data provided). Finally, where possible I have noted the underlying participatory task platform and the format of any related community platforms.

¹⁶ Including explicit rewards such as financial compensation, competitive elements like participant leaderboards and other visible markers of contributions, and recognition in scholarly publications.

¹⁷ The number of overall visitors to a site compared to the number who took part in at least one task.

¹⁸ Quinn and Bederson, 'Human Computation'.

¹⁹ Benkler, 'Coase's Penguin'.

²⁰ Naren B. Peddibhotla and Mani R. Subramani, 'Contributing to Public Document Repositories: A Critical Mass Theory Perspective', *Organization Studies* 28, no. 3 (1 March 2007): 327–46, doi:10.1177/0170840607076002.

²¹ Martyn Jessop, 'The Inhibition of Geographical Information in Digital Humanities Scholarship', *Literary and Linguistic Computing* 23, no. 1 (December 2007): 39–50, doi:10.1093/lc/fqm041.

Pre-digital participatory practice

A brief discussion of pre-digital participatory projects will set the scene for later analysis of digital practices and platforms, and highlight the impact of networked, digital technologies. The history of what became the *International Genealogical Index (IGI)* and *FamilySearch* provides a sense of the impact of changing technologies on historical research and participatory practices.²² The Church of the Latter Day Saints (LDS) began creating microfilm records in 1938,²³ allowing them to collect vast numbers of primary sources in a central location without the financial and logistical overhead of acquiring and managing physical copies of historical documents. In 1969 the LDS launched a system called *GIANT* (aka the Names Tabulation Program) that could centralise the data entry process, apply standardised place names and surnames, and de-duplicate indexed names.²⁴ A later system put copies of the *IGI* on compact discs, allowing people to run name searches on their own computers, markedly increasing productivity.²⁵ In 1987, a 'cooperative indexing' project to index the 1881 British census was initiated.²⁶ According to LDS material, nearly 'seven tons' of photocopies were assigned and distributed to local

²² As a religiously, rather than historically motivated project, the motivations behind *FamilySearch* may be slightly atypical, but their cooperative indexing projects also provide a good example of collaboration between different family history societies.

²³ James B. Allen, Jessie L. Embry, and Kahlile B. Mehr, 'Appendix I: Historical Milestones', *Brigham Young University Studies* 34, no. 2 (1994): 343–346, <https://journals.lib.byu.edu/spc/index.php/BYUStudies/article/viewFile/6198/5848>.

²⁴ While 'cooperative indexing' was introduced in 1988 to allow people outside the LDS Church to index records, *GIANT* was designed for internal use. James B. Allen, Jessie L. Embry, and Kahlile B. Mehr, 'Chapter 8: Automating the Records', *Brigham Young University Studies* 34, no. 2 (1994): 303–337, <http://www.jstor.org/stable/43042189>.

²⁵ Allen, Embry, and Mehr, 'Chapter 8'. p. 309.

²⁶ Stephen C. Young, 'Working Together', *Ensign*, March 1996, <https://www.lds.org/ensign/1996/03/working-together?lang=eng>. See also Richard Tice, 'Doing Temple Work at Home', *Ensign*, July 1990, <https://www.lds.org/ensign/1990/07/doing-temple-work-at-home?lang=eng>.

groups for double transcription and checking. After a second round of checks and corrections, each index entry was double-keyed into a bespoke software application, then redistributed to local groups for checking. Finally, the data was mailed to Salt Lake City on 'more than eleven thousand computer disks', where it was loaded, sorted and indexed on a county-by-county basis.²⁷ Online indexing interfaces for *FamilySearch* volunteers further transformed the work of distributing and validating records, while also making the transcription task easier for participants to access.²⁸ By 2009 'more than 100,000' volunteers were indexing half a million individual names per day.²⁹ On a smaller scale, participants working on 'family reconstitution' for the Cambridge Group for the History of Population and Social Structure moved from writing and collating paper records to computers. Computerisation made data entry more efficient and enabled automatic links between entries; later the data compiled by local scholars could be statistically analysed far more quickly compared to earlier technologies.³⁰

The sciences have a long history of participatory projects, particularly asking the public to contribute observations of natural phenomena,³¹ and a more recent history of public

²⁷ Young, 'Working Together'.

²⁸ For example, early contributors to *FamilySearch Indexing* recall posting 3 1/2" disks containing transcribed data to be checked then posted onwards. Chester R Heglund, 'Comment on 2014 Indexing Year in Review', *FamilySearch Blog*, 24 January 2015.

²⁹ Heather Whittle Wrigley, 'Technology Helps FamilySearch Volunteers Hit Major Milestone', *Liahona*, December 2009, <https://www.lds.org/liahona/2009/12/news-of-the-church?lang=eng>.

³⁰ E. A. Wrigley, 'Small-Scale but Not Parochial: The Work of the Cambridge Group for the History of Population and Social Structure', *Family & Community History* 1, no. 1 (1998): 27–36, <http://www.maneyonline.com/doi/abs/10.1179/fch.1998.1.1.003>.

³¹ See, for example, the special edition of *Science in Context* (2011) on 'Lay Participation in the History of Scientific Observation', Anne Secord, 'Corresponding Interests: Artisans and Gentlemen in Nineteenth-Century Natural History', *The British Journal for the History of Science* 27, no. 04 (1994): 383–408, doi:10.1017/S0007087400032416. Jonathan Silvertown, 'A New Dawn for Citizen Science.', *Trends in Ecology & Evolution* 24, no. 9 (September 2009): 467–71,

participation in scientific research.³² Pre-digital technologies that allowed the reproduction of records and transmission of information hint at the impact of digital networks on distributed participation. For example, Vetter discusses the impact of the telegraph (which enabled faster, but not instantaneous, long-distance communication) on field observation projects.³³ Goldstein and Brenna both describe the role of reference materials such as books and scientific journals in helping observers of natural history provide accurate reports;³⁴ modern projects are able to supply online images, references, open access articles, and community forums to help participants.

Digital participatory projects benefit from computational techniques that automate much of the coordination overhead of distributing tasks, providing feedback for volunteers, and aggregating and verifying their contributions. Online communication platforms allow projects to reach both broad and niche groups through loose networks. Online volunteer portals have replaced traditional volunteer bureaus in matching people to opportunities. The specialist information and support networks required to support participants are instantly available, and online platforms make help pages and discussion archives

doi:10.1016/j.tree.2009.03.017. Jeremy Vetter, 'Introduction: Lay Participation in the History of Scientific Observation', *Science in Context* 24, no. 02 (April 2011): 127–141,

doi:10.1017/S0269889711000032. Robert E Kohler, 'Finders, Keepers: Collecting Sciences and Collecting Practice', *History of Science* 45, no. 4 (2007): 428–455, doi:10.1177/007327530704500403.

³² Bonney et al., 'Public Participation in Scientific Research'.

³³ Jeremy Vetter, 'Lay Observers, Telegraph Lines, and Kansas Weather: The Field Network as a Mode of Knowledge Production', *Science in Context* 24, no. 02 (April 2011): 259–280, doi:10.1017/S0269889711000093.

³⁴ Daniel Goldstein, "Yours for Science": The Smithsonian Institution's Correspondents and the Shape of Scientific Community in Nineteenth-Century America', *Isis* 85, no. 4 (December 1994): 573–599, <http://www.jstor.org/stable/235279>. Brita Brenna, 'Clergymen Abiding in the Fields: The Making of the Naturalist Observer in Eighteenth-Century Norwegian Natural History', *Science in Context* 24, no. 02 (April 2011): 143–166, doi:10.1017/S0269889711000044.

searchable. Digital technologies have changed the speed, accuracy, and scale of participatory projects, while online communities have helped transform the volunteers' experience. Technology has also helped manage the limitations of physical space, conservation, location, and opening hours, all of which previously affected access to historical collections.

Types of participatory projects discussed

Before providing examples for different types of projects, I will briefly describe the different outputs from participatory projects I have focused on in this thesis. Many activities and projects produce multiple outputs, so these groups are necessarily non-exclusive.³⁵ Tasks range from simple 'microtasks' - small, self-contained actions that can be completed quickly - to complex research tasks.³⁶ Projects are grouped by output types as follows: partial text transcription; full text transcription; proofread and corrected versions of computationally-transcribed text; collections of historical materials, information including observations and personal or specialist knowledge; and other outputs. I have also included repositories with additional participatory functions.

Text-based transcription and correction tasks are generally relatively simple 'type what you see' tasks, and as such do not require much decision-making.³⁷ Text transcription or

³⁵ For example, entering text into a labelled database field in an indexing task is arguably a form of metadata creation as well as indexing or partial transcription, and the process of checking submitted transcriptions is a form of proofreading.

³⁶ The impact of task design is considered in the next chapter. I first encountered the term 'microtasks' in Jane McGonigal, 'Engagement Economy: The Future of Massively Scaled Collaboration and Participation' (Palo Alto: Institute for the Future, 2008).

³⁷ However, this varies with the source. For older texts, semi-diplomatic transcriptions, in which the transcription is made more readable than the original, can require judgement, while diplomatic transcriptions may require more technical skills to represent the source text exactly as it is written. 'Transcription Conventions', *Scriptorium: Medieval and Early Modern Manuscripts Online*, accessed

correction projects share some issues, including the relationship between task difficulty and unclear, faint or obscured handwriting or typefaces; archaic or foreign languages; and the quality of images.³⁸ They also share a requirement for interfaces which allow for the simultaneous display of transcribed and original texts on the same screen as the data entry form. Source materials for these tasks potentially include any artefact with text, including inscriptions, tombstones, books, manuscripts, newspapers, and structured records such as birth, death and marriages certificates or indexes. The amount of structure within a document (for example, tables or forms versus letters) varies, and is one factor in determining the transcription format.

24 February 2015, <http://scriptorium.english.cam.ac.uk/handwriting/materials/conventions/>.

³⁸ Including low-resolution images or those taken from damaged pages, microfilm or photocopies.

Table 1 Output types with sample projects

Text (type what you see)		
Partial text (indexing)		<i>FamilySearch Indexing</i> <i>World Archives Project</i> <i>FreeBMD</i> <i>Old Weather</i> <i>Operation War Diary</i> <i>Herbaria@Home</i>
Full text	Structured	<i>Transcribe Bentham</i> <i>DIY History</i> <i>Letters of 1916</i>
	Unstructured	<i>Project Gutenberg</i> <i>Flickr-based sites</i>
Proof-read / corrected text		
	OCR correction microtasks	<i>Trove newspaper correction</i> <i>BnF Correct</i> <i>TypeWright</i>
	Smooth reading tasks	<i>Distributed Proofreaders</i>
Collections		
	Information and/or digital records	<i>My Brighton and Hove</i> <i>HistoryPin</i> <i>Europeana 1914-1918</i> <i>History Harvest</i> <i>Billion Graves</i>
Information		
Information types	Metadata	<i>steve.museum</i> <i>Your Paintings Tagger</i>
	Georeferencing	<i>What's On The Menu Geotagger</i> <i>British Library Georeferencer</i>
	Personal and experiential	<i>Polar Bear Expedition Digital Collections</i>
	Expert	<i>Arago</i>
Information task formats	Annotations	<i>Pepys Diary</i>
	History mysteries	<i>Year of the Bay</i> <i>10 Most Wanted</i>
	Research tasks - distributed reading	<i>Civil War Pathways</i>
	Research tasks - research commissions	<i>Children of the Lodz Ghetto</i>
	Research tasks - research quests	<i>Family and Community Historical Research Society projects</i>
Repositories with participatory functions		
		<i>Wikimedia Commons</i> <i>Parallel Archive</i> <i>London Lives</i> <i>FindMyPast</i>
Other outputs		
		<i>Micropasts</i> <i>LibriVox</i>

Partial text transcription

Partial text transcription (also called 'indexing' when it is designed to create a searchable index of selected terms in a document) usually involves typing specific information from the source into a data entry form.³⁹ Indexing is usually a relatively uncomplicated task but the requirements to identify sections of text to be indexed for partial transcription, to decipher difficult handwriting, and to understand abbreviations may provide some intellectual challenges. Indexing is particularly common for biographical records,⁴⁰ and a range of commercial and grassroots (self-organised) projects are focused on indexing genealogical records. Projects may have once relied on participants physically visiting archives to access original or microfilmed documents but now generally work with online images.⁴¹

Some of the largest transcription projects are run by genealogy organisations indexing sources of biographical information. FamilySearch (which claims to be the 'largest genealogy organization in the world'⁴² and is run primarily for the benefit of The Church of Jesus Christ of Latter-day Saints' 'temple ordinances'⁴³) launched *FamilySearch Indexing* on the web in 2005⁴⁴ and claimed to have indexed over one billion historic records by 2014.⁴⁵ In 2014, almost 320,000 volunteers helped index and verify over 160 million

³⁹ Outside formal projects, of course, many researchers make partial transcriptions from source materials.

⁴⁰ In part because of the preponderance of family history projects.

⁴¹ This places greater weight on an organisation's ability to fund the digitisation of their holdings.

⁴² 'About FamilySearch', *FamilySearch.Org*, accessed 12 February 2015, <https://familysearch.org/about>.

⁴³ 'Family History-Ancestry or Genealogy', *Mormon.Org*, accessed 12 February 2015, <http://www.mormon.org/values/family-history>.

⁴⁴ Holley, 'Crowdsourcing: How and Why Should Libraries Do It?'

⁴⁵ Paul G. Nauta, 'Massive Online US Obituaries Project Will Help Find Your Ancestors',

records.⁴⁶ FamilySearch have the resources to experiment with new methods for encouraging participation, including special transcription challenges such as the 1940 US Census,⁴⁷ a community wiki for research advice,⁴⁸ and indexing applications for mobile phones.⁴⁹ While Ancestry is largely a commercial repository of family history records,⁵⁰ they also run the *World Archives Project*⁵¹ in which 'more than 130,000 contributors have helped to index more than 200 million records' since 2008.⁵² The project provides detailed guidance about the contents of different projects and how to transcribe the information contained within the records.⁵³ The project has an important role in helping save records from small organisations that might otherwise be lost to natural disaster or decay,⁵⁴ but

FamilySearch Blog, 1 October 2014, <https://familysearch.org/blog/en/massive-online-obituaries-project-find-ancestors/>.

⁴⁶ Spencer Ngatuvai, '2014 Indexing Year in Review', *FamilySearch Blog*, 2 January 2015, <https://familysearch.org/blog/en/2014-indexing-year-review/>.

⁴⁷ 1940 US Census Community Project, 'We Did It! The 1940 US Census Community Project', 2012, <http://us2.campaign-archive2.com/?u=bode542dc933cfcb848d187ea&id=c6e095aa92>.

⁴⁸ 'FamilySearch Wiki:Introduction', *FamilySearch.Org*, 28 March 2014, http://familysearch.org/learn/wiki/en/The_Vision:_Why_We_Built_FamilySearch_Wiki.

⁴⁹ Michael Judson, 'The Mobile App—What We've Learned', *FamilySearch Blog*, 14 December 2012, <https://familysearch.org/blog/en/mobile-app-what-we-learned/>.

⁵⁰ They claim to provide '9 billion family history records' as of 2012, including biographical sources such as birth, marriage and death records, censuses, passenger lists. 'About Ancestry.Co.Uk', Ancestry, accessed 26 February 2015, <http://www.ancestry.co.uk/cs/legal/Overview>.

⁵¹ 'About the Ancestry.Com World Archives Project', Ancestry, accessed 26 December 2013, <http://landing.ancestry.com/wap/learnmore.aspx>.

⁵² 'Ancestry World Archives Project Profile Page', Facebook, accessed 11 February 2015, <https://www.facebook.com/AncestryWorldArchivesProject>.

⁵³ Examples are accessible from http://www.ancestry.com/wiki/index.php?title=Category:World_Archives_Project. Last accessed 6 June 2015.

⁵⁴ Ancestry.com's Tim Sullivan paraphrased in Kimberly Powell, 'World Archives Project vs. FamilySearch Indexing', *About.Com Genealogy*, 8 September 2008, <http://genealogy.about.com/b/2008/09/09/world-archives-project-vs-familysearch-indexing.htm>. In February 2015 active World Archives Projects include French electoral cards (1900-1932), New

clearly also provides a commercial advantage by supplementing *Ancestry's* records.

Grassroots indexing projects include the UK-based FreeUKGenealogy⁵⁵ group of projects, *FreeBMD* (transcribing the General Register Office indexes of Births, Marriages and Deaths for England and Wales), *FreeCen* (transcribing 19th Century census returns), and *FreeReg* (transcribing parish records). *FreeBMD* began in 1998 after the founders negotiated permission from the Office for National Statistics to publish the indexes to births, marriages and deaths for England and Wales online.⁵⁶ By 2014 over 12,000 volunteers had transcribed over 350 million records for *FreeBMD*.⁵⁷ Syndicates are led by volunteers who manage the process of acquiring and assigning documents to participants, who then transcribe records with Microsoft Excel or specialised software.⁵⁸ *Online Parish Clerks* (*OPC*) is another grassroots indexing project, made up of volunteers who 'adopt' parishes and 'collect, collate and transcribe' as many sources of historical data as they can find for those parishes.⁵⁹ *OPC* began in Cornwall in 2001.⁶⁰ Most Cornish parishes now have

York United Methodist Church Records (1791-1945), North Western Railroad records (1935-1970), Irish Famine Relief Commission Papers (1844-1847) and selected Holocaust Records from Czechoslovakia (1939-1945) for the United States Holocaust Memorial Museum.

⁵⁵ Previously known as FreeGEN.

⁵⁶ Camilla von Massenbach, 'RootsWeb: FreeBMD-Admins-L FreeBMD Project', *FreeBMD-Admins-L Archives*, 25 September 1998, <http://archiver.rootsweb.ancestry.com/th/read/FreeBMD-Admins/1998-09/0906727973>.

⁵⁷ 'About', *Free UK Genealogy*, accessed 11 February 2015, <http://www.freeukgenealogy.org.uk/about/>. Allan Raymond, 'RootsWeb: FreeBMD-Admins-L FreeBMD *** 300 Million *** Records Milestone Achieved', *FreeBMD-Admins-L Archives*, 25 February 2014, <http://archiver.rootsweb.ancestry.com/th/read/FreeBMD-Admins/2014-02/1393290478>.

⁵⁸ 'Transcribers' Knowledge Base', *FreeBMD*, 9 February 2014, http://www.freebmd.org.uk/vol_faqs.html#13j.

⁵⁹ FamilySearch Wiki, 'Wiltshire Online Parish Clerks', *FamilySearch.Org*, 2 January 2011, http://familysearch.org/learn/wiki/en/Wiltshire_Online_Parish_Clerks.

volunteer *OPCs*, and the model has been tried in other counties including Kent, Essex and Wiltshire, with varying levels of success. Many local history societies across the UK have organised their own record transcription projects on a smaller scale.⁶¹ In the US, the *USGenWeb* project was founded in 1996 to collect transcriptions of public domain records such as census records, marriage bonds and wills online,⁶² aiming to create a 'global library for genealogy research'.⁶³ As with the *OPC* sites, projects are organised at a local level rather than through a centralised repository and transcription interface.⁶⁴ Some states also have more focused projects, and other specialist projects, such as the *Tombstone Transcription Project* (launched in 1996); *Census Project* (1997); *Pension Project* (1999) and the *United States Digital Map Library* (1990), are organised at a national level.⁶⁵ In projects like *OPC* or *USGenWeb* where volunteers can 'adopt' or 'own' a parish or county,⁶⁶ volunteers may have a greater sense of ownership than in larger, centralised projects,⁶⁷ but it also leaves projects more exposed to the motivation levels and availability of individual

⁶⁰ Cornwall Online Parish Clerks, 'About Us', *Cornwall Online Parish Clerks*, accessed 14 May 2015, http://www.cornwall-opc.org/Structure/about_us.php.

⁶¹ This is discussed further in *Chapter 4*, but the following site is typical of many: Keynsham & Saltford Local History Society, 'Archives', *Keynsham & Saltford Local History Society*, accessed 13 February 2015, <http://keysalthist.org.uk/archives.htm>.

⁶² 'About the USGenWeb Archives', *USGenWeb Archives*, accessed 12 February 2015, <http://www.usgwarchives.net/projecthistory.htm>.

⁶³ Joy Fisher et al., 'Overview of the USGenWeb Project', 10 April 1996, <http://usgenweb.org/about/whitepaper.shtml>.

⁶⁴ Projects are listed by state then by county (or parish). 'About the USGenWeb', *USGenWeb Project*, accessed 12 February 2015, <http://www.usgenweb.org/about/index.shtml>.

⁶⁵ These projects can all be found via the *USGenWeb* sites.

⁶⁶ FLGenWeb, 'Adopt a County', *FLGenWeb*, accessed 26 February 2015, <http://www.flgenweb.org/volunteers/adopt>.

⁶⁷ As forum posts demonstrate. For example, one said of her local *OPC* project, 'you decide what information you put on it. You get to look at your page & say "I did that!"'. Forum poster, 'Re: Wiltshire Online Parish Clerks', *Rootschat*, 6 April 2012, <http://www.rootschat.com/forum/index.php?topic=303500.msg4416364#msg4416364>.

volunteers. Niche indexing projects include the *Crew List Index Project (CLIP)*, which aims 'to improve access to the records of British merchant seamen for the last part of the nineteenth century' by indexing records from local record offices.⁶⁸

The *Zooniverse* group is largely known for their science-focused crowdsourcing (or 'citizen science') projects, particularly image classification projects such as *Galaxy Zoo*. Their first indexing project was *Old Weather (OW)*, which provides data for climate scientists by digitising the historical weather information contained in ships logs. However, it quickly became a classic example of 'citizen history' as participants uncovered interesting details in the logs, noted them on the data entry form and discussed them on the project forum. Launched in October 2010, by January 2011 'the database had additional information from almost 10,000 log pages, representing 126 different ships'.⁶⁹ Subsequent *Zooniverse* indexing projects include *Operation War Diary* which promises to 'create new "Citizen Historians"' through indexing military unit diaries from the First World War.

Herbaria@Home (launched August 2006)⁷¹ is another example of a non-genealogical

⁶⁸ 'What next for CLIP?', *CLIP - Crew List Index Project*, August 2013, <http://www.crewlist.org.uk/aboutclip.html>.

⁶⁹ Fiona Romeo and Lucinda Blaser, 'Bringing Citizen Scientists and Historians Together', in *Museums and the Web 2011: Proceedings*, ed. Jennifer Trant and David Bearman (Museums and the Web 2011, Toronto, Canada: Archives & Museum Informatics, 2011), http://www.museumsandtheweb.com/mw2011/papers/bringing_citizen_scientists_and_historians_tog.

⁷⁰ This claim is examined in more detail in a later chapter. *Operation War Diary*, 'About Operation War Diary', *Operation War Diary*, 2014, <http://www.operationwardiary.org/#/about>.

⁷¹ Tom Humphrey, 'About Herbaria@home', *Herbaria@home*, accessed 29 March 2014, <http://herbariaunited.org/ahcontent/AHabout/>. Its forum launched slightly earlier, in June 2006. Tom Humphrey, 'Welcome!', *Herbariaunited.Org Forum Index -> News and Announcements*, 4 June 2006, <http://herbariaunited.org/core/forum/viewtopic.php?t=6>.

transcription project. The site aims to catalogue and transcribe historical herbarium specimen sheets from collections held by universities and museums.⁷² The project was inspired by an in-person volunteer digitisation project, and was conceived of as an online distributed project to overcome 'the problem of giving large numbers of volunteers physical access to a museum's collection, limited numbers of computers and limited space'.⁷³ By May 2015, over 400 participants had transcribed over 150,000 sheets.⁷⁴

Several projects aim to index other types of documents, including menus, theatre programmes, maps and musical scores. The New York Public Library's *What's On The Menu?* was launched in 2011 to create a searchable database from images of the Buttolph Menu Collection.⁷⁵ *Cymru1900Wales* asked participants to transcribe place names found on Ordnance Survey maps. The New York Public Library's *Ensemble* project (theatre programmes) and the Bodleian Library's *What's the Score* (musical scores) are both built on *Zooniverse* software that asks the participant to classify the text while they transcribe it.⁷⁶

Full-text transcription

Full-text transcription, unsurprisingly, involves typing in or computationally recording all the text on a given document.⁷⁷ Typical source documents include letters, diaries, journals,

⁷² Humphrey, 'About Herbaria@home'.

⁷³ Humphrey, 'About Herbaria@home'.

⁷⁴ The latest stats are available at <http://herbariaunited.org/atHomeUserList/>. Last accessed 6 June 2015.

⁷⁵ Michael Lascarides and Ben Vershbow, 'What's on the Menu?: Crowdsourcing at the New York Public Library', in *Crowdsourcing Our Cultural Heritage*, ed. Mia Ridge (Farnham, Surrey, UK: Ashgate, 2014), <http://www.ashgate.com/isbn/9781472410221>.

⁷⁶ For example, classifying the text 'Words by Glen MacDonough' as being about a person and assigning the text to the role of playwright.

⁷⁷ Both partial and full transcription can be created from tabular or narrative text sources, but

recipes, administrative documents, newspapers, and fiction and non-fiction books. Sources are more likely to be hand-written, as optical character recognition (OCR) technologies are increasingly able to cope with typed text.⁷⁸ In some projects, one volunteer may be able to transcribe an entire document page-by-page; in others the pages may be split and assigned to different volunteers. Full-text transcriptions can be structured or unstructured. Unstructured text transcriptions contain all the words that appear on a document, but other than whitespace (for example, paragraph breaks) and punctuation marks, no information about the appearance or meaning of the text is recorded. Structured full-text transcriptions may use an encoding language such as the Text Encoding Initiative (TEI)⁷⁹ expressed in Extensible Markup Language (XML) to wrap tags around text; the tags can describe the meaning or appearance of the enclosed text. For example, this line of text from the Letters of 1916 project has been encoded as a date, and two characters have been further encoded as superscript: <date>19<hi rend="superscript">th</hi> March, 1916.</date>.⁸⁰ Other platforms use simple wiki-style markup to create links and generate an index of subjects, people and places mentioned within the text.⁸¹ Structured data requires more resources to create and manage, and is a slightly more complex task, but it enables a wider range of future uses.⁸² Many of the examples discussed in this section are

narrative texts are more likely to be fully transcribed.

⁷⁸ With the caveat that many OCR technologies have high error rates, as the existence of OCR correction projects shows, with error rates partly linked to typeface and paper quality. Handwritten Text Recognition (HTR) technology is also rapidly advancing through projects such as *tranScriptorium*, which aims to produce 'cost-effective solutions for the indexing, search and full transcription of historical handwritten document images'. *tranScriptorium*, 'Objectives', *TranScriptorium*, accessed 15 May 2015, <http://transcriptorium.eu/pagina/objectives/>.

⁷⁹ TEI Consortium, 'TEI: FAQ', *Text Encoding Initiative Consortium*, 29 June 2014, <http://www.tei-c.org/About/faq.xml>.

⁸⁰ Taken from <http://dh.tcd.ie/letters1916/diyhistory/scripto/transcribe/231/579> in May 2015.

⁸¹ For example, *Scripto* and *FromThePage* use wiki markup.

⁸² As discussed by those working on *Old Bailey Online*, such as Tim Hitchcock and Robert

both platforms and projects, as software developed for one particular project may be generalised and made available for others to use.⁸³

Project Gutenberg began in 1971⁸⁴ and is possibly the earliest full-text transcription project. *Gutenberg* publishes plain text versions of previously published, out of copyright texts, prioritising those they think the general public will want over esoteric texts or scholarly editions.⁸⁵ While it began as a manual transcription project, some contributors now scan pages and run OCR software over them rather than hand-transcribe an entire book.⁸⁶ *WikiSource* began in 2003 as a collection of supporting texts for use in *Wikipedia*.⁸⁷ *WikiSource* also hosts some historical documents 'of national or international interest'⁸⁸ for transcription, and can be used by projects looking for a transcription platform.

Shoemaker, 'Digitising History From Below: The Old Bailey Proceedings Online, 1674-1834: The Old Bailey Proceedings Online, 1674-1834', *History Compass* 4, no. 2 (March 2006): 193-202, doi:10.1111/j.1478-0542.2006.00309.x. and Sharon Howard, 'Bloody Code: Reflecting on Ten Years of the Old Bailey Online and the Digital Futures of Our Criminal Past', *Early Modern Notes*, 18 May 2013, <http://earlymodernnotes.wordpress.com/2013/05/18/bloody-code-reflecting-on-ten-years-of-the-old-bailey-online-and-the-digital-futures-of-our-criminal-past/>.

⁸³ For a discussion of this process, see Sharon M. Leon, 'Build, Analyse and Generalise: Community Transcription of the Papers of the War Department and the Development of Scripto', in *Crowdsourcing Our Cultural Heritage*, ed. Mia Ridge (Farnham, Surrey, UK: Ashgate, 2014), <http://www.ashgate.com/isbn/9781472410221>. pp. 89-111.

⁸⁴ Michael Hart, 'The History and Philosophy of Project Gutenberg', *Project Gutenberg*, August 1992, http://www.gutenberg.org/wiki/Gutenberg:The_History_and_Philosophy_of_Project_Gutenberg_by_Michael_Hart.

⁸⁵ Hart, 'The History and Philosophy of Project Gutenberg'.

⁸⁶ 'Volunteers' FAQ', *Gutenberg*, 7 December 2014, https://www.gutenberg.org/wiki/Gutenberg:Volunteers%27_FAQ.

⁸⁷ Originally it was modelled on *Project Gutenberg*. Wikisource, 'What Is Wikisource?', *Wikisource*, 10 July 2014, https://en.wikisource.org/wiki/Wikisource:What_is_Wikisource%3F.

⁸⁸ Wikisource, 'What Is Wikisource?'

More recently, specialist manuscript transcription platforms have emerged. *FromThePage*, a wiki-like application for crowdsourcing the transcription of handwritten documents, launched as a personal project in 2008⁸⁹ and was subsequently adopted by other projects.⁹⁰ In 2010, University College London (UCL) launched the *Bentham Papers Transcription Initiative*, generally known as *Transcribe Bentham*,⁹¹ to see whether they could interest the public in the 'somewhat recondite practice of manuscript transcription'⁹² while transcribing the manuscript papers of the philosopher Jeremy Bentham (1748-1832).⁹³ In addition to transcribing Bentham's difficult handwriting, the project also asks participants to mark-up attributes of the text including additions, deletions and marginal notes.⁹⁴ To date, over 450 participants⁹⁵ have transcribed over 13,000 manuscripts.⁹⁶

Another platform, *Scripto*, was launched by the Roy Rosenzweig Center for History and

⁸⁹ Ben W. Brumfield, 'Itinera Nova in the World(s) of Crowdsourcing and TEI', *Collaborative Manuscript Transcription*, 29 April 2013, <http://manuscripttranscription.blogspot.co.uk/2013/04/itinera-nova-in-worlds-of-crowdsourcing.html>.

⁹⁰ Ben W. Brumfield, 'Progress Report: GitHub, Archive.Org Integration, and General Availability', *Collaborative Manuscript Transcription*, 4 January 2011, <http://manuscripttranscription.blogspot.co.uk/2011/01/progress-report-github-archiveorg.html>.

⁹¹ Tim Causer and Melissa Terras, 'Crowdsourcing Bentham: Beyond the Traditional Boundaries of Academic History', *International Journal of Humanities and Arts Computing* 8, no. 1 (2014): 46-64, <http://www.eupublishing.com/doi/abs/10.3366/ijhac.2014.0119>.

⁹² Martin Moyle, Justin Tonra, and Valerie Wallace, 'Manuscript Transcription by Crowdsourcing: Transcribe Bentham', *Liber Quarterly* 20, no. 3-4 (2011): 347-356, <http://liber.library.uu.nl/index.php/lq/article/view/7999>.

⁹³ 'About Us', accessed 12 January 2012, <http://www.ucl.ac.uk/transcribe-bentham/about/>.

⁹⁴ 'Getting Started', *Transcribe Bentham: Transcription Desk*, 16 July 2013, http://www.transcribe-bentham.da.ulcc.ac.uk/td/Getting_Started.

⁹⁵ As of July 2014. 'Hall of Fame', *UCL Transcribe Bentham*, 22 July 2014, <http://blogs.ucl.ac.uk/transcribe-bentham/hall-of-fame/>.

⁹⁶ As of May 2015. Tim Causer, 'Progress Update, 9 to 15 May 2015', *UCL Transcribe Bentham*, 15 May 2015, <https://blogs.ucl.ac.uk/transcribe-bentham/2015/05/15/progress-update-9-to-15-may-2015/>.

New Media (CHNM) in March 2011,⁹⁷ and was first used in their own *Papers of the War Department* transcription project.⁹⁸ *Scripto* can be integrated with *MediaWiki* (the software platform that runs *Wikipedia*) or CHNM's *Omeka*, a content management system designed for collections.⁹⁹ It has subsequently been used by projects including the *UrCrowdsource* project, which aims to transcribe archaeological documentation, and *Texas Manuscript Cultures*. The *Scripto* software was extended with a custom theme and transcription plugin¹⁰⁰ by the University of Iowa (UI) Libraries' *DIY History* project, which launched their transcription platform with the Szathmary Culinary Manuscripts and Cookbooks collection in October 2012,¹⁰¹ and subsequently used it on several large collections, receiving their 50,000th transcription in November 2014.¹⁰² Their *DIYHistory|transcribe* software package is now used by a number of other museums and libraries,¹⁰³ including the *Letters of 1916* project which launched in September 2013.¹⁰⁴ *Letters of 1916* aims to collect and transcribe letters related to the Irish Easter Rising held in private and institutional collections.¹⁰⁵ By April 2015 92 letters had been fully proofed,

⁹⁷ Sharon Leon, 'Scripto Alpha Launches with the Papers of the War Department', *Scripto*, 17 March 2011, <http://scripto.org/scripto-alpha-launches-with-the-papers-of-the-war-department/>.

⁹⁸ Leon, 'Build, Analyse and Generalise'. p. 89

⁹⁹ At the time of writing it can also be integrated with *WordPress* and *Drupal*.

¹⁰⁰ Themes and plugins are software components that can change the appearance and functionality of a platform.

¹⁰¹ Jen Wolfe, 'UI Libraries Launches New Crowdsourcing Site with Manuscript Cookbooks and More', *The University of Iowa Libraries*, 9 October 2012, <http://blog.lib.uiowa.edu/drp/2012/10/09/diyhistory/>.

¹⁰² Jen Wolfe, 'DIY History Celebrates 50,000th Transcription!', *The University of Iowa Libraries*, 11 November 2014, <http://blog.lib.uiowa.edu/drp/2014/11/11/diyh50k/>.

¹⁰³ Jen Wolfe, 'Contributing in Code', *The University of Iowa Libraries*, 18 November 2014, <http://blog.lib.uiowa.edu/drp/2014/11/18/contributing-in-code/>.

¹⁰⁴ Their platform also uses the *Transcribe Bentham* TEI Toolbar to support the addition of TEI markup.

¹⁰⁵ Karolina Badzmierowska, 'Letters of 1916 Celebrates Its First Birthday!', 29 September 2014,

with a further 1586 transcribed.¹⁰⁶

The *Smithsonian Transcription Center* was available as a 'beta' (pre-official release) version from July 2013 and formally launched in July 2014.¹⁰⁷ By May 2015, over 90,000 pages had been transcribed and reviewed.¹⁰⁸ As the project includes hugely varied material from across the organisation, the platform has customisable templates and provides tutorials specific to the relevant source to ensure the best results for each type of record, whether indexing or full-text transcription. The project uses social media rather than on-site forums for community discussion, and actively encourages discussion about participant discoveries and questions.¹⁰⁹

Other tools include the *Transcribr* distribution of Drupal (an open source content management system) used for the US National Archives and Records Administration (NARA) platform when it launched in January 2012.¹¹⁰ The Transcription for Paleographical

<http://dh.tcd.ie/letters1916/letters-1916-celebrates-first-birthday/>.

¹⁰⁶ Letters of 1916, 'Progress Update, April 2015', *Letters 1916*, April 2015, <http://dh.tcd.ie/letters1916/progress-update/april-2015/>.

¹⁰⁷ SI Transcription Ctr, '@mia_out June 2013 Beta Then July 2014 (Though Always Iterating/Improving) #thanksforasking', microblog, @TranscribeSI, (27 February 2015), <https://twitter.com/TranscribeSI/status/571133725406302208>.

¹⁰⁸ SI Transcription Ctr, '@mia_out Great Q! We Update by e-Mail & Blogpost (& New Section, Shhh) - Today's Numbers: 722 Projects Comp and/or 91,135 Pages Completed!', microblog, @TranscribeSI, (15 May 2015), <https://twitter.com/TranscribeSI/status/599234883030921216>.

¹⁰⁹ For example, once a month the site moderator asks participants to share discoveries on the social media sites *Tumblr*, *Facebook* or *Twitter* by posting with the hashtag #MyTCdiscovery. Smithsonian Transcription Center, '#MyTCdiscovery', *Facebook*, 27 April 2015, <https://www.facebook.com/SmithsonianTranscriptionCenter/photos/a.341352889336961.1073741829.339997839472466/517754628363452/?type=1>.

¹¹⁰ 'Transcribr', *Drupal.Org*, 31 July 2012, https://www.drupal.org/project/transcribe_distribution. The Drupal-based site was closed within a few years, and the transcription function was later

and Editorial Notation (*T-PEN*) project is designed to work with documents uploaded to their platform or hosted in linked repositories.¹¹¹ The University of Alabama Libraries developed a system called *Acumen*, a 'digital archives explorer' that (as of 2015) offers transcription and tagging functions.¹¹² Some projects are designed to work with non-textual sources. NARA uses a tool called *Amara* to create subtitles from speech in videos.¹¹³ Some image annotation tools can also be used for text transcription. For example, *Recogito* was developed to transcribe place references in early geospatial texts and maps.¹¹⁴

The *Marine Lives* project, launched in 2012,¹¹⁵ aims to create a 'fully searchable semantic web based corpus of over 20 million words by 2017' from manuscripts of the High Court of Admiralty, London, 1650-1669.¹¹⁶ Participants work in small groups lead by academic historians volunteering as facilitators for phases of 3-4 months.¹¹⁷ While many cultural

provided within the archive catalogue interface.

¹¹¹ James Ginther, 'T-PEN Version 2.0', *The Center for Digital Theology at Saint Louis University*, 5 May 2012, <http://digital-editor.blogspot.co.uk/2012/05/t-pen-version-20.html>.

¹¹² Kate Matheny, 'Come Transcribe Our Items!', *Digital Services @ the University of Alabama*, 9 February 2015, <http://apps.lib.ua.edu/blogs/digitalservices/2015/02/09/transcribe-our-items/>. They had previously used *Scripto*.

¹¹³ Mary, 'Calling Citizen Archivists to Crowdsourcing Video Captions!', *NARAtions*, 26 March 2014, <http://blogs.archives.gov/online-public-access/?p=9222>.

¹¹⁴ 'Beginner's Tutorial', *Recogito*, accessed 15 May 2015, <http://pelagios.org/recogito/static/documentation/index.html>.

¹¹⁵ Colin Greenstreet, 'MarineLives - Call for Project Volunteers', *Academia.Edu*, 2012, http://marinelives.academia.edu/ColinGreenstreet/Posts/351069/MarineLives_-_Call_for_Project_Volunteers.

¹¹⁶ Jill Wilcox, 'About MarineLives', *MarineLives Project Manual*, 14 May 2014, http://marinelives.wikispot.org/About_MarineLives.

¹¹⁷ The length of the phases was apparently modelled on university terms, in support of the project's training goals. Unlike most crowdsourcing projects, this also means each phase of participation has a known end date. Discussion at Various, 'AHRC Crowd Sourcing Study: Scoping Seminar' (King's College London, 18 October 2012), <http://crowds.cerch.kcl.ac.uk/>.

heritage and academic sector projects have some public engagement goals,¹¹⁸ *Marine Lives* goes further, aiming to provide 'extramural research training' in 17th century history for people not in a position to undertake postgraduate study, and to help historians learn to work collaboratively on historical research.¹¹⁹

By breaking the transcription task into single letter or word-sized tasks some projects even make it possible to transcribe texts in languages the participant might not be familiar with. *Ancient Lives*, another *Zooniverse* project, aims to transcribe fragments of ancient Greek papyri,¹²⁰ and *DigitalKoot* was a crowdsourcing game that allowed anyone to transcribe or correct transcriptions in Finnish newspapers.¹²¹

Some projects eschew specialist platforms and turn to social media platforms, such as the image-hosting site *Flickr*. Various institutions, including the St Fagans National History Museum in Wales, London's Horniman Museum and the State Library of North Carolina have experimented with asking the public to help transcribe documents posted on *Flickr* by leaving comments with transcribed text on the image page.¹²² However, as posters cannot edit earlier comments, they cannot easily iteratively or collaboratively improve on transcriptions. It is also not clear how the institutions will manage the process of compiling and reconciling different versions of the text, or subsequently import this text into their collections or digital asset management systems. This use of *Flickr* highlights the

¹¹⁸ For example, *DIY History*: Wolfe, 'UI Libraries Launches New Crowdsourcing Site with Manuscript Cookbooks and More'.

¹¹⁹ Discussion at Various, 'AHRC Crowd Sourcing Study: Scoping Seminar'.

¹²⁰ In some ways, *Ancient Lives* works as a pattern-matching exercise where characters from the provided alphabet can be matched to characters on the papyrus.

¹²¹ 'Digitalkoot: Crowdsourcing Finnish Cultural Heritage', accessed 9 January 2012, <http://blog.microtask.com/2011/02/digitalkoot-crowdsourcing-finnish-cultural-heritage/>.

¹²² Links are in *Appendix A*, as for all sites listed.

benefits of specialist manuscript transcription platforms. Specialist software usually provides functionality for managing iterative transcriptions and moving records between 'transcription', 'review' and 'completed' stages. Some software also allows administrators to give trusted volunteer accounts the ability to moderate or approve submitted content, reducing the institutional workload and potentially mitigating against delays in approving records.

Proofread and corrected text

Text correction is possibly the smallest, most discrete task in participatory history, as it allows a user to make a meaningful contribution by correcting a single character in a single word. Text in need of correction has often been generated by optical character recognition (OCR) software, which can have accuracy rates as low as 70% on early newspapers (meaning 30% of the characters would be incorrect).¹²³ Text correction can be closely connected to the process of reading text for other purposes, but, as the examples discussed below show, some people choose to correct text for the intrinsic value or pleasure of the task. Some projects, such as the Bibliothèque nationale de France's *Gallica Correct*, are specifically designed for proofreading or correcting text, while some proofreading and correction is also undertaken by employees or volunteers as they review contributions to transcription projects.

Distributed Proofreaders was founded in 2000 to assist *Project Gutenberg* by proofreading OCRed text from scanned texts; as of April 2015 participants have fully processed over 30,000 books (including formatting checks and conversion to e-book formats).¹²⁴ The project also offers a simpler task called 'smooth reading', which involves reading books

¹²³ Holley, 'How Good Can It Get?'

¹²⁴ Distributed Proofreaders, 'DP: Welcome', *Distributed Proofreaders*, accessed 27 July 2014, <http://www.pgdp.net/c/>.

that are almost ready for posting to *Project Gutenberg* 'attentively, as for pleasure, with just a little more attention than usual to punctuation'.¹²⁵

Trove, the National Library of Australia's database of online resources, is a 'discovery service' for Australian content. It harvests metadata from over a thousand Australian libraries and other organisations,¹²⁶ and is particularly known for its freely available digitised newspapers. *Trove* originally began showing uncorrected (or 'raw') OCR to readers and providing the option to correct errors as it was cheaper than hiring expensive contractors.¹²⁷ The basic prototype was released to libraries in December 2007,¹²⁸ and the Australian newspapers site was formally launched in August 2008.¹²⁹ By June 2015, *Trove* users had contributed over 3.5 million tags, almost 93,000 comments, and over 163 million newspaper corrections.¹³⁰ *Trove* supports two key groups of users: those who intrinsically enjoy the text correcting task, and those who correct text, tag or add articles to lists etc. while using the newspapers for their research; both groups contribute towards these figures.

Other newspaper correction examples include Veridian's *Crowdsourced User Text Correction* module¹³¹ as used by the *California Digital Newspaper Collection*. In May 2015,

¹²⁵ Distributed Proofreaders, 'Smooth Reading Pool Preview', *Distributed Proofreaders*, 13 May 2014, http://www.pgdp.net/c/tools/post_proofers/smooth_reading.php.

¹²⁶ Rose Holley, 'Resource Sharing in Australia: Find and Get in Trove – Making "Getting" Better', *D-Lib Magazine* 17, no. 3/4 (2011), doi:10.1045/march2011-holley.

¹²⁷ Holley, 'How Good Can It Get?'

¹²⁸ Holley, 'How Good Can It Get?'

¹²⁹ Holley, 'Crowdsourcing: How and Why Should Libraries Do It?'

¹³⁰ Overall contributor numbers cannot be supplied as registration is optional; in May 2015 over 5000 registered users had been active and there were 158,000 registered users overall. Figures obtained from <http://trove.nla.gov.au/system/stats?env=prod>, last accessed 11 June 2015.

¹³¹ Veridian, 'Crowdsourced User Text Correction (UTC)', *Veridian*, accessed 21 July 2014,

over 1,800 users had corrected over 4,100,000 lines of text.¹³² The platform used for the *British Newspaper Archive* also has OCR corrections functions.¹³³ The *18thConnect* site allows users to correct documents from the digital collections Early English Books Online (*EEBO*) and Eighteenth Century Collections Online (*ECCO*) with a tool called *TypeWright*.¹³⁴

The *Dickens Journals Online* project is an example of a focused approach to OCR correction. The project aimed to present a 'complete online edition of Dickens's weekly magazines, *Household Words* and *All the Year Round*'¹³⁵ (about 1100 items) and launched the *Online Text Correction* project to clean up OCR from scanned pages by the deadline of Dickens's bicentenary.¹³⁶ Despite the relatively large 'units of work' - magazines of c. 24 pages, each estimated to take 'about 10 minutes to review and correct each page' for a total of '240 minutes or 4 hours' work¹³⁷ per task - over 3,200 registered users contributed to

<http://www.veridiansoftware.com/knowledge-base/crowdsourced-user-text-correction/>.

¹³² Figures obtained from <http://cdnc.ucr.edu/cgi-bin/cdnc?a=p&p=textcorrectstats&e=-----en--20-1--txt-txIN-----> in May 2015, page last accessed 11 June 2015. The page also states that over 3300 users have registered; suggesting that over 1500 users have registered but have not yet corrected any text. The reasons for this gap are not clear.

¹³³ 'FAQ', *British Newspaper Archive*, accessed 1 March 2015, <http://www.britishnewspaperarchive.co.uk/help-faq/what-should-i-do-if-i-notice-a-mistake-in-the-ocr-text#1>.

¹³⁴ 'What Is 18thConnect?', *18thConnect*, accessed 4 September 2014, <http://www.18thconnect.org/about/>.

¹³⁵ John Drew, 'About Us', *Dickens Journals Online*, accessed 14 February 2015, <http://www.djo.org.uk/about-us.html>.

¹³⁶ John Drew, 'The Online Text Correction (OTC) Project', *Dickens Journals Online*, accessed 14 February 2015, <http://web.archive.org/web/20130426050435/http://www.djo.org.uk/about-us/community-projects/online-text-correction.html>.

¹³⁷ Drew, 'The Online Text Correction (OTC) Project'.

complete the task by the deadline.¹³⁸

Collections

Collecting projects aim to gather new items, whether physical or digitised materials.

Collecting projects might be focused on specific types of objects or documents, events, places or cultural groups. Projects focused on specific places include *Know Your Place*, which invited people to contribute stories, images, films and objects to a 'community layer' on the map of Bristol.¹³⁹ *My Brighton and Hove* collects old and new photos, memories and information, and in 2007-8 organised a special appeal to the public to contribute personal letters and diaries related to Brighton and Hove.¹⁴⁰ Projects like *WhatWasThere* collect photographs linked to specific locations. *HistoryPin* hosts 'photographs, documents, sounds and moving images' from members of the public and from GLAMs,¹⁴¹ having expanded its original focus on the act of uploading and 'pinning' photos to a map.¹⁴² Users of mobile applications designed to 'digitise' historic images can contribute records to family history sites.¹⁴³

Collecting projects focused on specific topics include *Wir Waren so Frei*, which collected

¹³⁸ John Drew, 'Thanks to You All for a Dickens of a Job Done', *The Guardian*, 21 June 2012, sec. Letters, <http://www.theguardian.com/books/2012/jun/21/thanks-you-dickens-job-done>.

¹³⁹ Now *Know Your Bristol*. 'About', *Know Your Bristol*, accessed 1 March 2015, <http://knowyourbristol.org/about-us/>.

¹⁴⁰ Jack Latimer, 'Letter in the Attic: Lessons Learnt from the Project', *My Brighton and Hove*, 25 February 2009, <http://www.mybrightonandhove.org.uk/page/letterintheatticlessons?path=op116p1543p>.

¹⁴¹ Galleries, libraries, archives and museums.

¹⁴² 'Historypin', *Shift*, accessed 1 March 2015, <http://www.shiftdesign.org.uk/products/historypin/>.

¹⁴³ Rudy Adler, 'Ancestry.Com Acquires 1000memories', *1000memories*, 3 October 2012, <http://blog.1000memories.com/170-ancestry-acquires-1000memories>.

private pictures from the German public that represent the changes of 1989-90,¹⁴⁴ the Victoria and Albert Museum's *Wedding Fashion* project, which collected photos of 'clothes worn for weddings from all cultures between 1840 and the present',¹⁴⁵ the British Library's *Sound Map*,¹⁴⁶ and the Irish *Letters of 1916*.¹⁴⁷ *Europeana 1914-1918* is both a repository of 'official histories' and a collecting project for 'memories and memorabilia'.¹⁴⁸ The RunCoCo group have run a range of community collection projects, including the *Great War Archive* (2008) and *Project Woruldhord*, collecting material on Anglo-Saxon England from 'academics, museums, and members of the public'.¹⁴⁹ *Wiki Loves Monuments* and *Open Plaques* respectively ask people to contribute images of monuments and commemorative 'blue plaques' taken in public places. The Guardian newspapers' *GuardianWitness* project collects user-generated content around specific 'assignments', which sometimes include historical topics such as the First World War.¹⁵⁰ While material in projects based on GLAM and academic collections is generally regarded as having good provenance, projects collecting material from the public may encounter questions about the authenticity or accuracy of the content provided, or the intentions of its contributor. My interviews with historians (discussed in *Chapter 4*) showed that transcriptions without images of the

¹⁴⁴ 'Project Description', *Wir Waren so Frei*, accessed 15 February 2015, <https://www.wir-waren-so-frei.de/index.php/About/Index>.

¹⁴⁵ 'Wedding Fashion', *Victoria and Albert Museum*, 2011, <http://web.archive.org/web/20110818221004/http://www.vam.ac.uk/things-to-do/wedding-fashion/home>.

¹⁴⁶ Maureen Pennock and Chris Clark, 'Saving the Sounds of the UK in the UK SoundMap', *Ariadne*, no. 66 (30 January 2011), <http://www.ariadne.ac.uk/issue66/pennock-clark>.

¹⁴⁷ 'Kildare Launch of the "Letters of 1916"', *Maynooth University*, 28 April 2014, <https://www.maynoothuniversity.ie/foras-feasa/news/kildare-launch-letters-1916>.

¹⁴⁸ *Europeana 1989* follows a similar model.

¹⁴⁹ Stuart Lee, 'Woruldhord Web Site Now Launched!', *Project Woruldhord*, 14 March 2011, <http://blogs.it.ox.ac.uk/woruldhord/2011/03/14/woruldhord-web-site-now-launched/>.

¹⁵⁰ 'GuardianWitness', accessed 2 March 2015, <https://witness.theguardian.com/moreabout>.

original document will not be trusted, but they seem to take it on trust that digitised images accurately depict documents that exist in the real world.

Some platforms developed for specific projects have been repurposed and made available to other projects. The Community Sites software underlying *My Brighton and Hove* has been used for other sites, including *Herts Memories* (Hertfordshire) and a community history of St Hellier Estate, London. In the US, *History Harvest* has undergraduates work with local communities to collect artefacts and stories, which are then published in *Omeka*. In New Zealand, *Kete Horowhenua* aims to release their *Kete* code to help other local communities host contemporary and historical content.¹⁵¹ Some projects aimed at collecting historic photographs have used *Flickr* rather than build their own systems, including *The Great War Archive*,¹⁵² Royal Museums Greenwich's *Beside the Seaside* and *Picture Australia*. As *Flickr* is optimised for uploading and displaying photos, this is a better match between platform, material and purpose than projects attempting to use *Flickr* to transcribe handwritten documents, as discussed previously.

While most collecting projects seek material held in family or community collections, gravestone sites provide an interesting example of the 'collecting' dynamic when applied to objects in the world. Gravestones can be an important source of biographical information not available elsewhere, and the range of different approaches to the same core output (locations, images and transcriptions of gravestones) is an example of the variety of approaches in the wider field of participatory history projects. The Irish *Historic Graves*

¹⁵¹ Joann Ransom, 'Kete Horowhenua: The Story of the District as Told by Its People', accessed 15 June 2014, http://www.valaconf.org.au/vala2008/papers2008/31_Ransom_Final.pdf.

¹⁵² The *Great War Archive* used a Flickr group to collect images after their project funding had ended. 'The Great War Archive Flickr Group Rules', *Flickr*, accessed 15 February 2015, <https://www.flickr.com/groups/greatwararchive/rules/>.

project trains local community groups to digitally record field surveys of historic graveyards.¹⁵³ The *Gravestone Photographic Resource* began in 1998 as a personal project, in part to record gravestones that were being lost to weather and vandalism.¹⁵⁴ *Billion Graves*, launched in 2011, focuses on mobile applications designed for recording the locations of photos taken in cemeteries, which are then uploaded for transcription by any registered participant.¹⁵⁵ *Find A Grave* began in 1995 because the creator 'could not find an existing site that catered to his hobby of visiting the graves of famous people'.¹⁵⁶ While most gravestone sites concentrate on collecting as many transcribed records as possible, *Find A Grave* participants often focus on collecting ownership of the 'memorials' associated with the person recorded by a gravestone. Memorials 'owned' by one contributor cannot be edited by others, and some contributors are reportedly reluctant to hand memorial pages over to close relatives of the deceased; the urge to 'collect' and share has been transferred to collecting and controlling as many memorials as possible.¹⁵⁷ The same problematic 'collecting and controlling' dynamic can be seen on other sites where the community and/or interface has focused on 'owning' the record for a specific item rather than collecting new items or collaboratively improving existing records.¹⁵⁸ Gravestone sites can

¹⁵³ 'About Us', *Historic Graves*, accessed 2 March 2015, <http://historicgraves.com/content/about-us>.

¹⁵⁴ 'About the Gravestone Photographic Resource Project Grave Directory', *Gravestone Photographic Resource*, 2015, <http://www.gravestonephotos.com/information/about.php?r=1&width=1920&browser=1110>.

¹⁵⁵ 'Launching BillionGraves.Com', *BillionGraves Blog*, 25 May 2011, <http://blog.billiongraves.com/2011/05/launching-billiongraves-com/>.

¹⁵⁶ 'Who Is behind Find A Grave?', *Find A Grave*, accessed 14 February 2015, <http://www.findagrave.com/cgi-bin/fg.cgi?page=whois>.

¹⁵⁷ See the comments discussing 'hoarders' or 'squatters' on Kerry Scott and Various commenters, 'Why Is There So Much Drama On Find A Grave?', *Clue Wagon*, 13 April 2011, <http://www.cluewagon.com/2011/04/why-is-there-so-much-drama-on-find-a-grave/>. Some participants 'manage' hundreds of thousands of memorials: <http://findagrave.com/cgi-bin/fg.cgi?page=mp>. Last accessed 6 June 2015.

¹⁵⁸ Sites like *WikiTree* try to prevent this dynamic emerging by emphasising collaboration over

also be the focus of tensions between those looking to 'memorialise' the dead and those who regard gravestones as an artefact with text useful to genealogical research.¹⁵⁹

Institutional relationships with collecting projects vary, and collecting projects with other participatory functions can further blur the boundaries between project types. For example, *Letters of 1916* is both a collecting and a transcription project. Projects that collect stories in addition to physical or digital objects, such as the *Bracero History Archive*, could also be considered oral history projects. Collecting projects can be seen as a form of audience engagement by GLAMs rather than an attempt to record cultural history,¹⁶⁰ and as such, the items collected may be regarded as disposable content rather than accessionable collections material. For example, the V&A Museum's *Wedding Fashion* project has disappeared from their website while the *World Beach Project*, another collecting project which was regarded as an artwork rather than an exercise in audience engagement, remains. There are many potential explanations for this, but it is indicative of a broader pattern that sees 'user-generated content' as ephemeral. This can be problematic if the public expect an institution to preserve the material they have contributed. The commercial sector has already failed to protect contributed content through various company failures, mergers and acquisitions.¹⁶¹

control. 'Ownership and Control', *WikiTree*, 18 August 2014, http://www.wikitree.com/wiki/Ownership_and_control.

¹⁵⁹ For example, the owner of one cemetery site was forced to explain, 'I realize that if you saw your mother's burial info on this website, and noticed incorrect information, it becomes an emotional issue [...but...] Interment.net was not meant to memorialize the deceased, rather it's meant to be a genealogical reference'. Steve Johnson, 'Correcting Errors in Transcriptions', *Interment.Net*, 17 April 2009, <http://www.interment.net/column/2009/04/correcting-errors-in-transcriptions.html>.

¹⁶⁰ The difference between 'Web 2.0'-style user-generated content projects and purposive crowdsourcing is discussed briefly in the next chapter.

¹⁶¹ See for example this account of the fate of family histories stored on Ancestry's MyFamily.com: Jon Christian, 'Deleting the Family Tree', *Slate*, 23 April 2015,

Information

Information gathered through participatory projects has many forms and sources, from short impressionistic text tags to deeply researched specialist knowledge, from strongly evidenced factual statements to personal histories and observations. Information can be collected through tags, comments, image annotations, text markup, interviews, and individually- or collaboratively-authored articles. The same piece of information - for example, the identification of people, places or events in an image - might come from research that people specifically undertook to complete the task, from their disciplinary knowledge, or as the result of personal knowledge (for example, in 'living memory' or community history projects). Each context may raise different questions about verifying and crediting information, and about the role of institutional or community authority. This section discusses information types including: metadata, georeferencing, personal and experiential information and expert information; information task formats such as annotations and 'history mysteries'; and emergent forms of research tasks including distributed reading, research commissions, and research quests.

Online communities of historians have developed their own patterns of interaction around information exchange, including 'look-up requests', which are requests for another researcher to look up information in a specific archive or resource. Public requests and offers are particularly common on genealogy sites; one grassroots project, *Random Acts of Genealogical Kindness*, was created to coordinate the impulse to help strangers with their research demonstrated on many genealogy sites. Sites like *WeRelate*, *WikiTree* and *OneGreatFamily* aim to create collaborative, global family trees that could be seen as a form of proto-prosopography. Arguably, scholarly sites like *H-Net Commons* whose goal is

http://www.slate.com/articles/technology/future_tense/2015/04/myfamily_shuttered_ancestry_com_deleted_10_years_of_my_family_history.html.

to be a 'public square where multiple disciplines meet and share information'¹⁶² also promote the exchange of information by hosting discussion lists. The exchanges are usually scholar-to-scholar, but in hosting and archiving these discussions in public, they are creating a resource that is discoverable by others. Spaces where work-in-progress can be shared for open peer review, such as *History Working Papers*¹⁶³ also produce information, although they may not be as participatory as other projects.

In 2008 the *Flickr Commons* site was launched for organisations to share photographs with no known copyright restrictions as a way to increase public awareness of their collections.¹⁶⁴ Because the *Flickr* site offered the ability for users to add tags and comments to images, cultural heritage institutions have been able to use both *Flickr* and *Flickr Commons* to collect information about their images. The Library of Congress found that, in addition to over 10 million page views and almost 67,000 tags between January and October 2008, a group of 20 or so contributors regularly provided 'place names, more precise dates, event names, and fuller names for individuals previously identified only by surname'.¹⁶⁵ The Smithsonian also found that images previously hosted on their own sites got more attention and gathered more information when posted on *Flickr*,¹⁶⁶ helping their

¹⁶² Peter Knupfer, 'The Coming of the Commons: An Important Message from H-Net', *H-Announce*, 19 August 2013, <http://h-net.msu.edu/cgi-bin/logbrowse.pl?trx=vx&list=H-Announce&month=1308&week=c&msg=P5u9Jgjp/raThoo1UrHp3Q&user=&pw=>.

¹⁶³ 'History Working Papers Project', *History Working Papers Project*, accessed 19 February 2015, <http://www.historyworkingpapers.org/>.

¹⁶⁴ 'Library of Congress Photos on Flickr - FAQ', *Prints and Photographs Reading Room, Library of Congress*, 22 October 2010, http://www.loc.gov/rr/print/flickr_pilot_faq.html.

¹⁶⁵ Michelle Springer et al., 'For the Common Good: The Library of Congress Flickr Pilot Project' (Library of Congress, 2008), http://www.loc.gov/rr/print/flickr_report_final.pdf.

¹⁶⁶ Elizabeth Olson, 'Smithsonian Expands Its Reach Through Social Media and the Public', *NYTimes.Com*, 16 March 2011, <http://www.nytimes.com/2011/03/17/arts/design/smithsonian-expands-its-reach-through-social-media-and-the-public.html>.

images reach new audiences and providing more information for the benefit of all.

Subsequently, other organisations have posted images in the hope of gathering information, include the UK National Archives' *Africa Flickr* set.

Metadata

Many projects focus on collecting metadata, especially information that helps describe, classify or provide access to objects, images and catalogue records, particularly where extra data can aid resource discovery or analysis.¹⁶⁷ Metadata terms can be applied to objects to describe their visual appearance and other characteristics, relate to words present in image or media files (for example, place names on a map), or identify entities such as people, places, events and concepts. Metadata can be selected from lists of known terms or entered as free text. Known terms are based on pre-determined classifications and are created through activities ranging from classifications assigned through simple decision trees (for example, is the item simply rounded or oblong?) to terms chosen from an existing vocabulary (for example, a list of possible place names or artistic genres). The outcome is sets of items grouped into categories.

Metadata terms suggested by the public can bridge the 'semantic gap' between the language used in catalogues and the public when added to discovery interfaces.¹⁶⁸ Some

¹⁶⁷ Both search and browse discovery mechanisms rely on metadata. A report from the Association of Research Libraries stated that 'the lack of any online description virtually amounts to no access'. ARL Working Group on Special Collections, 'Special Collections in ARL Libraries: A Discussion Report from the ARL Working Group on Special Collections' (Washington, DC: Association of Research Libraries, March 2009), <http://www.arl.org>. See also Max J. Evans, 'Archives of the People, by the People, for the People', *American Archivist* 70, no. 2 (2007): 387-400, <http://archivists.metapress.com/index/d157t6667854536g.pdf>.

¹⁶⁸ Jennifer Trant, 'Tagging, Folksonomy and Art Museums: Results of Steve.Museum's Research' (Archives & Museum Informatics, 7 January 2009), <http://conference.archimuse.com/files/trantSteveResearchReport2008.pdf>.

scholarly library catalogues offer the ability to add 'Notes & Tags'¹⁶⁹ about items, but they appear to be very little used, perhaps because neither the outcome, nor the purpose, of the action is clear. Some versions, such as *PennTags* are more clearly modelled on 'social bookmarking' seen in sites like *del.icio.us*, and may have a higher rate of participation because this gives users an indication of how the tags they add might aid resource discovery. The simplicity of most tagging processes may mask subtle differences in the intellectual activity involved.¹⁷⁰ Free-text image tagging projects designed to improve the discoverability of cultural heritage collections include *steve.museum*, Tiltfactor's *Metadata Games*, *Artigo*, *Museum Metadata Games* and *Your Paintings Tagger*. Audio-visual tagging projects include *Waisda?* and NARA's use of *Amara*.¹⁷¹ These projects generally elicit descriptions of things depicted in the image, and as such can be verified by comparing the most commonly added tags.¹⁷² The *Your Paintings Tagger* project has two types of tagging. The public are asked to tag paintings using controlled vocabularies for 'Things or Ideas', People, Places and Events, while self-nominated 'expert taggers' can provide information on dates and artistic styles.

The process of marking up full-text transcriptions to record entities like people, places, events or concepts creates a form of metadata. However, there is a trade-off between the advantages of marked-up data, as structured data supports better search and browsing

¹⁶⁹ Wording taken from the British Library catalogue.

¹⁷⁰ For example, artworks may be tagged with subjective tags (opinions and interpretations, useful for self-expression), personal tags (useful for organisation) and factual tags. Arends et al., 'Analysing User Motivation in an Art Folksonomy'. See also Shilad Sen et al., 'Tagging, Communities, Vocabulary, Evolution', *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work - CSCW '06*, 2006, 181, doi:10.1145/1180875.1180904.

¹⁷¹ Mary, 'What's New On Amara?', *NARAtions*, 24 February 2015, <http://blogs.archives.gov/online-public-access/?p=9397>.

¹⁷² Barring orchestrated vandalism campaigns designed to overwhelm verification systems, but even ordinary acts of vandalism seem rare in heritage crowdsourcing.

interfaces but requires more effort to record. For example, the process of marking up text in *Transcribe Bentham* adds to the complexity of the overall task, which may lower participation rates.¹⁷³ Similarly, the data entry form for the Open University's *Reading Experience Database* asks the participant to make subjective decisions about the genre or subject matter of a piece of text; the project is unlikely to encourage casual participants, but has succeeded in meeting the needs of specialists.

Georeferencing

Applying spatial coordinates to a map by georeferencing the image (for example, by matching points on a historic map to a modern map) could be considered a form of metadata creation, as could geocoding (applying geographic coordinates to a place name). Specialist interfaces support the creation of geospatial metadata. The New York Public Library (NYPL) supplement their menu indexing project with the *What's On The Menu Geotagger*, which asks the viewer to help 'Locate our menus on the Earth' by finding location information (street addresses, ship names, etc.) on a given menu, looking up that location on the provided map search and selecting from the list of possible results, and selecting the right level of zoom to represent the precision of the location information. The British Library *Georeferencer* project asks participants to visually match points on historical maps with modern maps. The NYPL *Map Warper* is a tool for rectifying (aligning) historical maps against precise modern maps. These can be relatively complex tasks, but present an enjoyable puzzle for some people.

Personal and experiential information

Some projects seek to access information held in family or personal memory; in some

¹⁷³ Tim Causer and Valerie Wallace, 'Building A Volunteer Community: Results and Findings from Transcribe Bentham', *Digital Humanities Quarterly* 6, no. 2 (2012), <http://www.digitalhumanities.org/dhq/vol/6/2/000125/000125.html>.

cases the information gathered can subsequently be supported with evidence, but in other cases it is impossible to verify. One of the aims of the Next Generation Finding Aid Project's *Polar Bear Expedition Digital Collections* project (launched in 2006) was to collect information in the form of comments.¹⁷⁴ The project found that most 'information sharing' comments offered either descriptive data or error corrections.¹⁷⁵ When documentation (e.g. a death certificate or discharge papers) was provided by a contributor, staff updated the archive system; when evidence was not provided, the system was not updated but the comment was left 'to allow for multiple voices'.¹⁷⁶

Expert information

Organisations with specialist collections have used websites to enable experts (whether amateur or professional) to record information about their items or topics. For example, the US National Postal Museum's 'virtual museum' project, *Arago*, asks those with expert knowledge 'in a specific area of philately or postal history'¹⁷⁷ to share that knowledge with others by creating content for their website. The project began in 2006, and in 2009 they reported that '75% of the site has been written by non-museum employees'.¹⁷⁸ The expertise required (or gained) for these tasks is a long way from simple 'type what you see' tags.

Annotations

Information recorded through markup, comments or annotations can be difficult to

¹⁷⁴ Elizabeth Yakel et al., 'Polar Bear Expedition Digital Collections: Enhancing Online Use through Digital Curation' (DigCurr 2007, Chapel Hill, North Carolina, 2007).

¹⁷⁵ Yakel et al., 'Polar Bear Expedition Digital Collections'.

¹⁷⁶ Yakel et al., 'Polar Bear Expedition Digital Collections'.

¹⁷⁷ 'Welcome to the Arago Volunteer Researcher Program', *National Postal Museum*, accessed 26 February 2015, <http://postalmuseum.si.edu/Arago/Researchers.html>.

¹⁷⁸ MeredithMJ, 'Looking for MCN 2009 Co-Presenter, Topic: Crowdsourcing', *MCN-L*, 9 April 2009, <https://www.mail-archive.com/mcn-l@mcn.edu/msg00162.html>.

categorise. Bradley and Vetch point out that annotations (defined as the 'engaged, cognitive process of writing a meaningful note') range in purposes from aide-memoire to interpretive act.¹⁷⁹ Annotations on entries on *Pepys Diary*¹⁸⁰ include questions, definitions, background information about specific people, places, events or concepts, and commentary on specific lines of the text.¹⁸¹ The *Suda On Line* project (launched in 1998) which translated a 10th century Byzantine Greek historical encyclopaedia into English through a process open to anyone who 'possessed the ability to translate ancient Greek, regardless of formal credentials and specialization', also encouraged contributors to annotate the translations with links and further information to contextualise and disambiguate entries for modern readers.¹⁸²

History mysteries

'History mysteries' combine serendipitous discovery with requests for specific information. In some instances, the right person - perhaps a family member or a local with a long memory - who happened to stumble across an item might be able to instantly identify it, but an answer might also be reached cooperatively as people with different types of expertise share their knowledge. For example, Te Papa Museum's blog has 'Photo-Detective' posts asking for help identifying specific things within the image. They have received responses from experts in New Zealand's military history, historical fashion, and vintage cars, and from people who remember the fashions and cars from the time and can

¹⁷⁹ John Bradley and Paul Vetch, 'Supporting Annotation as a Scholarly Tool-Experiences From the Online Chopin Variorum Edition', *Literary and Linguistic Computing* 22, no. 2 (June 2007): 225-41, doi:10.1093/llc/fqm001.

¹⁸⁰ A grassroots website that presented entries from Pepys' diary in 'real time' from 2003 to 2012.

¹⁸¹ For an example, see <http://www.pepysdiary.com/diary/1662/02/25/>. Last accessed 6 June 2015.

¹⁸² Anne Mahoney, 'Tachypaedia Byzantina: The Suda On Line as Collaborative Encyclopedia', *Digital Humanities Quarterly* 3, no. 1 (2009), <http://www.digitalhumanities.org/dhq/vol/003/1/000025.html>.

add information from their own experience.¹⁸³ The Museum of Design in Plastics' *10 Most Wanted* site (launched in November 2013) aims to involve the public in researching plastic objects from their collection. The research challenges (for example, find the designer, manufacturer or place of manufacture for an object) are difficult for the casual participant who may not know which sources to begin to turn to; however they have succeeded in attracting some experts and answering some of their questions.¹⁸⁴ Other projects include *HistoryPin's Year of the Bay* and *Putting Art on the Map*¹⁸⁵ mysteries, which post requests to find information such as dates and locations for photographs or paintings on their website and social media.¹⁸⁶ While projects that require expert participation can attempt to recruit existing communities to help, projects that seek personal identifications are more reliant on serendipitous encounters between the appropriate image and the rare person with the right knowledge.¹⁸⁷

Research tasks

During my site review, I collected several forms of research tasks that do not seem to have been categorised in the existing literature - distributed reading, research commissions, and

¹⁸³ Athol McCredie, 'Help Wanted: Who, What, When?', *Te Papa's Blog*, 3 February 2015, <http://blog.tepapa.govt.nz/2015/02/03/help-wanted-who-what-when/>. See also Athol McCredie, 'More Help Wanted: Photo-Detective #2', *Te Papa's Blog*, 7 February 2015, <http://blog.tepapa.govt.nz/2015/02/07/more-help-wanted-photo-detective-2/>. Athol McCredie, 'You Can Play Photo-Detective #3', *Te Papa's Blog*, accessed 4 March 2015, <http://blog.tepapa.govt.nz/2015/02/19/you-can-play-photo-detective-3/>.

¹⁸⁴ Susan Lambert, Marcus Winter, and Phil Blume, 'Getting to Where We Are Now', *10most.Org.Uk*, 26 March 2014, <http://10most.org.uk/content/getting-where-we-are-now>.

¹⁸⁵ Rebekkah Abraham, 'Reflections on Putting Art on the Map', *Historypin Blog*, 20 April 2014, <http://blog.historypin.com/2014/04/20/reflections-on-putting-art-on-the-map/>.

¹⁸⁶ Kerri Young, 'Year of the Bay Project Officer Evaluation' (San Francisco: HistoryPin, May 2014).

¹⁸⁷ Given later discussion of the importance of search engines, any information or activity that helps make a record more discoverable will increase the odds. However, it can take time for records to be indexed by search engines, which creates a challenge for short-term projects.

research quests. Research tasks themselves are not new,¹⁸⁸ but these forms seem relatively unstudied as categories of participatory tasks. In these research tasks, unlike user-generated content style 'please share any information you have about this item' comment boxes or the history mysteries discussed above, participants are first recruited then assigned particular research tasks.¹⁸⁹ One form of research task, which I have called 'distributed reading', involves reading assignments set on particular topics, periods, and/or sources, in which any relevant information discovered is sent back to the organiser. 'Research commissions' are requests for specific individuals to research specific topics; this is much like appointing a research assistant except the role is voluntary. I also found community history societies setting 'research quests', research projects based on the records available and interests of their members, including the many projects of the Family and Community Historical Research Society (FACHRS)¹⁹⁰ and other local history societies. FACHRS are perhaps unusual in having documented their 'major' and 'mini' project models. Mini projects generally run for a month and can be completed with

¹⁸⁸ For example, the Public Catalogue Foundation's *Art Detective* project set 123 questions and resolved 59 in their first year. Edward Stone, 'Happy Birthday Art Detective!', *The Public Catalogue Foundation*, 19 March 2015, http://www.thepcf.org.uk/what_we_do/48/reference/951.

I have also experimented with crowdsourcing research tasks through a (not very successful, overall) experimental 'trivia' game that explored emergent game-play around longer forms of content that required some form of research or personal reference. Mia Ridge, 'Playing with Difficult Objects: Game Designs for Crowdsourcing Museum Metadata' (MSc Dissertation, City University London, 2011), <http://www.miaridge.com/my-msc-dissertation-crowdsourcing-games-for-museums/>.

¹⁸⁹ However, in the case of some FACHRS projects, membership of the society may be enough for a historian to be 'opted in' and asked to undertake a small research task. For example, all registered FACHRS members were sent information about a specific historical individual to research, as reported in Sue Smith, 'Work in Progress Victorian Stationmasters: Origins and Outcomes', *Family & Community History* 15, no. 2 (October 2012): 113–17, doi:10.1179/1463118013Z.0000000007.

¹⁹⁰ FACHRS focus on 19th and 20th century community history and appeal to family and local historians who wish to expand their research focus. FACHRS, 'Projects', *Family and Community Historical Research Society*, accessed 10 June 2014, <http://www.fachrs.com/projects.html>.

records available online, while major projects run for a year or more, require visits to archives and 'become a real labour of love'.¹⁹¹

The Washington State Historical Society's *Civil War Pathways* project (launched in 2013) is an example of 'distributed reading' in which participants read specific sources. They organised a 'Read-In Volunteer Program' that assigned participants specific primary documents ('newspapers, published classic histories, and archival material of all sorts'¹⁹²) to read online or in physical museums, libraries or archives. Participants uploaded any traces of the 'forgotten Civil War experience in Washington Territory' they found to an 'on-line searchable database of Civil War-era citations and documents'.¹⁹³ Participants were supported through an initial in-person training session and an online forum, and the project resulted in 2800 records being made available.¹⁹⁴ The FACHRS *Swing Riots* project is another example of a distributed reading task.¹⁹⁵

I have included here three quite different examples of volunteer 'research commissions'. *Founders and Survivors* is an academic prosopographical project presented as 'a partnership between historians, genealogists, demographers and population health researchers' that aims to trace the histories and descendants of over 70,000 people

¹⁹¹ FACHRS, 'Projects'.

¹⁹² Civil War Pathways, 'About', *Civil War Pathways*, accessed 27 March 2014, <http://web.archive.org/web/20130515144248/http://pathways.omeka.net/about>.

¹⁹³ Civil War Pathways, 'About'.

¹⁹⁴ Civil War Pathways, Survey: Crowdsourcing and public participation in digital history, interview by Mia Ridge, June 2014.

¹⁹⁵ Michael Holland, 'Swing Revisited: The Swing Project', *Family & Community History* 7, no. 2 (2004): 87–100, <http://www.maneyonline.com/doi/abs/10.1179/fch.2004.7.2.002>. See also FACHRS, 'The Swing Riots Project', *Family and Community Historical Research Society*, accessed 22 January 2015, <http://www.fachrs.com/Information/Swing.html>.

transported to Tasmania.¹⁹⁶ Launched in February 2009,¹⁹⁷ the site allowed family historians to enter details of convict ancestors into a database, along with details of any related documents. Some participants were invited to participate in the Ships Project and asked to trace the lives of every convict on a given ship, tracing marriages, births and deaths of the convicts, their children and any descendants who fought in the First World War.¹⁹⁸ The United States Holocaust Memorial Museum's *Children of the Lodz Ghetto* is another example of the research commission pattern. It was set up in 2008 as a 'citizen history' project designed to encourage more people to become historians, or 'at least make history and historical thinking more accessible to participants'.¹⁹⁹ Participants are challenged to find out what happened to 14,000 students from the Lodz Ghetto who signed a school album in 1941.²⁰⁰ The research task is not easy, particularly as sources use mid-20th century European conventions, and students may have gone by Polish or Yiddish versions of their names, and signed the album with full names or just initials.²⁰¹ The final example of a research commission task is *Lives of the First World War (LFWW)*, launched in 2014 by the Imperial War Museum and the commercial firm DC Thomson Family History. The primary message, 'Find your connection',²⁰² seems to be aimed at family

¹⁹⁶ Founders and Survivors, 'About the Project', *Founders and Survivors*, accessed 27 May 2014, <http://www.foundersandsurvivors.org/project>.

¹⁹⁷ The University of Melbourne, 'Is There a Tasmanian Convict Settler in Your Family Tree? Participants Sought for Study into Australia's Founding Population.', *The University of Melbourne*, 9 February 2009, <http://archive.uninews.unimelb.edu.au/view-54040.html>.

¹⁹⁸ Founders and Survivors, 'Ships Project News', *Founders and Survivors*, 22 October 2012, <http://foundersandsurvivors.org/content/ships-project-news>.

¹⁹⁹ Elissa Frankle, 'More Crowdsourced Scholarship: Citizen History', *Center for the Future of Museums*, 28 July 2011, <http://futureofmuseums.blogspot.co.uk/2011/07/more-crowdsourced-scholarship-citizen.html>.

²⁰⁰ Frankle, 'More Crowdsourced Scholarship: Citizen History'.

²⁰¹ Frankle, 'More Crowdsourced Scholarship: Citizen History'.

²⁰² In May 2015 the most visually prominent message on the front page is 'Who will you Remember?'

historians, but the site also encourages anyone who can to create 'Life Stories' for more than 8 million people 'whose contribution to the First World War is recorded in official documents'.²⁰³ Some participants have created records for thousands of individuals²⁰⁴ sometimes linked to particular communities or war memorials.²⁰⁵ Projects like these are an interesting study because participation requires undertaking reasonably complex tasks, whether providing evidence about convict ancestors, taking on assignments to research individual lives (with all the usual difficulties of tracing working-class lives, particularly those who may have changed biographical details to disguise their convict origins) or to trace the whole cohort of convicts on transportation ships in the Ships Projects.²⁰⁶

Repositories with additional participatory functions

Repositories are generally designed to host historical collections or scholarly outputs such as articles, and are an important resource of primary and secondary sources for digital history.²⁰⁷ Some repositories are similar to printed anthologies, taking the form of

²⁰³ Lives of the First World War, 'About', *Lives of the First World War*, 2014, <https://livesofthefirstworldwar.org/about>.

²⁰⁴ e.g. <https://livesofthefirstworldwar.org/profile/321> or <https://livesofthefirstworldwar.org/profile/27542>; both last accessed 6 June 2015.

²⁰⁵ e.g. <https://livesofthefirstworldwar.org/community/1568>. Last accessed 6 June 2015.

²⁰⁶ Janet McCalman et al., 'Building a Life Course Dataset from Australian Convict Records; Founders & Survivors: Australian Life Courses in Historical Context, 1803-1920', in *Proceedings Workshop Population Reconstruction* (Amsterdam, 2014), http://socialhistory.org/sites/default/files/docs/mccalman_kippen_silcot_smith_-_building_a_life_course_dataset.pdf. Another Australian project has asked family historians to 'donate' their research to a project studying The Benevolent Society's Asylum but does not seem to have asked specific family historians to undertake new research on unrelated individuals. Tanya Evans and Patricia Curthoys, 'Family History, Identity, and Public History: Writing a History of The Benevolent Society in Its 200th Year', *Journal of Australian Studies* 37, no. 3 (September 2013): 285-301, doi:10.1080/14443058.2013.814582.

²⁰⁷ For those who regularly use archives or work with GLAM specialists, the term 'archive' invokes specific disciplinary practices and the casual use of the term can be confusing. For discussion of the

interpretive exhibits or explanatory sites curated around a particular topic or question, while others aim to be comprehensive repositories of digitised sources or reflect the collections of the related physical archive.²⁰⁸ Some repositories that support participatory functions or outputs are mentioned here, as well as non-participatory projects that have significantly influenced the conceptual models or boundaries of digital history.

Until 2007, historians were able to deposit research data with the Arts and Humanities Data Service (AHDS), where it would be catalogued, preserved and optionally made available to other researchers.²⁰⁹ Research can now be deposited with the *UK Data Service*, which provides online access and downloadable research.²¹⁰ In contrast to highly-focused sites developed around a specific topic, period or research question, *Wikimedia Commons* will accept any material that is 'realistically useful for an educational purpose'²¹¹ and

different uses of the word 'archive', see archivist Kate Theimer. Kate Theimer, 'A Distinction Worth Exploring: "Archives" and "Digital Historical Representations"', *Journal of Digital Humanities* 3, no. 2 (Summer 2014), <http://journalofdigitalhumanities.org/3-2/a-distinction-worth-exploring-archives-and-digital-historical-representations/>.

²⁰⁸ One senior historian I interviewed calls the former 'synthetic' sites. The *Journal of American History's* information on Digital History Reviews also distinguishes between 'archives' (a body of primary documents) and essays/exhibits while the British Reviews in History site lists only 'Digital resources' with no further distinction. Jeffrey W. McClurken, 'Digital History Reviews', *Journal of American History*, accessed 11 February 2015, <http://www.journalofamericanhistory.org/submit/digitalhistoryreviews.html>.

²⁰⁹ AHDS History, 'AHDS History - Depositing Data with the AHDS', *AHDS History*, University of Essex, March 2007, <http://www.ahds.ac.uk/history/depositing/index.html>.

²¹⁰ While their focus is 'social and economic data', some historical datasets are available. Access to some data is restricted to UK-based academics, though they encourage data owners to 'identify and remove all unnecessary barriers to access'. UK Data Service, 'Open Access Data', *UK Data Service*, accessed 16 May 2015, <http://ukdataservice.ac.uk/get-data/open-data.aspx>.

²¹¹ The number of records is related to the fact that *Wikimedia Commons* does not have the same 'notability' requirements as *Wikipedia*. Wikimedia contributors, 'Commons:Project Scope', *Wikimedia Commons*, 2015, https://commons.wikimedia.org/wiki/Commons:Project_scope.

contains over 25 million files available for re-use.²¹² *Zotero Commons* was launched in 2007 to provide a 'central home' and permanent archive for scanned primary sources, photos or 'other useful scholarly documents'; the text from items shared would also be OCR'd by the underlying *Internet Archive* platform. However, it had very little use while still actively supported,²¹³ suggesting the difficulties in convincing scholars to share their data publicly. Meanwhile, the relatively small but consistently active *Parallel Archive* has over 2800 items, perhaps because the organisation behind it actively encourages scholars on research fellowships to 'digitize, tag, comment, and make publicly available' the material they access in the physical archive.²¹⁴ This is a rare example of a successful 'participant digitisation' project (a term coined for my original thesis proposal to describe the act of collecting and sharing digital records and information created when researchers access primary materials), and its success may be due to its embedding in existing relationships of trust rather than innovative technical or interface solutions.

The *London Lives* site is a repository with participatory functions that not only provides access to over '240,000 manuscript and printed pages', but allows registered users to save search results, create 'sets' of documents²¹⁵ and write biographies of 18th century Londoners on the *London Lives Wiki* (launched in October 2010).²¹⁶ Three repository

²¹² Wikipedia contributors, 'Wikimedia Commons', *Wikipedia, the Free Encyclopedia*, 2013, https://en.wikipedia.org/w/index.php?title=Wikimedia_Commons&oldid=581836800. In May 2015. Statistics from <https://commons.wikimedia.org/wiki/Special:Statistics>. Page last accessed 6 June 2015.

²¹³ 565 items in May 2015, including spam items. Statistics from <https://archive.org/details/zoterocommons&tab=about>. Page last accessed 6 June 2015.

²¹⁴ OSA Archivum, 'About Us', *Open Society Archives Archivum*, accessed 6 June 2015, <http://www.osaarchivum.org/about-us>.

²¹⁵ London Lives, 'About This Project', *London Lives*, April 2012, <http://www.londonlives.org/static/Project.jsp>.

²¹⁶ Sharon Howard, 'London Lives Wiki', *London Lives News*, 5 October 2010,

sites²¹⁷ based on the same *Collex* software for federated sites, Networked Infrastructure for Nineteenth-Century Electronic Scholarship (*NINES*), *18thConnect* and the Medieval Electronic Scholarly Alliance (*MESA*) offer the same ability to search metadata or full texts from a range of sites. Scholarly digital projects can apply to be listed in these sites; the peer review process they must pass before inclusion is designed to help establish the credibility of digital humanities projects,²¹⁸ and projects that pass peer review are eligible for a letter (e.g. from the Director of *18thConnect*) that can be submitted in applications for academic promotion.²¹⁹ Users who have created accounts on the sites have access to several actions for items: tagging, adding private annotations, discussing them on a forum, 'collecting' items, and optionally publishing them as an 'exhibit'. *18thConnect* also provides transcription correction through *TypeWright*. Virtual research environments or scholarly 'workbenches' such as *AustESE*, *TextGrid* and *CENDARI* are spaces in which researchers can use tools for collecting, automatically or manually transcribing, linking and annotating records within collections of data.²²⁰

Commercial genealogy sites provide other examples of repositories with additional functions. Sites like *Ancestry*, *MyHeritage* and *Findmypast* encourage users to create and publish family trees based in part on the records found in their repositories. For example,

http://www.hrionline.ac.uk/londonliveswiki/tiki-view_blog_post.php?postId=2.

²¹⁷ Strictly speaking they are aggregators (collecting links to items held in other repositories) but, as the distinction is not obvious to the casual user and some participatory functions are hosted on the centralised sites, I have included them here.

²¹⁸ *18thConnect*, 'Peer Review', *18thConnect*, accessed 16 May 2015, <http://www.18thconnect.org/about/scholarship/peer-review/>.

²¹⁹ *18thConnect*, 'About', *18thConnect*, accessed 19 February 2015, <http://www.18thconnect.org/about/>.

²²⁰ JISC, 'Virtual Research Environment Programme', 2013, <http://webarchive.nationalarchives.gov.uk/20140702233839/http://www.jisc.ac.uk/whatwedo/programmes/vre.aspx>.

records on *Ancestry* can be saved to individuals in a tree or added to a 'shoebox' of unsorted material. Within a particular site, family trees are often 'public' by default; the owner of a tree must grant another account permission to view a 'private' tree. Family trees on commercial genealogy sites are often only visible to people logged into accounts on that site, though accounts may not require a paid subscription.

As sites that have grown with the interests of their users over the years, local history societies often hold a mixture of document indices, repositories and local information. For example, groups such as the *Oxfordshire Family History Society*, *Manchester and Lancashire Family History Society*, and the *Hanslope and District Historical Society* have transcribed parish registers, wills, census records and hearth taxes, and created indexes for other local records and books relevant to the area or local occupations. These records may be sold as CDs, or made available online. Several village history websites in the Milton Keynes area use specialised software written by the late Robert Dymond to link different types of local records to specific people and addresses.²²¹ Some societies have online forums but most seem to rely on in-person meetings or private correspondence to manage participatory tasks.

Other outputs

My review uncovered a range of unique and emerging tangible outputs that did not naturally fit into other categories. The *Micropasts* project has a photomasking task that asks participants to draw the outline of an artefact in a photograph; the data can then be used to create 3D models. *Pybossa*, the platform underlying *Micropasts*, can be used to create mobile phone data collection projects and tasks including text transcription, image,

²²¹ Robert Dymond, personal communication, March-April 2014. Robert Dymond, 'Community Heritage Digital Archive: Functional Description', n.d. Unfortunately Mr Dymond passed away before we could meet to discuss his work for local history societies.

audio and video classification and georeferencing. This multi-purpose platform may provide a model for future projects.²²²

Since 2005 *LibriVox* volunteers have been recording themselves reading books aloud to create free public domain audiobooks.²²³ *Distributed Proofreaders* also offers a 'smooth reading' task where participants can read the texts for pleasure,²²⁴ and report any proofing errors they encounter. In some ways, this purposive 'noticing' is similar to science observation projects such as *iSpot* that ask participants to report on wildlife and nature. The Oxford English Dictionary's 'Appeals' to identify uses of words or find sources could be viewed as research tasks or directed noticing.²²⁵

Conclusion

The participatory digital history projects presented here were created for a number of reasons. Some were initiated by museums, libraries or archives seeking to provide public access to their collections, while others were designed for and by academic historians. Commercial projects are particularly well represented in family history and genealogy. While many projects were initiated by institutions who asked the public to help digitise, transcribe, describe or otherwise enhance collections, several grassroots projects were created by people who chose to collaborate on a shared task, and a few are the work of dedicated individuals.

²²² The Zooniverse *Panoptes* platform will provide image-based functions such as classification, marking points or drawing shapes on images, but might provide other functions in future. Chris Lintott, 'Open Zooniverse: Beta Testers Required', *Citsci-Discussion-L*, June 2015.

²²³ 'About LibriVox', *LibriVox*, accessed 27 February 2015, <https://librivox.org/pages/about-librivox/>.

²²⁴ In contrast to reading text while checking transcribed text against the original document.

²²⁵ Peter Gilliver, "Your Dictionary Needs You": A Brief History of the OED's Appeals to the Public', *Oxford English Dictionary*, 4 October 2012, <http://public.oed.com/the-oed-appeals/history-of-the-appeals/>.

The review of over 400 projects that aim to engage the public and/or collect, create or enhance records about historical materials for scholarly and general audiences presented in this chapter led to the development of a simple classification of participatory digital history projects. As discussed in the opening pages of this chapter, this classification, in which projects were loosely grouped according to their output types, allowed seemingly disparate projects to be compared. Analysis of the contrasting choices made by projects seeking a common output - in platforms, community involvement, organisational structure, participant motivation and task types, etc. - supported the examination of the impact of project and interface design that underlies this thesis.²²⁶ This classification, together with interviews with nearly 30 historians, also led to the definition of distributed reading, research quest and research commission projects.

Looking back at the projects reviewed, a general movement from individual, bespoke software or websites, to the use of social media platforms such as *Flickr* or *Facebook*, to the adoption of common platforms such as *DIYHistory|transcribe*, *FromThePage*, *Pybossa* or *Panoptes*) is apparent. Regional and specialist communities are still important, but it is now easier for people to discover projects in which they might be interested. Project designs are influenced by research and organisational practices (e.g. transcription, look-up requests, tasks that aid discoverability), modelled on projects from other non-profit, academic and commercial domains, and take lessons from the apparent successes or failures of other projects.

Revising projects reviewed early in my research for the final stages of writing this thesis proved, as expected, that the field of participatory digital history moves quickly. Many of

²²⁶ My broad classification focused on projects gathered for my review, rather than attempting to comprehensively address the fast-moving field of non-profit or heritage crowdsourcing.

the projects I reviewed changed over time - some added new content, or new types of content; removed or introduced new functionality; or abandoned their official project blog for updates on social networks, code repositories or journal articles; others disappeared after a website was restructured, domain name expired or a social media platform closed. Other projects were difficult to track as they changed their name or focus between initial funding announcements and the final product; others turned out to be 'vapourware' - software that is announced but never comes into existence. Acquisitions and partnerships in the commercial sector added another layer of complexity, as smaller sites were folded into larger ones.²²⁷ A final check of project links while compiling the related *Appendix* of websites reviewed hints at the extent to which sustainability and digital preservation will be an issue for future historians. Over 30 project links were unavailable except through the *Internet Archive*,²²⁸ and some of this research would not have been possible had the *Internet Archive* not existed. Sustainability and preservation is a key issue for the field of digital and participatory history.

The process of reviewing this significant body of projects has lead to several questions that inform the development of the next two chapters. The definition of 'success' is clearly contextual, depending on a project's goals, potential audience and material; but can common characteristics for effective designs be determined from this corpus? Successful projects were often based around a particular locality, topic, research community or type

²²⁷ For example, finding material in the USGenWeb Project Archives can be difficult where older links point to RootsWeb, which was bought by Ancestry. Some sites have moved but their older versions remain on merged sites; others used domain names that have since expired.

²²⁸ For example, Ancestry closed several services in 2014. As they did not provide effective data export services, the content of these sites has been lost to its creators as well any future researchers as discussed in the comments about data lost from MyFamily, MyCanvas, Genealogy.com, Mundia and more on Eric Shoup, 'Ancestry.Com Focuses on Core Offerings', 4 June 2014, <http://blogs.ancestry.com/ancestry/2014/06/04/ancestry-com-focuses-on-core-offerings/>.

of source, suggesting that 'niche' projects are capable of motivating dedicated activity,²²⁹ a question taken up in later chapters. Other successful projects featured highly polished microtasks, reflecting the importance of appropriate interaction design for participatory history projects. What is the effect of attributes such as task complexity, participant motivation, communities of practice, the tone of communication with participants, or the organisational model on the success of a project?

This chapter has set out a landscape of participatory history projects; the next focuses on projects that are specifically designed to engage the public in tasks related to creating or enhancing heritage resources. *Chapter 2* investigates the common factors that help crowdsourcing projects work with the public to enhance records about historical newspapers, maps, menus and more. Several participatory projects are discussed further in the third chapter on citizen history: *Herbaria@Home*, *Old Weather*, *Children of the Lodz Ghetto*, *Operation War Diary* and *Marine Lives*. These projects were selected because they offer insights into the development of 'citizen scientists' or 'citizen historians'.

²²⁹ Similarly, Melissa Terras found that 'standalone virtual amateur museums' often had a 'narrow, delineated focus, and very specific remit'. Terras, 'Digital Curiosities'.

Chapter 2: History with the public: crowdsourcing

Why are some heritage crowdsourcing projects more successful than others? Based on my analysis of voluntary crowdsourcing projects in this chapter, I suggest that the most important factors are good publicity, appropriate task design, and connecting the project to both participant motivations and a shared, significant goal. Following the broad overview of participatory digital history projects in the previous chapter, this chapter focuses on projects specifically designed to engage the public in the process of creating digital resources. The results inform the wider question of how digital technologies have shaped public participation and scholarly practices in historical research. The results also contribute to the growing body of knowledge about the most effective ways to encourage people to enhance historical resources through crowdsourcing. Heritage crowdsourcing projects are shaping the fields of digital and/or public history in several ways.¹

Crowdsourcing projects provide opportunities for the public to actively engage with the materials and research practices of related disciplines such as science and history. This impact on participants may be as important as their ability to process large collections of material for institutions. In this chapter, I discuss ways in which the success of a crowdsourcing project can be measured, relate the literature on motivations to

¹ Not least because they may be more likely to receive funding than traditional digitisation projects. For example, Tim Causer and Melissa Terras note that 'no organisation would ever give the Bentham Project sufficient funding to contract out the transcription', but they were able to get funding to crowdsource the transcription of the remaining manuscripts. Tim Causer and Melissa Terras, "Many Hands Make Light Work. Many Hands Together Make Merry Work": Transcribe Bentham and Crowdsourcing Manuscript Collections', in *Crowdsourcing Our Cultural Heritage*, ed. Mia Ridge, Digital Research in the Arts and Humanities (Farnham, Surrey, UK: Ashgate, 2014), <http://www.ashgate.com/isbn/9781472410221>. p. 85.

participation to specific projects, and then present an analysis of the characteristics of successful projects.

'Crowdsourcing' is becoming an over-used and often vaguely applied term. For the purposes of my research, I have used the following definition: crowdsourcing in cultural heritage asks the public to help with meaningful tasks that contribute to a shared, significant goal or research interest related to cultural heritage collections or knowledge.² As a voluntary activity, the tasks and/or goals should be inherently rewarding. This definition focuses on the necessity for rewards that match each party's motivations for participating, while recognising that those rewards may take many forms. It also draws on Rose Holley's definition of crowdsourcing as a group cooperating to achieve a 'shared, usually significant, and large goal'³ and Stuart Dunn and Mark Hedges' statement that 'humanities crowdsourcing' requires 'a clearly-defined humanities direction and/or research question'.⁴ Unlike user-generated content projects,⁵ crowdsourcing is inherently productive in intent: each activity should contribute to a meaningful, collective goal. At a macro level, crowdsourcing projects are designed as collaborative projects; however, an individual user completing a crowdsourced task may not realise that they are contributing

² While my research here is focused on history, this definition was created to reflect the wider field of cultural heritage.

³ Holley, 'Crowdsourcing: How and Why Should Libraries Do It?' Holley is the former manager of the Australian Newspapers Digitisation Program and Trove and her publications on crowdsourcing in libraries did much to stimulate interest and projects in the field.

⁴ Dunn and Hedges, 'Crowd-Sourcing Scoping Study'.

⁵ For example, comments on newspaper articles allow the public to discuss the issues raised, but the main benefit for the newspaper or journalist of this activity is the increased site traffic rather than the content of the comments themselves. For further discussion of the difference between user-generated content and crowdsourcing, see Mia Ridge, 'From Tagging to Theorizing: Deepening Engagement with Cultural Heritage through Crowdsourcing', *Curator: The Museum Journal* 56, no. 4 (October 2013).

to a larger task.⁶

Like other recently fashionable technological terms, the definition of crowdsourcing has been appropriated to suit many other goals. Its application to a variety of new and pre-existing related projects and commercial ventures has pushed at the boundaries of its definition. Following their 2012 overview of definitions of crowdsourcing, researchers Estellés-Arolas and González-Ladrón-de-Guevara noted that 'crowdsourcing' was evolving to the extent that the label might be applied to almost any internet-based collaborative activity.⁷ Discomfort with the term 'crowdsourcing' means that some have turned to descriptions such as 'community-sourcing',⁸ 'nichesourcing',⁹ 'micro-volunteering'¹⁰ or 'targeted crowdsourcing'.¹¹ These names acknowledge that often the 'crowd' of participants is neither large nor truly anonymous, and may also suggest institutional discomfort with

⁶ Here my views differ from Holley, who says that crowdsourcing involves 'sustained input from a group of people working towards a common goal'. More recent crowdsourcing projects have been designed to benefit from casual and one-off contributions as well as more sustained contributions. Holley, 'Crowdsourcing: How and Why Should Libraries Do It?'

⁷ Enrique Estelles-Arolas and Fernando Gonzalez-Ladron-de-Guevara, 'Towards an Integrated Crowdsourcing Definition', *Journal of Information Science* 38, no. 2 (March 2012): 189–200, doi:10.1177/0165551512437638.

⁸ Amy Sample Ward, 'Crowdsourcing vs Community-Sourcing: What's the Difference and the Opportunity?', *Amy Sample Ward's Version of NPTEch*, 18 May 2011, <http://amysampleward.org/2011/05/18/crowdsourcing-vs-community-sourcing-whats-the-difference-and-the-opportunity/>.

⁹ Victor de Boer et al., 'Nichesourcing: Harnessing the Power of Crowds of Experts', in *Proceedings of the 18th International Conference on Knowledge Engineering and Knowledge Management, EKAW 2012* (Springer, 2012), 16–20, http://dx.doi.org/10.1007/978-3-642-33876-2_3.

¹⁰ A term used by the New York Public Library. Lascarides and Vershbow, 'What's on the Menu?'

¹¹ For example, the National Library of Wales' The Welsh Experience of the First World War project. Lyn Lewis Dafis, Lorna M. Hughes, and Rhian James, 'What's Welsh for "Crowdsourcing"? Citizen Science and Community Engagement at the National Library of Wales', in *Crowdsourcing Our Cultural Heritage*, ed. Mia Ridge (Farnham, Surrey, UK: Ashgate, 2014), <http://www.ashgate.com/isbn/9781472410221>.

the vastness and anonymity conjured by the 'faceless'¹² crowd. While the term 'crowdsourcing' is often applied to projects that have neither outsourced work nor involve crowds, it is difficult to coin another with the same levels of immediate recognition. Forms of crowdsourcing occur in many sectors but I have generally excluded commercial projects, crowdfunding and design competitions¹³ from this research because the motivations and rewards for participation in commercial projects are different from those in the not-for-profit heritage sector.¹⁴

Citizen science projects involve members of the public assisting professional scientists with research,¹⁵ most commonly through data processing tasks like image classification but potentially also through fieldwork or observation tasks, data analysis or research design.¹⁶ Citizen science projects represent a significant body of prior work for public participation in scholarly research, and provide useful examples of crowdsourcing projects as a form of disciplinary education. As there are still relatively few historical crowdsourcing projects, let alone 'citizen history' projects, the inclusion of citizen science projects provides a wider range of approaches for consideration (an aspect that will be discussed further in *Chapter 3*). I have used the term 'heritage crowdsourcing' for cultural heritage and humanities projects¹⁷ to distinguish them from citizen science projects, and I have used 'heritage'

¹² de Boer et al., 'Nichesourcing'.

¹³ Payment for services rendered is common in the commercial sector, most notably through platforms like Amazon's Mechanical Turk, which brokers the distribution of and payment for 'Human Intelligence Tasks', or 'HITs'. In most design contests, every entrant completes the task but only those with winning designs are financially compensated.

¹⁴ Some historical projects invoke altruistic goals in the service of for-profit companies; the effects of this are discussed at various points in the following chapters.

¹⁵ M. Jordan Raddick et al., 'Galaxy Zoo: Exploring the Motivations of Citizen Science Volunteers', *Astronomy Education Review* 9, no. 1 (2010): 18, <http://dx.doi.org/10.3847/AER2009036>.

¹⁶ Bonney et al., 'Public Participation in Scientific Research'.

¹⁷ Some previous research has focused on either cultural heritage or humanities crowdsourcing. For

rather than 'history' to include projects based on historical collections initiated by individuals or organisations outside the discipline of history. While my focus is largely on historically-focused projects from community groups, academic or memory institutions,¹⁸ I have included some 'citizen science' and digital humanities projects in the site reviews that fed into my analysis.

Defining and measuring the success of heritage crowdsourcing projects

The commercial sector has developed metrics such as 'conversion rate' ('the number of potential workers aware of a task that choose to accept the task') to measure effective task design in terms of the number of potential participants who respond to an advertised task.¹⁹ Researchers who view collaborative digital projects as technology-mediated social systems have suggested measures such as member satisfaction, the type and degree of cooperation, network dynamics and 'community efficacy'.²⁰ These could be adapted for heritage projects, but, as the goals of heritage crowdsourcing might go beyond pure productivity or efficiency to encompass participant engagement and learning outcomes,²¹ additional metrics tailored to these goals should also be considered. Based on my research, potential quantitative metrics for measuring the success of heritage crowdsourcing

example, Dunn and Hedges, 'Crowd-Sourcing Scoping Study'. Oomen and Aroyo, 'Crowdsourcing in the Cultural Heritage Domain'.

¹⁸ That is, museums, libraries and archives.

¹⁹ Jason T. Jacques and Per Ola Kristensson, 'Crowdsourcing a HIT: Measuring Workers' Pre-Task Interactions on Microtask Markets', in *First AAAI Conference on Human Computation and Crowdsourcing*, 2013, <http://www.aaai.org/ocs/index.php/HCOMP/HCOMP13/paper/view/7540/o>.

²⁰ Robert Kraut et al., 'Scientific Foundations: A Case for Technology-Mediated Social-Participation Theory', *Computer* 43, no. 11 (November 2010): 22–28, doi:10.1109/MC.2010.324.

²¹ Here I am thinking of 'generic learning outcomes' which include changes in 'knowledge and understanding', 'attitudes and values' and 'activity, behaviour and progression' among participants. Museums, Libraries and Archives Council, 'Generic Learning Outcomes', *Inspiring Learning*, 2008, <http://www.inspiringlearningforall.gov.uk/toolstemplates/genericlearning/>.

projects include: the number of hours participants have spent on a project; initial and sustained participation rates;²² participant retention; the form and extent of use for community discussion platforms; the number of tasks completed; and the percentage of tasks validated against required quality standards. Valuable but less easily measured outcomes include attitudinal change, the extent to which participants gain related skills and knowledge, or the number of new research questions or discoveries that emerge during a project. Qualitative measures include the extent to which participants express support or appreciation for the project, the number of participants who pursue activities related to their new interest, or some wider impact on participants' behaviour or attitudes.²³

Some of these metrics are inherently opposed: time spent posting on community platforms or learning about the underlying discipline means less time is available to spend on the core task. Accordingly, measurements of success should be judged and weighted according to the overall goals of an individual project. Different sections or functions of a project will also require appropriate measures of success. For example, the efficiency of individual task and interface elements may be judged in terms of how many input items are processed per minute of participant time, while the effectiveness of design and text on a website's front page may be judged in terms of the proportion of first-time visitors who complete one or more tasks.

²² The number of participants could be provided as a raw figure or, ideally, as a percentage of people who viewed the project and went on to complete at least one task (i.e. conversion rates). Raw numbers can be misleading as some material is more accessible or appeals to a broader range of participants than others, and some projects require specialist skills or tasks that cannot be broken into microtasks.

²³ However, this may not be known without longitudinal studies (a problem also faced by museum exhibitions and public history projects).

My research seeks to understand how different attributes of heritage crowdsourcing projects contribute to these various types of success. I examined sources including published statistics, project publicity material, site copy, and community discussion. However, it is difficult for an outsider to assess many of the success metrics outlined above without access to internal statistics and communication. Consequently, the large-scale review of projects outlined in the first chapter, and the more detailed analysis of projects discussed in this and subsequent chapters have drawn on analysis of the 'user experience' of different sites, and conversations and interviews with project stakeholders. My approach to analysing sites was informed by my previous experience building and evaluating participative sites, which provided me with domain-specific expertise,²⁴ and by Nielsen and Molich's usability heuristics.²⁵ Heuristic evaluations are an established method for uncovering usability problems.²⁶ There are precedents within human-computer interaction for devising category-specific heuristics by reviewing existing products and abstracting principles to explain the underlying usability problems found.²⁷ While domain

²⁴ Including my MSc dissertation project in which I designed, built and evaluated crowdsourcing games to create metadata for museum collections, and previous work designing, building and running public-facing digital projects in museums and cultural heritage (including a prototype heritage crowdsourcing project focused on the First World War, created during a short Fellowship with the CENDARI project at Trinity College Dublin).

²⁵ Nielsen, '10 Usability Heuristics for User Interface Design'. Jakob Nielsen and Rolf Molich, 'Heuristic Evaluation of User Interfaces', in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '90 (SIGCHI Conference on Human Factors in Computing Systems, New York, NY, USA: ACM, 1990), 249–256, doi:10.1145/97243.97281.

²⁶ Joseph S. Dumas and Janice C. Redish, *A Practical Guide to Usability Testing*, Revised edition (Chicago: University of Chicago Press, 1999), <http://www.press.uchicago.edu/ucp/books/book/distributed/P/b05545036.html>. pp. 77–82.

²⁷ Dykstra, D. J., 1993, cited in Jakob Nielsen, 'How to Conduct a Heuristic Evaluation', *Nielsen Norman Group*, January 1995, <http://www.nngroup.com/articles/how-to-conduct-a-heuristic-evaluation/>.

expertise helps find more usability issues,²⁸ heuristic inspections are usually more effective when performed by multiple people than by a single reviewer.²⁹ As multiple reviewers were not possible in this project, I supplemented my evaluation with analysis of comments posted by site participants on community platforms and on social media. As passionate users of these platforms (and, in the case of some experienced participants, as users of multiple platforms), crowdsourcing participants can give valuable insights into the usability and performance of various websites and platforms. However, as always, both the users who post on discussion platforms and the types of problems they identify may not be representative of the project as a whole; while the issues they raise are useful pointers for analysis, they cannot be regarded as comprehensive or exhaustive.³⁰ Reviewing multiple sites offering similar functionality supported the process of developing category-specific heuristics (or diagnostic attributes) for crowdsourcing projects; these are discussed below.

Nielsen's usability heuristics contain many principles relevant to crowdsourcing projects, including: keeping users informed of the system status through appropriate feedback (e.g. whether input has been saved); speaking the users' language; preventing errors; supporting recovery from error when errors do occur; following platform conventions; minimising memory load by making actions and options visible; and (where necessary) providing concrete instructions that focus on the users' task.³¹ In heuristic evaluations, the evaluator

²⁸ Nielsen found that 'double usability specialists' - those who also had experience with the particular kind of interface being evaluated found 60% of problems with an interface, while 'regular' usability specialists found 41% of problems. Nielsen, 'Finding Usability Problems through Heuristic Evaluation'.

²⁹ Nielsen, 'How to Conduct a Heuristic Evaluation'.

³⁰ Furthermore, any methods that rely on the input of people who decided to participate in a project risk a bias towards those who were not deterred by any barriers to participation.

³¹ Nielsen, '10 Usability Heuristics for User Interface Design'.

'inspects' the interface, using heuristics such as those summarised above as prompts.³²

During my inspections, undertaking the core crowdsourcing tasks of the site provided a representative 'scenario' that mimicked that of typical participants. When viewing live sites, I focused on key moments such as participant recruitment and the process of completing the first task. I noted the presence and type of participant communities and of appeals to different types of initial and on-going motivation. When reviewing projects that seemed to be failing to meet their goals, I paid particular attention to barriers to participation and factors that might demotivate participants.

The motivations of crowdsourcing participants

Recruiting and retaining participants are possibly the single most important factors in the success of any participatory project. Therefore, understanding the range of motivations that may apply to any one project is vital for encouraging initial and on-going participation. The sight of empty, unused comment boxes and tag fields on heavily used library, archive and museum catalogues is an indication that merely offering functionality like commenting or tagging is not enough to encourage participation.³³ To understand participant motivations I drew on research from related fields including citizen science, cultural heritage volunteering,³⁴ commercial crowdsourcing, contributions to *Wikipedia* and open source software, and the emerging literature on cultural heritage crowdsourcing. This literature provides useful insights into participation in heritage crowdsourcing but

³² Nielsen, 'How to Conduct a Heuristic Evaluation'.

³³ These un-tagged catalogue entries also speak to the need for the consequences of any user action to have predictable outcomes (particularly any public action), and minimal barriers to action.

³⁴ In 1998 it was reported that volunteers contribute 'more than 17 million hours of labor' in US archives, a figure that presumably pre-dated most digital opportunities. Reported in Victoria Irons Walch, 'Where History Begins. A Report on Historical Records Repositories in the United States' (Council of State Historical Records Coordinators, May 1998). See also Kirsten Holmes, 'Volunteers in the Heritage Sector: A Neglected Audience?', *International Journal of Heritage Studies* 9, no. 4 (December 2003): 341–355, doi:10.1080/1352725022000155072.

must be read with care, as conclusions cannot always be transferred directly to a different disciplinary or cultural context. For example, participation in open source software development provides career benefits by directly demonstrating that participants have the programming skills desired by recruiters; the same is not true for citizen scientists.

Research into volunteering by psychologists Clary et al. found six groups of motivations for volunteers: values ('altruistic and humanitarian concerns for others'), understanding (new learning experiences and the chance to practice knowledge, skills and abilities), social 'relationships with others', career-related benefits, ego-protective ('eliminating negative aspects surrounding the ego'), and enhancement (positive strivings for growth and development).³⁵ Research on contributors to open source software³⁶ found that 'enjoyment-related intrinsic motivations' linked to tasks that enabled 'a sense of creativity' were key.³⁷ Commercial crowdsourcing researcher Brabham reported on several studies that found the primary motivator of participation in open source projects was 'the pleasure found in doing hobbies'.³⁸ Quinn and Bederson's 2011 review of the field of human computation found that key motivations were pay, altruism, enjoyment, reputation and

³⁵ E. Gil Clary et al., 'Understanding and Assessing the Motivations of Volunteers: A Functional Approach', *Journal of Personality and Social Psychology* 74, no. 6 (June 1998): 1516–30, <http://www.ncbi.nlm.nih.gov/pubmed/9654757>.

³⁶ As a technologically-enabled form of voluntary collaboration, open source software has been seen as a useful body of research that addresses some issues relevant to non-commercial crowdsourcing

³⁷ Karim Lakhani and Robert Wolf, 'Why Hackers Do What They Do: Understanding Motivation and Effort in Free/Open Source Software Projects', in *Perspectives on Free and Open Source Software*, ed. Joseph Feller et al. (MIT Press, 2005), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=443040.

³⁸ Daren C. Brabham, 'Moving the Crowd at IStockphoto: The Composition of the Crowd and Motivations for Participation in a Crowdsourcing Application', *First Monday*, 2 June 2008, <http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2159/1969>.

implicit work.³⁹ Zooniverse projects have made a substantial contribution to research on motivations in citizen science. *Galaxy Zoo* participants tend to have a range of motivations but when asked to choose their main motivation, nearly 40% selected 'I am excited to contribute to original scientific research'.⁴⁰ The next most common primary motivations⁴¹ were: 'I am interested in astronomy' (12.4%); 'I can look at galaxies that few people have seen before' (10.4%); 'I enjoy looking at the beautiful galaxy images' (8.9%); 'I am amazed by the vast scale of the universe' (8.3%); and 'I am interested in science' (6.8%).⁴² Research with museum volunteers found that 'doing something enjoyable', an interest in the subject, meeting people and 'making friends' were the main reasons for volunteering.⁴³ Csikszentmihalyi and Hermanson make a strong case for learning as an intrinsic motivation, as does game designer and theorist Koster, who claimed that 'learning is the drug' that makes games fun, motivating us to learn through play.⁴⁴ Csikszentmihalyi's theory of 'flow'⁴⁵ posits that deep engagement in a task can be its own reward, and Cacioppo et al. found that some people ('chronic cognizers') are highly intrinsically motivated to engage in cognitive activity.⁴⁶

³⁹ Implicit work happens when the desired output is a side-effect of other user activity. Quinn and Bederson, 'Human Computation'.

⁴⁰ The statements given after the name used for each motivation are those given in the survey. Jordan Raddick et al., 'Galaxy Zoo'. The survey did not define 'contribute' but it may have been read as making 'a contribution to science' rather than 'undertaking scientific research'.

⁴¹ Each was chosen by over 5% of respondents

⁴² Jordan Raddick et al., 'Galaxy Zoo'.

⁴³ Deborah Edwards and Margaret Graham, 'Museum Volunteers and Heritage Sectors', *Australian Journal on Volunteering* 11, no. 1 (2006): 19–27.

⁴⁴ Raph Koster, *A Theory of Fun for Game Design* (Scottsdale, AZ: Paraglyph Press, 2005). p. 40.

⁴⁵ Mihaly Csikszentmihalyi, *Flow: The Psychology of Optimal Experience* (New York: Harper and Row, 1990).

⁴⁶ Conversely, those with low intrinsic motivation to engage in effortful thought are called 'chronic cognitive misers' and there is presumably very little point in trying to coax them into crowdsourcing. John T. Cacioppo et al., 'Dispositional Differences in Cognitive Motivation: The Life

There is a growing collection of research into motivations relevant to specific heritage crowdsourcing projects. Oomen et al reported that taggers on the video project *Waisda?* were motivated by altruism and interest in the television clips shown.⁴⁷ Some of the earliest users of the National Library of Australia's *Trove* OCRed newspaper correction system reported finding correcting text 'both addictive and rewarding'.⁴⁸ A more recent study by Alam and Campbell found that OCR text correctors in *Trove* site reported 'high but sustainable levels of self-motivation' stemming from a range of 'egoism-based reasons' including personal interest and a response to the trust placed in them by the library.⁴⁹ Some participants also reported a sense of 'addiction' to correcting text, a feeling of obligation stemming from gratitude for the library's provision of the resource, interest in the topics covered by the newspapers, the challenge created as new content was added, an interest in the topic, learning, the collegial and supportive environment on the site and its forum, and competition with themselves or with others on a leaderboard.⁵⁰ Community and enjoyment-based motivations including fun, the simplicity of the task, autonomy in task selection and passing time were important for some. Finally, the researchers found four non-monetary extrinsic motivations for participation: attribution, recognition and rewards, indirect feedback, and advocacy (or contributing to a greater good).⁵¹ While this is a single case study, it does provide a sense of the range of motivations a single project

and Times of Individuals Varying in Need for Cognition', *Psychological Bulletin* 119, no. 2 (1996): 197-253.

⁴⁷ Johan Oomen, Riste Gligorov, and Michiel Hildebrand, 'Waisda?: Making Videos Findable through Crowdsourced Annotations', in *Crowdsourcing Our Cultural Heritage*, ed. Mia Ridge (Farnham, Surrey, UK: Ashgate, 2014), <http://www.ashgate.com/isbn/9781472410221>.

⁴⁸ Holley, 'How Good Can It Get?'

⁴⁹ Alam and Campbell, 'Crowdsourcing Motivations in a Not-for-Profit GLAM Context'.

⁵⁰ Leaderboards are tables that display an ordered list of participants who have the highest score based on some quantifiable metric such as the number of tasks completed.

⁵¹ Alam and Campbell, 'Crowdsourcing Motivations in a Not-for-Profit GLAM Context'.

can support. In heritage crowdsourcing, access to source material appears to be a meaningful reward and powerful motivation for participation. *Old Weather* transcribers who have completed the most pages of a given ship's logs are awarded the rank of 'captain'. One participant explained that it's a 'huge motivator' - not only for the status, but also because the captain is able to see every page of that ship's logs: 'You get it first and get to post exciting things from it on the forum, it gets horribly competitive because of that'.⁵²

Some extrinsic motivations are financial, as with archive projects like the Dutch *WieWasWie*⁵³ and the National Archives of Australia's *arcHIVE*⁵⁴ that offer vouchers exchangeable for copies of archival documents as a reward for transcriptions.⁵⁵ The *18thConnect* site offers people who have corrected OCR through their *TypeWright* tool a copy of the text they helped correct,⁵⁶ and *Ancestry* offer 'active contributors'⁵⁷ access to original images and the ability to vote on which records are indexed.⁵⁸ Some activities are extrinsically motivated because the crowdsourced content is a side effect of game-play or tasks undertaken for other reasons. Other extrinsic motivations are informational. While

⁵² Various, 'AHRC Crowd Sourcing Study: Scoping Seminar'.

⁵³ 'WieWasWie Project Informatie', *VeleHanden*, accessed 1 August 2014, http://velehanden.nl/projecten/bekijk/details/project/wiewaswie_bvr.

⁵⁴ National Archives of Australia, 'ArchIVE – Homepage', *ArchIVE*, accessed 18 June 2014, <http://transcribe.naa.gov.au/>.

⁵⁵ As I was not following this line of enquiry I have not sought research into the effectiveness or impact of this design decision. However, I would note that rewards with monetary value (or easily convertible into monetary value) may demotivate participants and risk comparisons with the work required to earn that compensatory value.

⁵⁶ Presumably the result of negotiations with the commercial companies ProQuest and Gale that own the digitised collections. 'What Is 18thConnect?'

⁵⁷ Those who have indexed more than 900 documents within the past three months.

⁵⁸ 'About the Ancestry.Com World Archives Project'.

most respondents to a 2011 survey of participants in *Transcribe Bentham* reported taking part because of their interests in 'Bentham's life and thought', 'history and philosophy', altruism, or 'taking part in something which will ultimately benefit the wider community', some were interested in 'crowdsourcing and the technology behind the project'.⁵⁹

Similarly, of those 74% of users who gave a reason for requesting an account for *Papers of the War Department (PWD)*, 8% were registering because they were interested in how *Scripto* (the transcription platform) and the overall transcription process worked.⁶⁰

Motivations for participation can change over time. Holmes' research with museum volunteers found that while they may have initially volunteered in order to pursue a particular interest, they continued to volunteer for their colleagues, social opportunities and general 'enjoyment/recreation'.⁶¹ Rotman et al found that participants in citizen science projects were initially interested in opportunities for education, 'individual gain and personal interest', but drew on memories of the 'ongoing appreciation and acknowledgment of volunteers' by scientists when deciding whether to continue their participation in a project.⁶² Alam and Campbell also found that motivations changed over time, though less markedly. Again, motivations like 'recognition and rewards'⁶³ had an

⁵⁹ Causer and Terras, "Many Hands Make Light Work..." p. 67.

⁶⁰ 34% of *PWD* participants were searching for material on a specific place, event or person, with a further 29% 'engaged in genealogical research'; the final group interested in the subject of the source material had a 'general interest in the American Revolution and the early national period'. Teachers and students made up an additional 6% while 10% participated out of a sense of 'civic duty'. Leon, 'Build, Analyse and Generalise'.

⁶¹ Holmes, 'Volunteers in the Heritage Sector'.

⁶² Rotman et al., 'Dynamic Changes in Motivation in Collaborative Citizen-Science Projects'.

⁶³ Here it is important to note Haythornthwaite's distinction between recognition ('a visible summation of contribution') and reward ('the outcome of a recognition and reputation system'). Caroline Haythornthwaite, 'Crowds and Communities: Light and Heavyweight Models of Peer Production', in *Proceedings of the 42nd Hawaii International Conference on System Sciences - 2009*,

impact on levels of long-term participation in the project as some participants moved from initial involvement to sustained involvement.⁶⁴ Crowston and Fagnot's research on motivations for 'massive virtual collaboration' outlined a 'motivational arc' that spanned stages of initial contribution, sustained (i.e. repeated) contribution and meta contribution ('contributions that structure and enable further contributions').⁶⁵ Together, these various views of changes in motivation over time suggest that different stages of participation might require different rewards.⁶⁶

For the purposes of this research, I have grouped motivations relevant to heritage crowdsourcing into extrinsic, intrinsic⁶⁷ and altruistic motivations.⁶⁸ As most projects do not support extrinsic motivations, such as tangible rewards, I have not focused on them.

2009, 1–10, http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4755627.

⁶⁴ Alam and Campbell, 'Crowdsourcing Motivations in a Not-for-Profit GLAM Context'.

⁶⁵ Kevin Crowston and Isabelle Fagnot, 'The Motivational Arc of Massive Virtual Collaboration', 2013, 1–35.

⁶⁶ However, particular care must be taken when making decisions based on studies of changes in motivation, as it is difficult (or resource-intensive) to avoid a bias towards the preferences of participants who continued to be involved with a project and who were disposed to respond to surveys and other research protocols. Various human-computer interaction theories also remind us that systems shape people as much as people shape systems, so long-term participants may have nudged the project design towards their own preferences.

⁶⁷ Intrinsic motivations are in effect when an activity is worth doing for its own sake, regardless of external rewards. Mihaly Csikszentmihalyi and Kim Hermanson, 'Intrinsic Motivation in Museums: Why Does One Want to Learn?', in *Public Institutions for Personal Learning: Establishing a Research Agenda*, ed. John Falk and Lynn D. Dierking (Washington D.C.: American Association of Museums, 1995), 66–77.

⁶⁸ I could also have followed Fugelstad et al and grouped motivations for volunteering into those that relate to the benefits for others, and those that relate to the benefits gained by the individual, but I suspect the boundaries between private and public benefit are less clear in heritage work. Paul Fugelstad et al., 'What Makes Users Rate (Share, Tag, Edit...)? Predicting Patterns of Participation in Online Communities', in *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work* (Seattle, 2012), 969–978, <http://dl.acm.org/citation.cfm?id=2145349>.

Altruistic motivations include those related to the 'collective' or greater good, 'the importance attributed to the project's goals',⁶⁹ and ideological values or principles.

Intrinsic motivations, such as fun, an interest in the subject and socialising, are inherently rewarding.⁷⁰ While source materials or the discussion around them can feed an interest in the subject, 'fun' is often dependent on design decisions. For example, Alam and Campbell reported that factors that helped make *Trove's* text correction fun included simplicity and task autonomy.⁷¹ In my review of design attributes, aspects related to altruistic (or value-led) reasons, and intrinsic motivations including recognition and feedback, enjoyment, an interest in the topic or the source materials, the chance to learn or master skills, and participation in a community, seemed particularly important in the success of a project.

Participants in heritage crowdsourcing

Participants vary widely in terms of their skills, knowledge and experience, their relationship to the project and related organisations, their interests in particular topics or tasks, their level or type of participation, and by their location, age, gender and other demographic characteristics. To pick examples from the contrasting subject areas of sciences and the arts, a *Galaxy Zoo* (GZ) survey found over '80% of respondents to the gender question self-reported as male',⁷² while nearly 70% of respondents to a *Your Paintings Tagger* (YPT) survey were women.⁷³ 65% of the *Galaxy Zoo* sample were from the

⁶⁹ Nov, Arazy, and Anderson, 'Technology-Mediated Citizen Science Participation'.

⁷⁰ Following Csikszentmihalyi and Hermanson, 'Intrinsic Motivation in Museums'.

⁷¹ That is, participants had a high level of freedom to choose what they corrected as well as when and how they participated in text correction. Alam and Campbell, 'Crowdsourcing Motivations in a Not-for-Profit GLAM Context'.

⁷² Jordan Raddick et al., 'Galaxy Zoo'.

⁷³ Kathryn Eccles and Andrew Greg, 'Your Paintings Tagger: Crowdsourcing Descriptive Metadata for a National Virtual Collection', in *Crowdsourcing Our Cultural Heritage*, ed. Mia Ridge (Farnham, Surrey, UK: Ashgate, 2014), <http://www.ashgate.com/isbn/9781472410221>.

US or UK; the rest were from 116 other countries or territories. Participants in both GZ and YPT tend to be well educated; almost 70% of American GZ respondents aged over 25 had a bachelor's degree or higher,⁷⁴ and nearly 80% of YPT respondents were educated to at least degree level.⁷⁵ While it may seem too obvious to mention, the two most common characteristics of crowdsourcing participants are free time and access to a computer (or other tablet/mobile device) with internet connectivity.⁷⁶

Most crowdsourcing projects report that up to 80-90% of the work is done by 10% of participants, while many other participants contribute a small amount each. Projects or summaries mentioning this 'power law' include *Red een Portret*,⁷⁷ *Old Weather*,⁷⁸ *Papers of the War Department*,⁷⁹ *Waisda?*,⁸⁰ *steve.museum*,⁸¹ *Your Paintings Tagger*,⁸² *Galaxy Zoo*

⁷⁴ Jordan Raddick et al., 'Galaxy Zoo'.

⁷⁵ 80% of YPT respondents self-report visiting art galleries 'at least every few months' and 28% work or volunteer in the 'art world' or museums, further suggesting the role an interest in the topic plays. Eccles and Greg, 'Your Paintings Tagger'.

⁷⁶ Demographic data from a wider range of projects and further research into the relationship between participation, leisure time in retirement, formal education and access to the internet would be useful.

⁷⁷ Noordegraaf, Bartholomew, and Eveleigh, 'Modeling Crowdsourcing for Cultural Heritage'.

⁷⁸ From October 2010 to July 2012, '94% of participants contributed in aggregate 15% of project input'. Alexandra Eveleigh et al., 'Designing for Dabblers and Deterring Drop-Outs in Citizen Science' (ACM Press, 2014), 2985-94, doi:10.1145/2556288.2557262. See also Philip Brohan, 'There's a Green One and a Pink One and a Blue One and a Yellow One', *Old Weather Blog*, 5 September 2012, <http://blog.oldweather.org/2012/09/05/theres-a-green-one-and-a-pink-one-and-a-blue-one-and-a-yellow-one/>.

⁷⁹ Leon, 'Build, Analyse and Generalise'. p. 101

⁸⁰ Johan Oomen et al., 'Emerging Institutional Practices: Reflections on Crowdsourcing and Collaborative Storytelling', in *Museums and the Web 2010: Proceedings*, ed. Jennifer Trant and David Bearman (Museums and the Web 2010, Archives & Museum Informatics, 2010), <http://www.museumsandtheweb.com/mw2010/papers/oomen/oomen.html>.

⁸¹ Trant, 'Tagging, Folksonomy and Art Museums: Results of Steve.Museum's Research'.

⁸² Eccles and Greg, 'Your Paintings Tagger'.

projects⁸³ and the Library of Congress on *Flickr Commons*.⁸⁴ Participants who contribute the majority of work may be called 'super-contributors' or variants thereof. For example, *Transcribe Bentham*'s success is in large part due to 17 super-transcribers.⁸⁵ Holley reported that some super-volunteers on projects like *Trove* were working as if the project was their full-time job.⁸⁶ Participation rates in other activities presumably also vary, but there is less research on activity levels for discussion, moderation and other tasks.⁸⁷ Given the role super-contributors play in a projects' productivity, it could be tempting to optimise designs for their needs. However, projects must cater for both casual and super contributors, as well as for those in-between who may become super-contributors at some point. Discussion with *Old Weather* participants suggests that putting super-volunteers in competition with others may be more detrimental than positive. Discussing the status and access to material gained when a 'captain' of a ship, one said, 'you get terribly paranoid, hoping and praying that [a specific super-volunteer] won't join your ship'.⁸⁸ Further research to understand the impact of super-volunteers on the participation of other contributors would be useful.

Terms such as 'nichesourcing' or 'targeted crowdsourcing' may describe the number of

⁸³ For citizen science examples such as Planet Hunter, see Mugar et al., 'Planet Hunters and Seafloor Explorers'.

⁸⁴ Springer et al., 'For the Common Good: The Library of Congress Flickr Pilot Project'.

⁸⁵ Causer and Terras, "Many Hands Make Light Work..." p. 73

⁸⁶ Holley, 'Crowdsourcing: How and Why Should Libraries Do It?'

⁸⁷ However, one paper surveying several *Zooniverse* projects found that a tiny percentage of those who contributed 'were responsible for over two thirds of the total number of Talk entries, and one third of the total Tasks performed'. Markus Luczak-Roesch et al., 'Why Won't Aliens Talk to Us? Content and Community Dynamics in Online Citizen Science', in *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media*, 2014, <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/download/8092/8136>.

⁸⁸ Various, 'AHRC Crowd Sourcing Study: Scoping Seminar'.

participants, their relationship to the organisation, and how they were recruited or vetted. These terms additionally recognise that the more specialised the skills, knowledge or equipment required to complete tasks, the smaller the potential pool of project participants becomes.⁸⁹ Some projects allow participants to winnow naturally according to their skills and abilities following an open call, while other projects vet potential participants through registration approval processes. *Transcribe Bentham* investigators have described the process by which a smaller group of self-selected individuals who are willing and able to take on the complex project task emerge from those who answered the original open call as 'self-sifting'.⁹⁰ Projects that vet participants may do so because they are less trusting of 'amateurs', unaware of systems for validating contributed content, have tasks which cannot easily be computationally checked, or tasks they feel will be too difficult for the general public. The *Medici Archive Project* has a deliberate 'gatekeeping' policy for their community-sourcing project, as they believe it requires specialist palaeographic, historical and linguistic skills that are only to be found in 'high-level academic researchers'.⁹¹ They also feel gatekeeping will help reassure other academics that those transcribing, contextualising and disambiguating text have sufficient expertise.⁹² *Suda On Line* approves almost everyone who requests authorisation, perhaps because their requirement to nominate a specific entry to work on weeds out spam bots and the

⁸⁹ This also makes the process of finding and recruiting potential contributors, and of putting the right content under the nose of the right expert more difficult. Previously discussed in Ridge, 'Playing with Difficult Objects: Game Designs for Crowdsourcing Museum Metadata'.

⁹⁰ Causer and Terras, "Many Hands Make Light Work..." pp. 73-4.

⁹¹ Lorenzo Allori and Lisa Kaborycha, 'Opening Aladdin's Cave or Pandora's Box? The Challenges of Crowdsourcing the Medici Archives' (Digital Humanities 2013, Lincoln, Nebraska, 2013), <http://dh2013.unl.edu/abstracts/ab-312.html>.

⁹² However, this is not the only reason for this restriction of access to the community - it is also designed to reduce resistance to sharing unpublished work and to inhibit any transgressions of academic norms around citing others' work. Allori and Kaborycha, 'Opening Aladdin's Cave or Pandora's Box? The Challenges of Crowdsourcing the Medici Archives'.

unqualified.⁹³ A Dutch research project called *Accurator* aims to gather specialist tags by computationally identifying experts posting on niche topics on social media,⁹⁴ and then inviting them to annotate images from museum collections.⁹⁵

Many of these semi-closed projects rely on academic qualifications as markers of expertise, but in future the need for specific skills may be less of a barrier to participation. Some projects are attempting to expand the pool of potential participants by teaching the public the skills required to contribute. The Folger Shakespeare Library is planning to teach participants how to read the secretary hand used on early modern manuscripts through online tutorials,⁹⁶ and the *Marine Lives* (ML) project has provided extensive resources and personal support for learning palaeography for participants who have committed to the project.⁹⁷ ML's Colin Greenstreet believes this learning can be best achieved through small groups of volunteers working with a facilitator.⁹⁸

⁹³ Mahoney, 'Tachypaedia Byzantina'.

⁹⁴ Chris Dijkshoorn et al., 'Personalized Nichesourcing: Acquisition of Qualitative Annotations from Niche Communities', in *Proc. 6th Workshop on Personalized Access to Cultural Heritage - PATCH'13* (Rome, 2013), <http://www.few.vu.nl/~snaz10/pubs/paper4.pdf>. The project aims to explore 'nichesourcing' as an extension of crowdsourcing in which 'complex tasks are distributed amongst a small crowd of amateur experts'. de Boer et al., 'Nichesourcing'. However, in May 2015, links from the demonstration site fail and the status of the project is uncertain.

⁹⁵ For example, experts might be able to provide the specific species of a bird or vegetable depicted in a painting. Rob Blaauboer, 'Nichesourcing helpt het Rijksmuseum collecties in kaart brengen', *Frankwatching*, 20 March 2014, <http://www.frankwatching.com/archive/2014/03/20/nichesourcing-helpt-het-rijksmuseum-collecties-kaart-brengen/>.

⁹⁶ Personal communication, August 2014, and 'Early Modern Manuscripts Online (EMMO) - Folgerpedia', July 2014, [http://folgerpedia.folger.edu/Early_Modern_Manuscripts_Online_\(EMMO\)](http://folgerpedia.folger.edu/Early_Modern_Manuscripts_Online_(EMMO)).

⁹⁷ Colin Greenstreet, 'Progress Review Two', *The Shipping News*, 21 January 2013, <http://marinelives-theshippingnews.org/blog/2013/01/21/progress-review-two/>.

⁹⁸ Colin Greenstreet and Jill Wilcox, 'C17th London as Seen in the Archives of the High Court of Admiralty' (Docklands History Group, Museum of London Docklands, 3 September 2014), https://www.academia.edu/8188588/Text_and_Thumbnail_slides_Dockland_History_Group_Talk_

Design factors in the success of heritage crowdsourcing projects

In this section I discuss design factors that seem particularly important for heritage crowdsourcing projects. Basic usability (minimising dissatisfaction) is rarely enough; websites should both minimise annoyances for users and offer pleasing features that encourage users to return.⁹⁹ As crowdsourcing is a voluntary activity, it is vital to minimise barriers to participation, points of friction and demotivators. The quality of the user experience can have a positive effect in convincing visitors to try undertaking a site's tasks. For example, previous human-computer interaction research investigating how people join and become active on social media found that potential contributors looked for the ease of making small contributions, the visibility of their contributions, and how they would be recognised for the quality and quantity of their contributions.¹⁰⁰

Crowdsourcing projects are framed as sites where one goes to respond to requests for participation. It can be difficult to isolate the effects of this framing, and the effort put into marketing and participant recruitment from the design of the interface itself, but one example may illustrate the difference made by the combination of the three. The Biodiversity Heritage Library reported that their images¹⁰¹ had been viewed millions of

London_Dockland_Museum_September_3rd_2014.

⁹⁹ Zhang and Von Drang offer the useful labels 'satisfiers' and 'dissatisfiers' for website design attributes that either contribute to or diminish user satisfaction. Ping Zhang and Gisela M. von Dran, 'Satisfiers and Dissatisfiers: A Two-Factor Model for Website Design and Evaluation', *Journal of the American Society for Information Science* 51, no. 14 (2000): 1253–1268, [http://onlinelibrary.wiley.com/doi/10.1002/1097-4571\(2000\)9999:9999%3C::AID-ASI1039%3E3.o.CO;2-O/full](http://onlinelibrary.wiley.com/doi/10.1002/1097-4571(2000)9999:9999%3C::AID-ASI1039%3E3.o.CO;2-O/full).

¹⁰⁰ Jennifer Preece and Ben Shneiderman, 'The Reader-to-Leader Framework: Motivating Technology-Mediated Social Participation', *AIS Transactions on Human-Computer Interaction* 1, no. 1 (2009): 13–32, <http://www.cs.umd.edu/~ben/papers/Jennifer2009Reader.pdf>.

¹⁰¹ At the time of writing their *Flickr* account has nearly 100,000 images but it is not clear when they were all uploaded. <https://www.flickr.com/people/biodivlibrary/>. Last accessed 6 June 2015. It is

times on *Flickr*, and that 18% of those images had one or more user-contributed tags.¹⁰²

However, when posted on the crowdsourcing site, *Science Gossip*, 140,000 images were classified within one week.¹⁰³

The quality of the request for help can also have an effect at a micro level. Some projects are designed as repositories first, and crowdsourcing interfaces second. These projects have opportunities to recruit viewers of repository content as participants. A comparison of two newspaper sites may be illuminating. The *California Digital Newspaper Collection* (CDNC)¹⁰⁴ and *Trove* both offer links to correct OCR'd text from their digitised newspaper viewers. However, the 'Correct this text' link on the CDNC site is comparatively subtle - the colour matches the heading, the link is not underlined, and it is placed at the top of the transcribed article. The 'Fix this text' link on *Trove* pops up whenever the cursor is over the transcribed text, contains an image of a pencil, has underlined text and is in a brighter colour than the rest of the page. Together, it is easy to miss the option to correct text on CDNC and hard to ignore it on *Trove*.¹⁰⁵

also difficult to assess 'completion' rates on *Science Gossip* as they may load new material as earlier tranches are completed.

¹⁰² Trish Rose-Sandler, 'Crowdsourcing Your Cultural Heritage Collections: Considerations When Choosing a Platform' (Visual Resources Association conference, Denver, 13 March 2015), <http://www.slideshare.net/trosesandler/crowdsourcing-your-metadata>.

¹⁰³ Rose-Sandler, 'Crowdsourcing Your Cultural Heritage Collections'. See also Trish Rose-Sandler, 'Citizen Science Uses Art to Unlock Scientific Knowledge', *Biodiversity Heritage Library*, 31 March 2015, <http://blog.biodiversitylibrary.org/2015/03/citizen-science-uses-art-to-unlock.html>.

¹⁰⁴ Based on software by Veridian.

¹⁰⁵ While factors such as the availability of other sources and the effect of formal and word-of-mouth publicity make it impossible to directly compare the effect of interface decisions, the California Digital Newspaper Collection (CDNC) has relatively fewer registered users ('more than 2,000') and corrections (more than 2.5 million errors) in three years. Three years after launch, *Trove* had over 62,7000 registered users, 2,700 of whom were active in the last month, with a cumulative total of nearly 45 million lines corrected. Veridian, 'Crowdsourced User Text Correction (UTC)'.

Understanding participant motivations is important for designing appropriate user experiences that support people's reasons for taking part in a project. After a brief discussion of designing for participant motivations, I present design factors that may influence the success of a crowdsourcing project. These factors need empirical testing for validation, but nonetheless they suggest useful areas for focusing further research. They are organised to match the broad stages of hearing about a project, viewing it for the first time, trying out its tasks, learning more about the tasks and material, and communicating with other participants.

Designing for motivations

Finding appropriate rewards means balancing complex requirements. As previously discussed, matching rewards to motivations is important, but other factors also have an impact on the participants' experience. For example, competitive models like leaderboards are an easy way to recognise individuals who have completed more tasks, but they favour those with more free time, and there is some evidence that some participants are deterred by competition.¹⁰⁶ This suggests that any competitive reward scheme should be available to those who seek it, but should not be forced onto participants who do not find it rewarding.

Rewarding or emphasising certain activities will often cause participants to change their behaviour to focus on those activities; this may come at the expense of other tasks that

¹⁰⁶ Chris Preist, Elaine Massung, and David Coyle, 'Competing or Aiming to Be Average?: Normification as a Means of Engaging Digital Volunteers', in *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing* (CSCW 2014 - Computer Supported Cooperative Work and Social Computing, Baltimore, MD, USA.: ACM Press, 2014), 1222–33, doi:10.1145/2531602.2531615.

may be more important to the project but harder to quantify. Ben Brumfield reports that a crowdsourced project aimed at indexing species mentioned within a naturalist's field notes accidentally changed participant behaviour when it added lists of the top ten transcribers and editors: 'the minute this went up, the volunteers who previously had been transcribing and indexing every single page stopped indexing completely' in order to complete pages more quickly.¹⁰⁷ Emphasising raw numbers for tasks that can be completed to different levels of detail rewards those who complete pages quickly rather than those who take the time to record other items of interest. For example, pages of ships' logs in *Old Weather* (OW) can be submitted once the required quantitative data has been added, but participants can also note unusual incidents, transcribe names, etc., which takes more time but creates a richer dataset. A thread on the *Operation War Diary* forum provides an insight into the range of (self-reported) times taken to index diary pages, from '28 pages in an hour or so' to 'about an hour over each page' for a participant who took time to research place names and activities mentioned.¹⁰⁸ In short, any visible metrics must be carefully chosen to reward the desired behaviour on a project.

Public recognition for contributions is important, and can be built into many points of the project interface and communications. *Trove* newspaper pages show the number of corrections and the name of the most recent corrector. The link to 'Show corrections' on an article generates a page that shows the improvement in a page over time, with blocks of corrections attributed to named individuals. Some projects name contributors in project

¹⁰⁷ Ben W. Brumfield, 'Crowdsourcing at IMLS WebWise 2012', *Collaborative Manuscript Transcription*, 17 March 2012, <http://manuscripttranscription.blogspot.com.au/2012/03/crowdsourcing-at-imls-webwise-2012.html>.

¹⁰⁸ Forum posters, '2nd Batch of Unit War Diaries Added to Operation War Diary Today!', *Talk Operation War Diary History Board / Latest News*, March 2014, <http://talk.operationwardiary.org/#/boards/BWD0000003/discussions/DWD00006gw>.

updates¹⁰⁹ or list them as co-authors on journal articles.¹¹⁰ Describing, or even better, showing the impact of contributions on the goal of a project can be effective when the form of recognition is well chosen.

Project communications: marketing, publicity and outreach

Some content has a wider appeal, and consequently makes the work of recruiting participants easier. The New York Public Library (NYPL) chose the menu collection that became *What's on the Menu* because it had 'broad appeal to both the casual viewer and the serious scholar'.¹¹¹ *DIY History* selects handwritten, historically significant, 'interesting' and extensive materials.¹¹² They also note a preference for material that is 'old enough' to avoid copyright and privacy issues. It is easy to explain the value in name-rich sources to family and other historians. Historical material that is less immediately appealing may require more explanation through marketing messages designed to get people to a site.

Project communications that help with marketing and outreach include publicity material, posts on social media and text on project sites. Clary et al found that 'persuasive messages' (i.e., marketing) that resonate with 'the specific motivations important to individual recipients of the message' have enhanced 'persuasive impact' and help volunteers find more enjoyable and satisfying roles that match their motivations.¹¹³ For example, *Operation*

¹⁰⁹ For example, Philip Brohan, 'In Search of Lost Weather', *Old Weather Blog*, 18 August 2014, <http://blog.oldweather.org/2014/08/18/in-search-of-lost-weather/>.

¹¹⁰ For example, the authors whose affiliation is listed as 'Planet Hunter' in 'Planet Hunters. VI. An Independent Characterization of KOI-351 and Several Long', accessed 4 September 2014, <http://adsabs.harvard.edu/abs/2013arXiv1310.5912S>.

¹¹¹ Lascarides and Vershbow, 'What's on the Menu?'

¹¹² Michelle DiMeo, 'First Monday Library Chat: University of Iowa's DIY History', *The Recipes Project*, 3 February 2014, <http://recipes.hypotheses.org/3216>.

¹¹³ Clary et al., 'Understanding and Assessing the Motivations of Volunteers: A Functional Approach'. See also Fugelstad et al., 'What Makes Users Rate (Share, Tag, Edit...)?'

War Diary's front page text connects to both altruistic and intrinsic motivations by using emotive language to ask for help remembering the soldiers of the First World War, with stories 'waiting to be discovered in 1.5 million pages of unit war diaries'.¹¹⁴ Furthermore, volunteers whose experiences matched their motivations were more satisfied and more likely to intend to continue volunteering,¹¹⁵ suggesting that the text used to market and describe projects could be as important as interface and task design.

Projects like *Transcribe Bentham*,¹¹⁶ *Dickens Journals Online*¹¹⁷ and *Galaxy Zoo*¹¹⁸ demonstrate the value of mass media attention in reaching a large audience.¹¹⁹ More targeted publicity may reach a smaller number of people, but those reached may be proportionally more likely to participate. The *History Harvest* project found that outreach work should be 'tailored to the particular community or theme' of a planned collecting event.¹²⁰ *History Harvest* and *Letter in the Attic*¹²¹ found that 'face-to-face contact' at local events and groups is more effective than media attention at gaining contributions for their

¹¹⁴ Operation War Diary, 'Operation War Diary', Operation War Diary, accessed 6 June 2015, <http://www.operationwardiary.org/>.

¹¹⁵ Clary et al., 'Understanding and Assessing the Motivations of Volunteers: A Functional Approach'.

¹¹⁶ Causer and Wallace, 'Building A Volunteer Community'. Causer and Terras, "Many Hands Make Light Work..." p. 65, 67.

¹¹⁷ Sarah Crown, 'Good News! Dickens Detectives Flooding In', *The Guardian*, 5 August 2011, <http://www.theguardian.com/books/booksblog/2011/aug/05/dickens-detectives-found>.

¹¹⁸ Raddick et al., 'Galaxy Zoo: Exploring the Motivations of Citizen Science Volunteers'.

¹¹⁹ Converting people reached by marketing messages into participants relies in part on other interface factors discussed below.

¹²⁰ William G. Thomas, Patrick D. Jones, and Andrew Witmer, 'History Harvests: What Happens When Students Collect and Digitize the People's History?', *Perspectives on History*, January 2013, <https://www.historians.org/publications-and-directories/perspectives-on-history/january-2013/history-harvests>.

¹²¹ The project collected letters, diaries and items related to Brighton and Hove.

regionally-focused projects.¹²² The *Smithsonian Transcription Center's* launch post provided hooks for many different interests, which befits both their varied source data and good publicity practices.¹²³ One science project suggested finding a way to make the science 'romantic' to encourage people to want to support it.¹²⁴ It is difficult to disentangle the role of luck in getting media and popular attention but a quirky story or topic, relationships with an existing community, being the first of its type, or an opportunity to access highly-valued content or expertise seem to help.¹²⁵

A broad, open call may not be appropriate for all projects. The *Marine Lives (ML)* project aimed to find participants willing to commit time in advance. Their recruitment posts state the time commitment required (i.e. 'fifty hours of research time between September and December 2012') and the skills participants will gain (i.e. 'digital editorial, project management, semantic markup and data mining skills').¹²⁶ It seems to have been effective, with a self-reported conversion rate from 'an emailed expression of interest to a signed up

¹²² Latimer, 'Letter in the Attic: Lessons Learnt from the Project'. Face-to-face events might also help reach those not online. One First World War project heard from a potential contributor 'aged 89 and nearly blind' who had asked neighbour to email the project after hearing about it on the radio. Piers Dillon-Scott, 'How Europeana, Crowdsourcing & Wiki Principles Are Preserving European History', *The Sociable*, 31 March 2011, <http://sociable.co/business/how-europeana-crowdsourcing-wiki-principles-are-preserving-european-history/>.

¹²³ Including scientists' accounts of the importance of bumblebee data, and sections for 'art lovers', 'armchair archeologists' and 'bird lovers'. 'Volunteers Needed for Massive Smithsonian Digitization Project', *Newsdesk Newsroom of the Smithsonian*, 12 August 2014, <http://newsdesk.si.edu/releases/volunteers-needed-massive-smithsonian-digitization-project>.

¹²⁴ Trisha Gura, 'Citizen Science: Amateur Experts', *Nature* 496, no. 7444 (10 April 2013): 259–61, doi:10.1038/nj7444-259a.

¹²⁵ Future research on the relationship between task size and complexity (discussed later), and appropriate marketing and outreach techniques would be useful.

¹²⁶ Greenstreet, 'MarineLives - Call for Project Volunteers', 2012.

volunteer' of about three-to-one, with a 'relatively low' drop-out rate after starting.¹²⁷ The project also invested resources in publicising the project through specific channels, including news items in *History Today*¹²⁸ and the Institute of Historical Research website.¹²⁹ The *History Today* article is an effective combination of good storytelling and scholarly credentialing, and apparently prompted 'more than one third of the eventual thirty volunteers' into action.¹³⁰

The launch of *Operation War Diary* (OWD) shows how difficult it can be to juggle the last-minute details of launching a major technical project, media requests and simultaneously have a visible presence on the site itself.¹³¹ Comments on the OWD forum on the 'almost total lack of response from Moderators or anyone running the project on these boards today'¹³² suggest high expectations about appropriate response times, which may come as a shock to heritage institutions. Organisational models that fail to resource community management after project launch may be one reason for projects failing to thrive.

¹²⁷ Colin Greenstreet, 'Communicating MarineLives', *The Shipping News*, 24 November 2013, <http://marinelives-theshippingnews.org/blog/2013/11/24/communicating-marinelives/>.

¹²⁸ Colin Greenstreet, 'On the Crest of a Wave', *History Today*, accessed 24 May 2014, <http://www.historytoday.com/colin-greenstreet/crest-wave>.

¹²⁹ Colin Greenstreet, 'MarineLives - Call for Project Volunteers', *Institute of Historical Research*, 12 July 2012, <http://www.history.ac.uk/news/2012-08-31/marinelives-call-project-volunteers>.

¹³⁰ Greenstreet, 'Communicating MarineLives'.

¹³¹ An interesting small future research project might be to compare the first week and subsequent weeks of posts to project forums (paying special attention to weeks when traditional or social media attention may have brought a fresh influx of participants) to understand typical questions and reactions in order to better plan for required resources (technical, moderation, or scholarly) at different points.

¹³² Forum posters, '[FEATURE REQUEST] Ability to Go Back and Change Tags', *Talk Operation War Diary Help Board*, 2014, <http://talk.operationwardiary.org/#/boards/BWDoooooooo8/discussions/DWDoooooooo3>.

Once a project is up and running, marketing efforts may need to shift from participant recruitment to participant retention. Providing visible evidence of the impact of participation helps motivate continued participation. Rotman et al.'s study of motivation provides evidence for the importance of feedback to participants about how their contributions have been used.¹³³ The *What's on the Menu* project does an excellent job of updating their front page to tie-in with current affairs with menus transcribed through the project.¹³⁴ Some projects use *Twitter* hashtags like '#TranscribeTuesday' to encourage participation. However, on-going communications, whether simple quantitative progress updates, answering questions or liaising with experts to pass on information on the impact of the project, can require significant amounts of time. Finding project updates on social media for some projects in this review was not easy as they were not included or linked to in prominent locations on the main site.¹³⁵

Onboarding

In user experience design, 'onboarding' refers to orienting people to the features of a site and helping them start to use it.¹³⁶ It is particularly important for participatory sites, as convincing people to take action is harder than convincing them to look at information. Ideally, the first page that potential participants see shows (rather than tells) them what the project aims to do, how their help can make a difference, and where to start the task. The first page of *What's on the Menu* and many *Zooniverse* projects provide good examples: their titles describe the source material and sometimes the task - e.g. '*Worm*

¹³³ Rotman et al., 'Dynamic Changes in Motivation in Collaborative Citizen-Science Projects'.

¹³⁴ Including the sporting events like the Super Bowl, political events like the presidential inauguration dinner and the changing seasons.

¹³⁵ Additionally, some projects changed social media platforms without updating links across their sites.

¹³⁶ Whitney Hess, 'Onboarding: Designing Welcoming First Experiences', *UX Magazine*, 16 February 2010, <http://uxmag.com/articles/onboarding-designing-welcoming-first-experiences>.

Watch', *'Planet Hunters*', *'Notes from Nature*'. Straplines appear as project sub-titles or headings and can also help give a sense of the larger challenge that tasks will contribute to. The title and strapline ideally connect to probable motivations for action. For example, 'With a few keystrokes, you could bring a family together';¹³⁷ 'We know the names of these children; can you help us tell their stories?';¹³⁸ 'Kill Time. Make History.';¹³⁹ or simply, 'Historians need your help!'.¹⁴⁰ Together, the project title, strapline and micro-copy (small pieces of text on buttons, menu items, etc.) can encourage participation by succinctly explaining what a participant would do and why they should help. The 'call to action' (a term for the 'trigger' to act) is also important.¹⁴¹ Effective calls to action show participants how to get started. For example, *What's on the Menu* has an old-fashioned image of hands pointing to a button labelled 'Help transcribe'.

User experience design patterns, such as social proof, are starting to appear on some sites. Social proof is the idea that people look to others for cues on how to act and would, for example, be encouraged by the fact that others have already chosen to participate.¹⁴² For example, the front page of *What's on the Menu* prominently lists the number of dishes transcribed so far. *Notes from Nature* lists the number of participants and transcriptions.

¹³⁷ Ancestry's *World Archives Project*

¹³⁸ *Children of the Lodz Ghetto Research Project*

¹³⁹ *Building Inspector*

¹⁴⁰ *DIY History*

¹⁴¹ For more theory on triggering action, see B.J. Fogg, 'A Behavior Model for Persuasive Design', in *Proceedings of the 4th International Conference on Persuasive Technology - Persuasive '09* (New York, New York, USA: ACM Press, 2009), doi:10.1145/1541948.1541999.

¹⁴² Tanushree Mitra and Eric Gilbert, 'The Language That Gets People to Give: Phrases That Predict Success on Kickstarter' (CSCW, 2014), <http://comp.social.gatech.edu/papers/cscw14.crowdfunding.mitra.pdf>. See also Preist, Massung, and Coyle, 'Competing or Aiming to Be Average?' on the 'normalising' effect of displaying participant activity.

Trove lists the number of corrections already made on a given day, the number of items tagged that week, and the number of comments added that month, showing how updates can be tailored to the frequency of different tasks (a method that also allows low-frequency sites to tailor updates to best represent 'recent' activity).

Tutorials and instructions

Sites that need to provide specific information or skills may provide a step-through or interactive tutorial as part of the onboarding process. For example, the *GeoTag-X* project's interactive tutorial¹⁴³ first points out important features of the site interface. It then combines the provision of information with task questions that test that information; the feedback received is tailored to the material and the answer given. The *FamilySearch Indexing* 'test drive'¹⁴⁴ interface similarly introduces the concept of indexing while anticipating areas of uncertainty. Task previews help participants decide whether they would find the task enjoyable or not. NYPL's *Building Inspector* projects show short video tutorials that explain how and why to undertake the task; the videos also anticipate and address common questions and issues. In demonstrating the processes and information involved, previews may be particularly valuable for explaining more complex tasks. In contrast, you cannot view *HistoryPin* 'history mysteries' without registering or logging in, so it is difficult to know exactly what information and evidence they require to 'solve' a mystery. Gradual engagement¹⁴⁵ is another design pattern that allows participants to try a task while learning about the project.

¹⁴³ I tried the 'Crop identification for drought' project tutorial at <http://www.geotagx.org/project/cropid2/> in May 2015.

¹⁴⁴ Scott Flinders, 'The Future of Indexing', *FamilySearch Blog*, 1 November 2013, <https://familysearch.org/blog/en/future-indexing/>.

¹⁴⁵ Described as a combination of onboarding and marketing which aims to entice people to use a site while showing them how it works. Nathan Barry, 'A Lesson in Gradual Engagement', *UX Booth*, 18 December 2012, <http://www.uxbooth.com/articles/a-lesson-in-gradual-engagement/>.

Sites that require users to register can still provide a preview of their tasks. *Distributed Proofreaders* provide a simple but thoughtful 'walkthrough' of their correction task. *Free the Files*¹⁴⁶ used a 'glass door' - a slightly blurred image of the task screen - to convince people to sign up and contribute, claiming that it is 'hard to get to that screen and *not* want to see what's behind it'.¹⁴⁷ The *Zooniverse Chimp & See* project front page plays sample video from tasks in the background of the page.

Poorly designed tutorials can be worse than no tutorial at all. For example, showing multiple tips in a row can also make an interface appear 'overly complicated and daunting to new users'.¹⁴⁸ Research suggests that multi-screen tutorials that the user passively clicks through might have poor usability, because people must store information that cannot be applied immediately in their short-term memory.¹⁴⁹ Research on reducing the cognitive load for learners, and the techniques game designers use for including skills tests and tutorials may be relevant for projects that wish to teach specific skills or knowledge to participants undertaking new tasks.¹⁵⁰

¹⁴⁶ A project which crowdsourced the review of documents related to political ad spending.

¹⁴⁷ Al Shaw, 'No Windows. One Exit. Free Drinks: Casino-Driven Design for Crowdsourcing', *ProPublica*, 20 March 2013, <http://www.propublica.org/nerds/item/casino-driven-design>.

¹⁴⁸ According to usability experts at the Nielsen Norman Group. Aurora Bedford, 'Instructional Overlays and Coach Marks for Mobile Apps', *Nielsen Norman Group*, 16 February 2014, <http://www.nngroup.com/articles/mobile-instructional-overlay/>. For example, some *Zooniverse* tutorials contain many steps, each providing complex information that should be retained to complete tasks accurately. Further researching testing different tutorial designs would be useful.

¹⁴⁹ Bedford, 'Instructional Overlays and Coach Marks for Mobile Apps'. The possible limits of working memory - as few as 2 or 3 'novel interacting elements' is an important factor in tutorial design. Fred Paas, Alexander Renkl, and John Sweller, 'Cognitive Load Theory and Instructional Design: Recent Developments', *Educational Psychologist* 38, no. 1 (March 2003): 1-4, doi:10.1207/S15326985EP3801_1.

¹⁵⁰ While I explored this during my Masters project on crowdsourcing games, more recent work

Finally, instructions should be clear and unambiguous, and available at the point at which they are needed.¹⁵¹ They should also address 'boundary cases' and ideally provide examples of what is expected.¹⁵² An understanding of the questions participants have at different points can be gained while conducting user testing, but providing instructions for unusual, boundary cases requires a good knowledge of the material. Understanding the need for documentation from their own initial experiences, participants in *Zooniverse* projects sometimes compile 'Frequently Asked Questions' on project forums to supplement the official instructions and help pages.

Finding the first task

Some projects feed participants tasks from a queue of material, while others leave the choice of material up to the participant. Providing initial tasks from a queue minimises the number of decisions a participant has to make, which helps reduce cognitive load (the amount of mental effort required to operate a system or learn new information).¹⁵³ This, in turn, leaves more mental resources for learning the task.¹⁵⁴ Feeding the first tasks to participants also allows a project to begin with 'golden tasks' (tasks to which the answer is

should be available. See also Richard E. Mayer and Roxana Moreno, 'Nine Ways to Reduce Cognitive Load in Multimedia Learning', *Educational Psychologist* 38, no. 1 (2003): 43–52, http://www.tandfonline.com/doi/full/10.1207/S15326985EP3801_6.

¹⁵¹ Nielsen, '10 Usability Heuristics for User Interface Design'.

¹⁵² Aniket Kittur et al., 'The Future of Crowd Work', in *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, 2013, 1301–1318, <http://dl.acm.org/citation.cfm?id=2441923>.

¹⁵³ Kathryn Whitenon, 'Minimize Cognitive Load to Maximize Usability', *Nielsen Norman Group*, 22 December 2013, <http://www.nngroup.com/articles/minimize-cognitive-load/>.

¹⁵⁴ See also Paas, Renkl, and Sweller, 'Cognitive Load Theory and Instructional Design'. and Jeroen J. G. Van Merriënboer, Paul A. Kirschner, and Liesbeth Kester, 'Taking the Load off a Learner's Mind: Instructional Design for Complex Learning', *Educational Psychologist* 38, no. 1 (2003): 5–13, http://www.tandfonline.com/doi/full/10.1207/S15326985EP3801_2.

known) so they can assess the participant's performance.¹⁵⁵

Projects that let the participant decide what to transcribe face a design challenge in helping participants find records in need of work. Early repository-based projects presented the viewer with page after page of records, some of which had been completed and others which had not yet been started. Newer projects have found ways to present the records that need work first. For example, after clicking the 'Try Now!' button on *DIY History* you are automatically taken to a page to transcribe. If you choose 'transcribe by topic', the site puts the least complete items at the top of the topic page.¹⁵⁶ Some projects (for example, *Ancestry's World Archives Project*) group tasks by difficulty to help participants decide where to start. The levels of difficulty in the *Making History Searchable* project were based on the quality of the original document, the legibility of the handwriting, and the length of the document.¹⁵⁷ Projects such as *Letters of 1916* and the State Library of Queensland group material by topic,¹⁵⁸ while others group material by the required task (for example, transcription or review). The *Smithsonian Transcription Center* provides many ways for a participant to find content that they might be interested in,

¹⁵⁵ Tommaso De Benetti, 'The Secrets of Digitalkoot: Lessons Learned Crowdsourcing Data Entry to 50,000 People (for Free)', *Microtask*, 16 June 2011, <http://blog.microtask.com/2011/06/the-secrets-of-digitalkoot-lessons-learned-crowdsourcing-data-entry-to-50000-people-for-free/>.

¹⁵⁶ This feature is partly based on their experience with a previous transcription project, *Civil War Diaries and Letters*. Jen Wolfe, 'Crowdsourcing Continued', *The University of Iowa Libraries*, 16 March 2012, <http://blog.lib.uiowa.edu/drp/2012/03/16/crowdsourcing-continued/>. Shawn Averkamp and Matthew Butler, 'The Care and Feeding of a Crowd', in *Code4Lib Conference*, 2013, http://ir.uiowa.edu/lib_pubs/129.

¹⁵⁷ Dolph Briscoe Center for American History, 'Frequently Asked Questions', *Dolph Briscoe Center for American History*, accessed 28 February 2015, <http://www.cah.utexas.edu/projects/transcribe/faq>.

¹⁵⁸ State Library of Queensland, 'Text Correct', *State Library of Queensland*, accessed 12 May 2014, <http://www.slq.qld.gov.au/about-us/pitch-in/text-correct>.

including themes (such as 'Civil War Era' or 'Field Book Project'), source organisations (specific museums or archives), featured projects and those with recent activity. The *Notes from Nature* collection pages list the average time per record (ranging from 3 minutes to 15 minutes) as well as the average 'difficulty' (ranging from 'easy' to 'very hard').

Task design

Crowdsourcing is built on 'microtasks': small, self-contained actions that can be completed quickly. Some successful projects have more complex tasks, but this generally leads to lower participation rates and may require stronger calls to motivation. Recent research has found that microtasks lead to fewer mistakes and an 'easier' experience.¹⁵⁹ Task 'size' can be measured in terms of the amount of source material to process, the time per task, task modularity¹⁶⁰ and cognitive load. My research suggests that the combination of task size and its relevance to participant motivation interact to produce a task 'weight', but further research is needed to test this model.¹⁶¹ Reducing the amount of time required per task not only makes it easier for participants to find time to contribute, but other research has shown that it enables 'occasional' contributors,¹⁶² which creates a broader base of participants and increases overall productivity. For example, separate lines in a *Trove* article can be corrected by different contributors whose actions need not affect the other,

¹⁵⁹ The research compared macrotask and microtask versions of the same overall task. It also found that microtasks took more time combined than the equivalent macrotask. Justin Cheng et al., 'Break It Down: A Comparison of Macro- and Microtasks' (CHI2015, Seoul, Republic of Korea: ACM Press, 2015), 4061–64, doi:10.1145/2702123.2702146.

¹⁶⁰ Modular tasks can be 'independently and asynchronously produced'; they enable flexibility in the timing and extent of contributions. Benkler, 'Coase's Penguin'.

¹⁶¹ It seems that strong motivations reduce the perceived impact of task size on participation, in that some large, complex tasks (such as those in the *Dickens Journals Online* text correction project or *Children of the Lodz Ghetto*) can still attract participants if the motivation and/or challenge are strong enough.

¹⁶² Eveleigh et al., 'Designing for Dabblers and Deterring Drop-Outs in Citizen Science'.

but single pages of a document on *WikiSource* can only be edited by one contributor at a time without overwriting the work of another contributor. Cognitive load, or mental effort, can be minimised through the presentation of information or tasks.¹⁶³ For example, avoiding visual clutter and drawing on existing mental models in interaction design both reduce the overall cognitive load of a site.¹⁶⁴ Mental models are one's sense of how to interact with a system or beliefs about how it works, often based on analogies with other systems.¹⁶⁵ For example, following the navigation conventions used by other websites can reduce the amount of new information that participants have to absorb. 'Offloading' tasks by minimising the points at which participants have to read text, remember information or make a decision also reduces cognitive load,¹⁶⁶ as does breaking a task into smaller parts.¹⁶⁷

These design techniques can be seen in the *What's on the Menu* interface.¹⁶⁸ The interface is tightly focused on the task of transcribing menu items and prices, minimising any potential uncertainty about how to fill in the text boxes and complete the task. This focus makes the task both simple and enjoyable. *DIY History* helps reduce cognitive load by pre-populating the transcription page with any existing transcription created through OCR.¹⁶⁹ NYPL's next project, *Building Inspector*, followed the principles they established in earlier projects, with extremely focused, tiny microtasks embedded in a specialised interface.

Combining activities within a task seems to multiply the complexity or size of a task. For

¹⁶³ Paas, Renkl, and Sweller, 'Cognitive Load Theory and Instructional Design'.

¹⁶⁴ Whintont, 'Minimize Cognitive Load to Maximize Usability'.

¹⁶⁵ Sharp, Rogers, and Preece, *Interaction Design: Beyond Human-Computer Interaction*. pp. 116-9

¹⁶⁶ Whintont, 'Minimize Cognitive Load to Maximize Usability'.

¹⁶⁷ Van Merriënboer, Kirschner, and Kester, 'Taking the Load off a Learner's Mind'.

¹⁶⁸ Lascarides and Vershbow, 'What's on the Menu?'

¹⁶⁹ 'DIY History Code', *The University of Iowa Libraries*, 2012, <http://web.archive.org/web/20130218122809/http://diyhistory.lib.uiowa.edu/code.html>.

example, Tim Causer and Melissa Terras report that the requirement to transcribe and mark-up the text in *Project Bentham* appears to be an added 'aggravation'¹⁷⁰ that may have deterred participants or reduced the extent of their participation. Similarly, Doug Reside notes that while NYPL's *What's on the Menu* asks for a simple transcription, the theatre programme project *Ensemble* asks participants to examine the programme closely in order to 'identify relationships' between lines of text and determine, for example, whether a listed name is 'the playwright, the producer, or perhaps the director'.¹⁷¹ In these cases, participants must make a series of decisions, increasing the cognitive load and the time taken per task in addition to increasing potential anxiety about making the 'wrong' choice. Reside further points out that in some cases the system does not have an appropriate category to record the information on the page. Balancing the need for simplicity with the need for flexibility is a challenge for projects working with materials that may contain unexpected or inconsistent information.

Microtasks appear to play an important role in helping participants learn some of the skills required for projects with more complex tasks. For example, transcribing a word or a line of text is easier than transcribing an entire page, and offers an opportunity to practice palaeography. Microtasks can also help participants become familiar with new material. *Ancient Lives* is a *Zooniverse* project that asks participants to help transcribe ancient Greek texts on fragments of papyri. In addition to a tutorial for the character-by-character transcription task, the project also offers a measurement task, which involves marking the edge of a papyrus and the start of the block of writing (allowing margin sizes to be

¹⁷⁰ Causer and Terras, "Many Hands Make Light Work..." p. 67. See also Causer and Wallace, 'Building A Volunteer Community'.

¹⁷¹ This might also require some knowledge of the theatre industry. Doug Reside, 'Crowdsourcing Performing Arts History with NYPL's ENSEMBLE', 2014, <http://dharchive.org/paper/DH2014/Paper-131.xml>.

calculated). This exercise allows the participant to become familiar with the manuscripts. It is also possible to begin the transcription task by simply marking the centre of each character (to help map individual letters for the software). Again, this gives the participant time to start to recognise repeated characters and perhaps take on the transcription task as their skills grow. Microtasks can also be used to tempt audiences to take their first action on a site. *Waisda?* designed microtasks that led users to increase their level of activity, beginning with giving another user's tag a thumbs-up or a thumbs-down; users who press the thumbs-down button are asked to correct the label.¹⁷²

The pages in which tasks are presented can further influence participant behaviour. For example, *Typewright*'s OCR correction interface encourages the participant to stay on the page: the page does not offer any obvious navigation path away from the transcription interface,¹⁷³ and their transcription screen includes a preview of the next line of text in the document. This gives the participant a sense of how well the next line has been OCREd, feeding the impulse to correct 'just one more' line.

While quality control processes to check the results of crowdsourced processes are necessary,¹⁷⁴ they did not seem to have a relationship with the success or failure of an

¹⁷² Cees G. M. Snoek et al., 'Crowdsourcing Rock N' Roll Multimedia Retrieval', in *Proceedings of the International Conference on Multimedia* (New York, NY, USA: ACM, 2010), 1535–38, <http://doi.acm.org/10.1145/1873951.1874278>.

¹⁷³ These design features also occur in 'casino-driven design', which aims to eliminate distractions by 'actively discouraging cross-site exploration and page exits' and uses rewards to keep people working on the task, but this may be a coincidence. The increasing use of consumer psychology in interface design does raise interesting ethical issues about 'free choice' and participant agency. Shaw, 'No Windows. One Exit. Free Drinks: Casino-Driven Design for Crowdsourcing'.

¹⁷⁴ Most projects judge the quality of the contribution; manually in cases where the task is subjective and computationally where it is more straightforward. *Zooniverse* projects will assess the number of participants required for reliable data and adjust it as necessary, as described by a *Zooniverse*

interface, except when participants felt their efforts had been 'wasted' upon discovering that each input item is processed multiple times. Conversely, some participants are reassured by the knowledge that they are not the only ones to process an item.¹⁷⁵

Task feedback

Nielsen's usability heuristic of providing information on the state of a system¹⁷⁶ is particularly important for crowdsourcing tasks. Providing feedback on the first task (for example, with text or design elements showing that the data has been received and is valued) may be particularly important. Interfaces should respond to participant actions by making it clear that actions on a task have been received and accepted by the system. Some interfaces do not quite meet this requirement, which may cause participants to feel uncertain about the status of their task. For instance, on the *Letters of 1916's Scripto* site, there is only a small difference on-screen when transcribed text has been saved. The save button says 'Save edits' before saving, and 'Edit transcription' after saving transcriptions to the system; this small difference is the only sign that changes have been saved, and it is

developer posting as ttfnrob in Forum posters, '2nd Batch of Unit War Diaries Added to Operation War Diary Today!' See also Quinn and Bederson, 'Human Computation'. Derek L. Hansen et al., 'Quality Control Mechanisms for Crowdsourcing: Peer Review, Arbitration, & Expertise at FamilySearch Indexing', in *Computer Supported Cooperative Work and Social Computing (CSCW)* (San Antonio: ACM, 2013). Ben W. Brumfield, 'Quality Control for Crowdsourced Transcription', *Collaborative Manuscript Transcription*, 5 March 2012, <http://manuscripttranscription.blogspot.co.uk/2012/03/quality-control-for-crowdsourced.html>.

¹⁷⁵ For example, statements on an *OWD* forum thread range from 'I was a bit shocked to think I had spent hours repeating work already completed by someone else' and 'there's an excellent chance your work is irrelevant' to 'I know at least one of the other 4 or so "taggers" will enter the correct detail. [...] it's reassuring to know I'm not alone on the page/diary I'm working on'. Posts in Forum posters, '2nd Batch of Unit War Diaries Added to Operation War Diary Today!' Similar conversations about double-keyed transcription occur in the *FreeBMD* discussion list archives.

¹⁷⁶ Nielsen, '10 Usability Heuristics for User Interface Design'.

easily missed.¹⁷⁷ The *Louisville Leader Transcription Project* version of *Scripto* has been adapted to display the following message when a transcription is saved: 'Thank you for participating in the Louisville Leader transcription project! The article you transcribed has been sent to our Digital Initiatives Librarian'. This message not only explicitly thanks the participant; it also lets them know that their work has been received by the organisation. The combination of feedback and a clear goal helps create conditions in which participants can experience 'flow', which in turn motivates further participation.¹⁷⁸

Feedback is also important for another of Nielsen's heuristics - helping users recognise, diagnose and recover from errors.¹⁷⁹ Participants in some crowdsourcing projects have expressed a desire for feedback on the accuracy of their transcriptions based on the comparisons made with other transcriptions during data validation.¹⁸⁰ However, language used when giving feedback should be chosen carefully. The first version of *Ensemble* provided task feedback based on how closely the transcription matched other contributions. Unfortunately, the first transcribers of a piece of text were given a message that the system 'had a low "degree of confidence" in the work they had just submitted', which understandably made them feel 'either insulted or disheartened'.¹⁸¹

Participant communication

Some crowdsourcing projects provide ways for participants to communicate with each

¹⁷⁷ This small text change is an improvement on the original *Scripto* interface which displays 'Edit transcription' regardless of the system status.

¹⁷⁸ Flow is a state of deep, enjoyable focus or engagement reached when 'the challenges of the task meet the person's skills'. Csikszentmihalyi and Hermanson, 'Intrinsic Motivation in Museums'. Its role in crowdsourcing is discussed further in the next chapter.

¹⁷⁹ Nielsen, '10 Usability Heuristics for User Interface Design'.

¹⁸⁰ Various, 'AHRC Crowd Sourcing Study: Scoping Seminar'.

¹⁸¹ Reside, 'Crowdsourcing Performing Arts History with NYPL's ENSEMBLE'.

other, either in a general purpose discussion forum, on social media, or through comments on specific items. For example, transcribers in the *Smithsonian Transcription Center* can leave notes for other transcribers and reviewers on specific pages, or post questions on social media. After processing items on *Zooniverse* projects, participants can choose to discuss them in the *Talk* system. Mailing lists for grassroots projects such as *FreeBMD* date back to the late 1990s, providing an established model for community discussion.

The *Galaxy Zoo* forum was created in response to a 'flood of e-mails' after media attention brought many new participants to the project, in that hope that volunteers would 'communicate with one another and answer each other's questions'.¹⁸² Its success in doing so has influenced other crowdsourcing projects.¹⁸³ Community forums allow participants to discuss difficulties, help each other with specific queries, collate lessons learnt over time, share stories about interesting finds or potential discoveries, and provide feedback or suggestions for improvement to project stakeholders.¹⁸⁴ For some, being around other people while participating in a shared activity is inherently rewarding.¹⁸⁵ Alam and Campbell report that the community that formed on the *Trove* forum collaborated to create text correction guidelines after a lack of 'strict rules' from the organisation caused some issues.¹⁸⁶ Community discussion also provides opportunities for project staff to notice participants who could be deputised to take on more advanced tasks or

¹⁸² Raddick et al., 'Galaxy Zoo: Exploring the Motivations of Citizen Science Volunteers'.

¹⁸³ The role of participant forums in supporting participants in learning and skills development is discussed in the next chapter.

¹⁸⁴ For example, forum conversation led moderators to create an area for discussing and collecting 'First World War jargon in the diaries' Forum posters, 'Jargon Guide?', *Talk Operation War Diary History Board / First World War Jargon in the Diaries*, January 2014, <http://talk.operationwardiary.org/#/boards/BWDooooooooe/discussions/DWDooooooooo>.

¹⁸⁵ Robert A. Stebbins, 'Casual Leisure: A Conceptual Statement', *Leisure Studies* 16, no. 1 (January 1997): 17–25, doi:10.1080/026143697375485.

¹⁸⁶ Alam and Campbell, 'Crowdsourcing Motivations in a Not-for-Profit GLAM Context'.

responsibilities such as approving contributions or answering questions. Participants' expectations about the presence of project staff on discussion forums vary, and projects should be careful about how these forums are described so that expectations are not disappointed.

Some sites create forums only to find they are not used. It may be that a critical mass of participants is required to seed forum discussion,¹⁸⁷ or that some materials make better 'social objects' around which discussion can form.¹⁸⁸ Le Cornu and White's continuum of digital 'visitors' and 'residents' may be useful for anticipating the communication preferences of individuals within a given community.¹⁸⁹ It is clear that the tone of communication, the usability and other affordances of the community platform, the subject matter or type of research question, and the relative number of participants has an effect on the success or otherwise of project communities. Understanding the role that each of these play would be a useful topic for further research.¹⁹⁰

¹⁸⁷ For example, an early paper on Old Weather reported that fewer than 5% of participants used the site forum. Romeo and Blaser, 'Bringing Citizen Scientists and Historians Together'. Research on computer-supported cooperative work (CSCW) suggests that a critical mass of participants is necessary. See for example Mark S Ackerman, 'The Intellectual Challenge of CSCW: The Gap Between Social Requirements and Technical Feasibility', *Human-Computer Interaction* 15 (2000): 179–203.

¹⁸⁸ See for example the discussion of social discussions on *iSpot*. Doug Clow and Elpida Makriyannis, 'ISpot Analysed: Participatory Learning and Reputation', in *Proceedings of the 1st International Conference on Learning Analytics and Knowledge* (1st International Conference on Learning Analytics and Knowledge, Banff, Alberta, Canada, 2011), 34–43.

¹⁸⁹ While residents may go online purely to spend time with others, visitors regard the web as a set of tools 'which deliver or manipulate content' rather than a social space. David S. White and Alison Le Cornu, 'Visitors and Residents: A New Typology for Online Engagement', *First Monday* 16, no. 9 (2011), doi:10.5210/fm.v16i9.3171.

¹⁹⁰ Although their definition of 'leader' does not map directly onto roles in heritage crowdsourcing, the 'reader to leader' model presented by Preece and Shneiderman lists various 'usability and

The tone of the first discussions that newcomers encounter is important. Nonverbal content such as vocal inflection and visual, spatial, temporal cues are lost in written communication,¹⁹¹ so there is a greater chance that the message will be lost in translation (or, more formally, a 'misinterpretation or communication breakdown of the message or stimuli' between the sender and receiver will occur) unless attention is paid to tone.¹⁹² In addition to the impact of tone on the immediate sender and receiver(s) of a message, public messages on forums affect current and future readers of the discussion. Research on computer-supported cooperative work found that 'pre-existing contributions in an online news discussion site affect the thoughtfulness of contributions' made by newcomers.¹⁹³ Discussion on participatory sites might also have a negative effect on participant motivation. Perhaps due to the loss of nonverbal content, some volunteer moderator posts on the *Operation War Diary* forum appear to have a scolding tone, using phrases like 'what's the point' or 'worthless' while ostensibly encouraging participants to use hashtags to mark interesting content. This may be at odds with the project's goal of encouraging 'citizen historians'. The (over)use of common hashtags may represent an important initial stage of deeper engagement with the material rather than an inefficient use of time, and it is possible that statements discouraging hashtags introduce an element of uncertainty or anxiety that prevents exploratory tagging by newcomers.

sociability' factors that could be tested. Preece and Shneiderman, 'The Reader-to-Leader Framework'.

¹⁹¹ Together these make up to 93% of face-to-face communication, with only 7% of the message conveyed through words. Mehrabian, 1971, cited in Kristen Betts, 'Lost in Translation: Importance of Effective Communication in Online Education', *Online Journal of Distance Learning Administration* XII, no. II (2009), <http://www.westga.edu/~distance/ojdla/summer122/betts122.html>.

¹⁹² This also applies to messages from the system such as the *Ensemble* example mentioned earlier. Betts, 'Lost in Translation: Importance of Effective Communication in Online Education'.

¹⁹³ Sukumaran et al., 2011, cited in Preist, Massung, and Coyle, 'Competing or Aiming to Be Average?'

Finally, a potential disadvantage of community discussion should be considered.

Community discussion can be incredibly engaging, and therefore it can distract participants from the core goal of a project. In a telling example, the first post on a thread called 'Signs of OW addiction' said one of the 'Top Ten' signs of addiction might be that 'You spend more time on the forum than you do transcribing'.¹⁹⁴ Reed et al.'s research on *Zooniverse* projects found that 89% of users primarily spend time on the 'main research task of the project' while 11% of users primarily spend time on 'communication tools' of the project.¹⁹⁵ This is not a new issue - according to historian Samuel J. M. M. Alberti, the distractions of 'pleasure and cheerfulness' interfering with 'real observations' led to the separation of 'the social and scientific elements' of natural history societies in late Victorian Yorkshire.¹⁹⁶ However, some research suggests that these highly social participants who seek to express themselves and connect with other users are less likely to engage in the core tasks of a site,¹⁹⁷ so they may already be ideally placed to contribute by helping other participants.

¹⁹⁴ Forum posters, 'Signs of OW Addiction ...', *Old Weather Forum* » *Shore Leave* » *Dockside Cafe*, present 2010, <http://forum.oldweather.org/index.php?topic=1432.0>. Reddit moderators have also made comments like '...when you become a mod. You stop being a historian here, and spend most of your time moderating other historians' (or non-historians'!) content.' in Forum posters, 'Algernon_Asimov Comments on [Meta]100k Users, Eternal September, Rules, Moderators, and a Million Other Things.', *Reddit AskHistorians*, 2013, http://www.reddit.com/r/AskHistorians/comments/18sbnb/meta100k_users_eternal_september_rules_moderators/c8hxx6r.

¹⁹⁵ Reed et al., 'An Exploratory Factor Analysis of Motivations for Participating in Zooniverse'.

¹⁹⁶ Samuel J. M. M. Alberti, 'Amateurs and Professionals in One County: Biology and Natural History in Late Victorian Yorkshire', *Journal of the History of Biology* 34 (2001): 115-147.

¹⁹⁷ Fugelstad et al., 'What Makes Users Rate (Share, Tag, Edit...)?' Other research discusses the roles people adopt in discussion communities, including people who prefer to answer rather than post questions. Preece and Shneiderman, 'The Reader-to-Leader Framework'.

Fit between the platform, task and input material

Finding or creating software platforms that suit the crowdsourced task and the input material can help projects be more productive. For example, interfaces that break larger tasks into microtasks can increase the number of potential participants and the rate of participation. The software used also affects backend processes such as content workflow. Before the *DIY History* project started using *Scripto*, they manually copied emailed versions of transcriptions into their library software. Once *Scripto* was in place the process was streamlined, and participants could see their contributions appear live on the site immediately.¹⁹⁸ The library was also able to deputise some participants to approve records.¹⁹⁹

Re-using software created for one project can cause issues when the source material in the new project does not match the material for which the project was originally designed. As discussed earlier, some projects allocate the next task from a queue of unprocessed material, while others let participants choose from the material available on the repository. However, queue-based systems may not suit historical documents such as diaries or letters. For example, *Operation War Diary* was based on existing *Zooniverse* software,²⁰⁰ originally designed to process queues of images of galaxies. This underlying model, in combination with licensing restrictions, meant that participants could not move between diary pages. Participants expressed frustration at not being able to use their growing palaeographic experience, or the contextual knowledge gained on subsequent pages to go back and correct earlier jargon, places and personal name tags: 'This is more than just

¹⁹⁸ DiMeo, 'First Monday Library Chat: University of Iowa's DIY History'.

¹⁹⁹ DiMeo, 'First Monday Library Chat: University of Iowa's DIY History'.

²⁰⁰ First used for Old Weather. Arfon Smith, 'Making the Zooniverse Open Source', *Zooniverse*, 18 February 2013, <http://blog.zooniverse.org/2013/02/18/making-the-zooniverse-open-source/>.

tagging galaxy types, it's a story that you get better at interpreting as you go along'.²⁰¹ The desire to correct records on *OWD* may have a particular emotional resonance as they contain the names of specific individuals.²⁰² However, other projects report people returning to handwritten materials to edit previous transcriptions as they learn to interpret a particular hand, and adding new information as they re-read documents.²⁰³

More technical aspects can also have an effect. For example, systems should not enforce formats for historical names and dates, which may be only partially or imprecisely recorded, to the same standards as modern, born-digital data. Ideally, source material should be displayed at a reasonable image resolution, with image zoom, contrast and rotation functions available to all participants.

Some projects benefit from combining in-person events with online platforms. For example, community digitisation projects sometimes organise physical 'road show' events to collect personal documents. This allows them to explain the process, digitise material for collectors and obtain copyright permission, all tasks that would be more difficult for their target audiences if conducted purely online.²⁰⁴

²⁰¹ Forum posters, '[FEATURE REQUEST] Ability to Go Back and Change Tags'. Similar frustrations were also expressed in Forum posters, 'Why Is There No Back Button', *Talk Operation War Diary Help Board / I Need Some Help!*, January 2014, <http://talk.operationwardiary.org/#/boards/BWDooooood/discussions/DWDooooo3l>. Old Weather participants can edit or annotate previous transcriptions, which may have informed participants' expectations. Romeo and Blaser, 'Bringing Citizen Scientists and Historians Together'.

²⁰² This project is discussed further in the next chapter.

²⁰³ For example, Ellen Davis, 'Collaborative Transcription Project', 9 February 2012, <http://www.southwestern.edu/live/news/6475-collaborative-transcription-project>. Ben W. Brumfield, 'Wikilinks in FromThePage', *Collaborative Manuscript Transcription*, 14 March 2014, <http://manuscripttranscription.blogspot.co.uk/2014/03/wiki-links-in-fromthepage.html>.

²⁰⁴ Several projects have published useful material on the subject of community collecting projects.

Demotivators

As discussed above, participant motivations change over time, and projects that fail to match changing motivations may lose participants at a faster rate than those that provide for a range of motivations. Various factors can also cause participants to become demotivated. Discussions at an AHRC-sponsored workshop on humanities crowdsourcing attended by participants from several crowdsourcing projects illustrate some factors in demotivation.²⁰⁵ One reported frustration at delays in moderating her contributions to the *Dickens* project, suggesting that a lack of timely feedback on a completed task has a detrimental effect. Increasing the level of challenge may be one solution for avoiding the boredom that can arise as participant experience increases. As one participant said, solving puzzles like palaeography is 'only fun when you can't do it'.²⁰⁶ Others felt that the changing source material they worked with helped keep them motivated.²⁰⁷

Some projects have excellent reasons for compulsory registration before undertaking tasks, including spam prevention, tracking contributions in case of vandalism, and gatekeeping for specialist communities. However, compulsory registration is also a barrier to participation.²⁰⁸ Registration requires potential participants to decide whether they trust a site with their email address, attempt to find an available username, meet password

Latimer, 'Letter in the Attic: Lessons Learnt from the Project'. Thomas, Jones, and Witmer, 'History Harvests'. 'RunCoCo: How to Run a Community Collection Online' (Oxford: Oxford University Computing Services (OUCS), 2011), <http://projects.oucs.ox.ac.uk/runcoco/>. Dafis, Hughes, and James, 'What's Welsh for "Crowdsourcing"?'

²⁰⁵ Various, 'AHRC Crowd Sourcing Study: Scoping Seminar'.

²⁰⁶ Various, 'AHRC Crowd Sourcing Study: Scoping Seminar'. This is supported by Csikszentmihalyi's theory of 'flow', discussed further later. Csikszentmihalyi, *Flow: The Psychology of Optimal Experience*.

²⁰⁷ Various, 'AHRC Crowd Sourcing Study: Scoping Seminar'.

²⁰⁸ For example, this was discussed in Young, 'Year of the Bay Project Officer Evaluation'.

strength restrictions, and so on, forcing them to negotiate several small steps and decisions successfully. Sometimes they must do so before they have even seen the tasks and materials that they are signing up to work on. The result, as summarised in a headline by respected usability consultants Nielsen Norman Group is: 'Login Walls Stop Users in Their Tracks'.²⁰⁹ Compulsory registration not only deters some potential participants, it may also create enough of a delay between the intention to act and acting that some people who register on the site lose their spark of interest and never return.²¹⁰ *Zooniverse* data suggests that 10% of those registered never complete a single task.²¹¹ Gradual engagement could be used to let participants start a task without registering or logging in, then encouraging them to register later. Before their first task, participants on *Planet Hunters* are shown the text, 'Logged in users get to see the best stars and get credit for their work. Would you like to login?' and two options: a green button for 'yes' and a red one for 'no', making the preferred path very clear but letting participants choose to not register.

Anxiety and uncertainty

Feedback on the quality of their contributions helps participants understand where they can improve their work, and may also support motivations related to learning and mastering skills. *Old Weather* found that ceasing participation is strongly associated with

²⁰⁹ Raluca Budi, 'Login Walls Stop Users in Their Tracks', *Nielsen Norman Group*, 2 March 2014, <http://www.nngroup.com/articles/login-walls/>.

²¹⁰ Or they cannot recall their login details if they do return. Difficulties with validating accounts, email spam filters, supplying incorrect email addresses and delays in manual account approval may also account for some of those who register but never come back to complete even one task. Raluca Budi, 'Memory Recognition and Recall in User Interfaces', *Nielsen Norman Group*, 6 July 2014, <http://www.nngroup.com/articles/recognition-and-recall/>.

²¹¹ Based on the statement that if '90.8% completed at least one Task' in Luczak-Roesch et al., 'Why Won't Aliens Talk to Us?'

an anxiety about the quality of contribution,²¹² suggesting that providing feedback on tasks may improve retention. *Zooniverse* also reported that asking more 'abstract' questions could cause participants to lose confidence;²¹³ this may be because it is harder for participants to judge whether their response to an abstract question is 'correct'.

Arbitration is a process in which an arbitrator reviews transcriptions that do not match each other.²¹⁴ A *FamilySearch Indexing* blog post reports that inconsistent arbitration was the most common topic in over 600 reader comments discussing their indexing processes.²¹⁵ Comments on that post also demonstrate the importance of 'fair' arbitration to participants, including some that illustrate the demotivating effect of contradictory instructions that led to their work being marked as 'wrong' by some arbitrators.²¹⁶

Commercialising voluntary work

Projects in which the public domain status of the source material or project outputs is unclear may face difficulties attracting participants. For example, the *Tags en Uitleg* project may have suffered because potential participants were not sure whether their work would be commercialised.²¹⁷ The *Lives of the First World War (LFWW)* project, a partnership between the Imperial War Museum and genealogy company *Findmypast*, has

²¹² Eveleigh et al., 'Designing for Dabblers and Deterring Drop-Outs in Citizen Science'.

²¹³ Chris Lintott, 'Why the Zooniverse Is Easy to Use.', *Zooniverse*, 17 February 2013, <http://blog.zooniverse.org/2013/02/17/why-the-zooniverse-is-easy-to-use/>.

²¹⁴ Hansen et al., 'Quality Control Mechanisms for Crowdsourcing'.

²¹⁵ Jennifer Anderson, 'A Message from the Indexing Workforce Team', *FamilySearch Blog*, 6 May 2013, <https://familysearch.org/blog/en/message-indexing-workforce-team/>.

²¹⁶ One reports their contribution level falling from 'a couple of records a day' to '5 a month' while another who has indexed 'over 100,000 names' was discouraged by the effect of contradictory arbitration decisions on their reputation 'score'. Jennifer Anderson, 'The Key to Unlocking Their Stories', *FamilySearch Blog*, 28 August 2013, <https://familysearch.org/blog/en/key-unlocking-stories/>.

²¹⁷ Noordegraaf, Bartholomew, and Eveleigh, 'Modeling Crowdsourcing for Cultural Heritage'.

encountered resistance or confusion on several levels. Some object to the 'rampant commercialization of remembrance'²¹⁸ and conclude that they prefer to 'remember my relatives personally and not share the info I have'.²¹⁹ Others were expecting an archive and were surprised to find a 'commercial' site;²²⁰ subsequently, some became suspicious of the project motives: 'it seems that all you want is money and not the information!'.²²¹ The *LFWW* feedback section also contains posts from people reluctant to create content voluntarily that the site 'will charge other people to access'.²²² Rose Holley's 2010 summary of research on participation in *Distributed Proofreaders*, *FamilySearch Indexing*, *Wikimedia* and *Trove* reported that volunteers 'do not want to feel that their work can be commercially exploited'.²²³ Projects that commercialise, or appear to commercialise, existing work may face revolt. When *Ancestry* began including *FreeBMD* records in their research results, many *FreeBMD* volunteers objected to this perceived land grab, and

²¹⁸ Forum posters, 'Close the Site down. It Seeks to Commercialize Remembrance and Is Poorly Designed. – Member Feedback for Lives of the First World War', accessed 25 May 2014, <http://support.livesofthefirstworldwar.org/forums/246136-feedback/suggestions/5926545-close-the-site-down-it-seeks-to-commercialize-re>.

²¹⁹ Forum posters, 'Ease of Use and Linkage to Commercial Organisation – Member Feedback for Lives of the First World War', accessed 25 May 2014, <http://support.livesofthefirstworldwar.org/forums/246136-feedback/suggestions/5917863-ease-of-use-and-linkage-to-commercial-organisation>.

²²⁰ Forum posters, 'Not Impressed. This Is a Commercial Site Not a National Archive. Very Frustrated. Not How It Has Been Advertised. – Member Feedback for Lives of the First World War', accessed 25 May 2014, <http://support.livesofthefirstworldwar.org/forums/246136-feedback/suggestions/5918130-not-impressed-this-is-a-commercial-site-not-a-nat>.

²²¹ Forum posters, 'Why Have to Pay Again? – Member Feedback for Lives of the First World War', accessed 25 May 2014, <http://support.livesofthefirstworldwar.org/forums/246136-feedback/suggestions/5956766-why-have-to-pay-again>.

²²² Forum posters, 'The British Public Has Already Paid to Collect the Information Concerning Our Armed Forces Who Fought in the First World War.', *Member Feedback for Lives of the First World War*, accessed 25 May 2014, <http://support.livesofthefirstworldwar.org/forums/246136-feedback/suggestions/5920281-the-british-public-has-already-paid-to-collect-the>.

²²³ Holley, 'Crowdsourcing: How and Why Should Libraries Do It?'

worried that *Ancestry* might charge people for records that they had created for the common good.²²⁴

Emerging design patterns

In addition to individual design features, some broader design patterns (generic solutions for particular types of problems) are emerging. Some are responses to the challenge of retaining participants; others come from the desire to make tasks as small and simple as possible or to help manage the workload that crowdsourcing projects can create for institutions.

Crowdsourcing ecosystems

Crowdsourcing ecosystems are systems in which related applications are combined to process different aspects of the same source materials. For example, *Distributed Proofreaders* was created to proofread texts produced by *Project Gutenberg*.²²⁵ The New York Public Library has two ecosystems, one for their menus collection and another for their maps. *What's on the Menu* focuses on transcribing menu items, while the *What's on the Menu Geotagger* focuses on geolocating the documents through location information available on the menu. Their *Building Inspector* offers five tasks based on historical maps, each embedded in an interface dedicated to supporting participants in completing that specific task. Together, the tasks offered (checking or fixing automatically-detected building 'footprints', entering street numbers, classifying colours or finding place names) contribute to the larger goal of digitising the maps.

²²⁴ See posts to the *FreeBMD* discussion lists circa October 2001 (<http://archiver.rootsweb.ancestry.com/th/index/FreeBMD-Admins/2001-10> and <http://archiver.rootsweb.ancestry.com/th/index/FREEBMD-DISCUSS/2001-10>; both pages last accessed 11 June 2015).

²²⁵ Distributed Proofreaders, 'DP: Welcome'.

Ecosystems can also be used to process material in stages - for example, marking pages as containing no text, printed or type-written text, or as containing clear, difficult or very difficult hand-writing - could help process documents into batches based on relative difficulty or time per task.

By allowing tasks to be broken into microtasks, these ecosystems may be able to attract a larger number of casual contributors. Providing different tasks also helps prevent participants from becoming bored and leaving the project, while also keeping them motivated by offering new opportunities to learn or practice skills,²²⁶ which is important for retaining experienced participants. Minimising the attrition rates for project participants is an important factor in the success of projects. Retaining long-term participants can also help newcomers, sharing their experience in community forums.

Promoting participants

Projects that promote participants to roles that require more skills or responsibility provide a slightly different model for ecosystems. In this model, participants are taking on roles, such as discussion moderation or task review, that have a bigger influence on the overall project than the completion of a microtask. Participants given extra responsibilities may have requested them, or they may have been noticed for their 'high level of skill and dedication'.²²⁷ These promotions provide more experienced participants with an opportunity to gain new skills and become more deeply involved with the project.

Csikszentmihalyi's theory of flow, which posits that the challenge of a task must match the skills of an individual, helps explain why providing a range of tasks motivates participants

²²⁶ Csikszentmihalyi and Hermanson, 'Intrinsic Motivation in Museums'.

²²⁷ DiMeo, 'First Monday Library Chat: University of Iowa's DIY History'.

to keep contributing to a project.²²⁸ Asking volunteers to take on some quality control tasks can have a positive effect on the overall project by helping to reduce the demotivating delay before contributions are approved and made live on a site.²²⁹

Challenges

Challenges are activity drives based on targets set by crowdsourcing projects. They usually set a target goal to be reached by a specific time. They may be based on a specific topic or dataset, and can be tied to a particular motivation or events such as anniversaries. For example, *Ancestry's World Archives Project* uses social media posts to nominate projects for indexing, then rewards and publicly recognises the participants who indexed or reviewed the most records in each project.²³⁰ Challenges have also been successfully used by the *Smithsonian Transcription Center*,²³¹ *Free the Files*,²³² *Dickens Journals Online* and the 1940 US Census transcription project.²³³

²²⁸ Mihaly Csikszentmihalyi, *Creativity: Flow and the Psychology of Discovery and Invention* (New York, NY, USA: Harper Perennial, 1997).

²²⁹ Better methods for manually approving contributions may also reduce a reliance on simple 'tag agreement' validation methods (in which more commonly added tags are kept while unique tags are removed in an attempt to reduce spam or vandalism) that would remove unique specialist terms added by those with specific disciplinary or experiential knowledge.

²³⁰ For example, Ancestry World Archives Project, 'With about 90 Minutes Left There Are...', *Facebook*, 31 January 2015, <https://www.facebook.com/AncestryWorldArchivesProject/posts/10153104225303487>. Participant prizes include a three-month Ancestry membership. Anna Fechter, 'Let's Start 2015 Off With a Bang', *Ancestry World Archives Project News*, 2 January 2015, <http://blogs.ancestry.com/worldarchivesproject/?p=2322>.

²³¹ Smithsonian National Museum of Natural History, '#FWTrueLove Transcription Challenge', *Field Book Project*, 13 February 2015, <http://nmnh.typepad.com/fieldbooks/2015/02/fwtruelove-transcription-challenge.html>.

²³² Amanda Zamora, 'Crowdsourcing Campaign Spending: What We Learned From Free the Files', *ProPublica*, 12 December 2012, <http://www.propublica.org/article/crowdsourcing-campaign-spending-what-we-learned-from-free-the-files>.

²³³ 1940 US Census Community Project, 'We Did It! The 1940 US Census Community Project'.

Challenges can be a good way to focus on specific tasks that might not be part of the usual site activity. For example, *Ravelry* (a site where 'knitters and crocheters' can 'keep track of their yarn, tools, project and pattern information'²³⁴) held a week-long 'party' to enter metadata to support a structured search function; in that time 23,500 participants 'categorized and assigned attributes to nearly 160,000 patterns'.²³⁵

Mini-projects

One noticeable feature of the *Smithsonian Transcription Center* is the size of their 'projects' - some are as small as a single field notebook.²³⁶ Breaking larger projects into smaller ones has several advantages. Each mini-project has been described with the date, location, or purpose of the documents, providing more hooks to interest potential participants. Similarly, *Notes from Nature* has specialised landing pages for each collection, listing its temporal range, geographic context, taxonomic coverage, and why the collection is important. The small size of the projects, from tens to hundreds of pages, allows progress through the stages of transcription, review and approval to be made relatively quickly. *Notes from Nature* also divides their material into smaller batches as it gives them more opportunities to 'celebrate the success of completion'.²³⁷ Meanwhile, the V&A's progress bar on their image-cropping project shows the total number of items completed out of its target of 120,072 items; a goal so sizeable that the incremental progress is barely noticeable.

²³⁴ Ravelry, 'About Our Site', *Ravelry*, accessed 6 June 2015, <http://www.ravelry.com/about>. Many thanks to the enthusiastic knitter who showed me through the site.

²³⁵ rainydaygoods, 'Ravelers Rocked the Search Party!', *Unraveled*, 20 July 2010, <http://blog.ravelry.com/2010/07/20/ravelers-rocked-the-search-party/>.

²³⁶ By 15 May 2015, 722 projects had been completed. SI Transcription Ctr, 'Today's Numbers'.

²³⁷ Rob Guralnick, 'Making Progress Clear on Notes from Nature', *Notes from Nature*, 24 February 2014, <http://blog.notesfromnature.org/2014/02/24/making-progress-clear-on-notes-from-nature/>.

In providing a range of topics to work on, mini-projects may also help retain participants who like a sense of autonomy over what they work on.²³⁸ Mini-projects allow larger repositories to be grouped by topic, location, people or institutions and other categories that might help attract motivated participants. The more frequent milestones that smaller projects enable also help demonstrate the impact of participation.

Niche projects

Tightly-focused niche projects are closely related to the mini-projects described above, and feature specific temporal, geographical or topical foci. Terras found that amateur digitisation projects tended to be focused on 'novel, detailed, and niche content with a very specific scope'.²³⁹ Niche projects are inherently interesting to anyone with a passion for that particular topic, and tend to attract ardent contributors. The concentration of material in a niche project may also mean they are more discoverable in searches for specific specialist terms. Niche projects may lend themselves to better storytelling, making the task of marketing a project easier. In addition to breaking datasets into mini-projects, *Notes from Nature* attracts niche participants by grouping material into 'compelling themes' derived from their topic or 'other societal value'.²⁴⁰

Projects based on specific locations may have some immediate relevance for locals and those with connections to the area. *History Harvest* reports that projects that are 'responsive to local contexts and concerns are more likely to generate widespread support'

²³⁸ However, projects should also suggest tasks for those who prefer being directed. For example, *Papers of the War Department* have 200 documents nominated as suggested transcriptions. Leon, 'Build, Analyse and Generalise'. See also Holley, 'Crowdsourcing: How and Why Should Libraries Do It?'

²³⁹ Terras, 'Digital Curiosities'.

²⁴⁰ Guralnick, 'Making Progress Clear on Notes from Nature'.

from a community.²⁴¹ As for mini-projects, the specificity of niche projects may provide more hooks for the curious and the already interested. The *World Archives Project* lists the location, type of record (usually the source of the record or the document type), and date range. A *FamilySearch* blog post promises an upcoming feature where 'you will be able to look into certain projects and select a batch in a specific location or time period'²⁴², presumably in the knowledge that family historians will want to focus on particular places and times, and will therefore help process them quickly.

Working on closely related items may lead participants to notice potential patterns and become curious about unusual information present in records. Curiosity can both encourage participants to view more items and to discuss the items with others. Huvila hoped that 'specificity' might induce 'deep collaboration' between individual researchers.²⁴³ (This may be one reason for the success of projects like *Old Weather*.) Niche projects may be more likely to develop a constructive participant community based on a shared interest.²⁴⁴ Participant communities are discussed further in the next chapter.

The editors of *Suda On Line* credit the project's focus on a single, bounded text with its success because the bounded nature of the task created 'natural milestones' as various markers of progress were reached.²⁴⁵ The challenges, mini-projects and niche projects discussed here also suggest that a sense of boundedness helps motivate participants.

²⁴¹ Thomas, Jones, and Witmer, 'History Harvests'.

²⁴² Flinders, 'The Future of Indexing'.

²⁴³ Isto Huvila, 'Participatory Archive: Towards Decentralised Curation, Radical User Orientation, and Broader Contextualisation of Records Management', *Archival Science* 8, no. 1 (September 2008): 15–36, doi:10.1007/s10502-008-9071-0.

²⁴⁴ For example, Dunn and Hedges posit that 'passion for the subject' draws participants into communities. Dunn and Hedges, 'Crowd-Sourcing Scoping Study'.

²⁴⁵ Mahoney, 'Tachypaedia Byzantina'.

Conclusion

The field of crowdsourcing is changing rapidly, and it is difficult to predict the impact of future technologies or cultural movements, but the move to mobile devices and advances in technology could present future design challenges as well as opportunities. American data suggests that people increasingly use mobile phones and tablet devices in their leisure time.²⁴⁶ Design issues for mobile and tablet devices include limited screen space, uncertain data connectivity, and the lack of a physical keyboard.²⁴⁷

Advances in technology such as handwritten text recognition and machine learning could challenge current models for crowdsourcing projects. Broadly speaking, 'machine learning' is a label for technologies that allow computers to learn from the data available to them. Microsoft recently announced technology that can automatically identify objects in a picture,²⁴⁸ and the *tranScriptorium* project is developing tools to transcribe historical handwritten documents.²⁴⁹ *Your Paintings Tagger*, *Building Inspector*, and some *Zooniverse* projects are already combining crowdsourcing and machine learning to improve the abilities of computers to classify and identify items in images.²⁵⁰ This leads to the first issue: computers are getting better at microtasks such as text transcription and image

²⁴⁶ Jack Marshall, 'Mobile Isn't Killing the Desktop Internet', *CMO Today*, *The Wall Street Journal*, 26 May 2015, <http://blogs.wsj.com/cmo/2015/05/26/mobile-isnt-killing-the-desktop-internet/>.

²⁴⁷ Larger organisations such as *FamilySearch Indexing* have been working to address this for several years. See for example Judson, 'The Mobile App—What We've Learned'.

²⁴⁸ Allison Linn, 'Picture This: Microsoft Research Project Can Interpret, Caption Photos', *Next at Microsoft*, 28 May 2015, <http://blogs.microsoft.com/next/2015/05/28/picture-this-microsoft-research-project-can-interpret-caption-photos/>.

²⁴⁹ *tranScriptorium*, 'Objectives'.

²⁵⁰ Rachel Collings, 'The Art of Computer Image Recognition', *The Public Catalogue Foundation*, 5 May 2014, http://www.thepcf.org.uk/what_we_do/48/reference/862. Kyle Willett, 'New Paper: Galaxy Zoo and Machine Learning', *Galaxy Zoo*, accessed 31 March 2015, <http://blog.galaxyzoo.org/2015/03/31/new-paper-galaxy-zoo-and-machine-learning/>.

classification - tasks that many crowdsourcing participants find satisfying, and that may be important first steps in developing new historical interests. What impact will this have on crowdsourcing projects and participants? Secondly, organisations like the Citizen Science Alliance, whose core values include 'do not waste participants' time',²⁵¹ may find that this conflicts with participants' desires to engage in satisfying tasks even though those tasks could eventually be more efficiently processed computationally.

However, these new technologies can also be harnessed to make the initial crowdsourced microtasks even smaller.²⁵² The success of OCR correction projects like *Trove* shows that providing some pre-processed data might actually make tasks easier, and therefore more enjoyable.²⁵³ An emerging design pattern uses the results of crowdsourced tasks to help software learn about a dataset. The software can then make a first pass at classification or metadata creation for a new dataset. The results of this process are then presented to crowdsourcing participants for validation or correction. For example, you can speed through a page that presents all the paintings that software thinks contain a cow and mark the occasional image that mistakenly shows a horse,²⁵⁴ or vote on tags to describe recordings from the BBC World Service archive that have been suggested by software.²⁵⁵

²⁵¹ Chris Lintott and The Citizen Science Alliance, 'The Two-and-a-Bit Page Guide to Running a Zooniverse Project' (The Citizen Science Alliance, n.d.).

²⁵² See also research on 'social machines', in which people and computers are part of a larger integrated system. Paul R. Smart, Elena Simperl, and Nigel Shadbolt, 'A Taxonomic Framework for Social Machines', in *Social Collective Intelligence: Combining the Powers of Humans and Machines to Build a Smarter Society*, ed. Daniele Miorandi et al. (Berlin, Germany: Springer, 2014), <http://eprints.soton.ac.uk/362359/>.

²⁵³ This does not mean that projects should only offer 'easy' microtasks, as the relationship between challenge and enjoyment is complex, but they may allow for a broader range of participants and thereby create more opportunities for deeper engagement.

²⁵⁴ Collings, 'The Art of Computer Image Recognition'.

²⁵⁵ Tristan Ferne, 'Algorithms and Crowd-Sourcing for Digital Archives', *Create Hub*, 29 May 2014,

These tasks may be more enjoyable because correction tasks have a lower cognitive load, but it might also be that it is easier to recognise rather than recall information.²⁵⁶ Finally, machine learning technologies could be used to provide personalised feedback on participant tasks, helping them feel more confident and learn skills more quickly.²⁵⁷

My observations of crowdsourcing projects have revealed two interesting tensions between the needs of a project and the wishes of its participants. The first tension is between the desire for productivity and clean datasets, and the potential to support the interests that participants develop during a project. These interests commonly include discussing their findings in forums and undertaking related research projects, but they may be quite varied - Holley also discusses participants who started correcting and processing stories serialised in historical newspapers then uploading them to *Project Gutenberg*.²⁵⁸ The other tension also lies between project productivity and participants' own preferences. In one example, participants on the *Operation War Diary* forum debated the correct uses of hashtags over many discussion threads. While some believed that only tags that were useful to the

<http://www.create-hub.com/comment/algorithms-and-crowd-sourcing-for-digital-archives/>.

²⁵⁶ Budiu, 'Memory Recognition and Recall in User Interfaces'.

²⁵⁷ Projects such as *Herbaria@Home* and *Old Weather* provide newcomers with feedback upon request, but to be effective this requires newcomers who feel confident enough to ask for feedback and a sufficient number of experienced participants with time to respond to requests. Tom Humphrey, 'About Feedback Requests', *Herbariaunited.Org Forum Index -> Herbarium Sheet Discussion*, December 2007, <http://herbariaunited.org/forum/viewtopic.php?t=679>. Forum posters, 'Ask for Expert Advice', *Old Weather Forum*, 2013, <http://forum.oldweather.org/index.php?action=printpage;topic=3944.o>. Participant requests for task feedback include Forum posters, 'What Could We Improve?', *Talk Operation War Diary Help Board / Suggestions for Improvements*, January 2014, <http://talk.operationwardiary.org/#/boards/BWDooooo08/discussions/DWDooooo04>.

²⁵⁸ Rose Holley, 'From Little Things Big Things Grow – Gold Star to Newspaper Text Correctors for E-Books!', *Rose Holley's Blog*, 13 December 2011, <http://rose-holley.blogspot.co.uk/2011/12/from-little-things-big-things-grow-gold.html>.

project should be added, others wanted to use tags (or comments) to note things that were interesting to them (for example, the mention of trenches, horses and tasks related to specialist military units).²⁵⁹ Some participants take more time on a task than a project might wish because they are performing additional checks on the source material. For example, some forum posters mention that they check the names of individual people or places in relevant sources before tagging them. This is illustrated by an exchange on one project where a participant mentioned the material that he had collected, then concluded, 'None of the above is ready to send to you as I have more research to do'.²⁶⁰

However, these tensions can be resolved by considering the benefits of crowdsourcing projects for the organisations that run them. Crowdsourcing can enhance the discoverability of collection items by creating or enhancing metadata, bridging the 'semantic gap',²⁶¹ transcribing or correcting text and making it easier for external people to share their expert knowledge about specific items. These tangible benefits are important, particularly when funding for digitisation is limited. However, the intangible benefits may be equally important. Michael Lascarides and Ben Vershbow have reported that the New

²⁵⁹ For example, Forum posters, 'Casulties Tag - Room For Improvment', *Talk Operation War Diary Help Board / Suggestings for Improvements*, 2014, <http://talk.operationwardiary.org/#/boards/BWDooooo8/discussions/DWDoooo6eo>. and Forum posters, 'How Do I Use Hashtags?', *Talk Operation War Diary Help Board / FAQs - Tags and Tagging*, 2014, <http://talk.operationwardiary.org/#/boards/BWDoooooof/discussions/DWDoooo34v>.

²⁶⁰ Forum posters, 'In Person Meetings - Update', *Civil War Pathways*, 4 October 2013, <http://civil-war-pathways.12348.n7.nabble.com/update-td503.html>. Again, this is not a new issue. The Victoria County History also encountered contributors who wanted to do new research instead of compiling existing material. John Beckett and Charles Watkins, 'Natural History and Local History in Late Victorian and Edwardian England: The Contribution of the Victoria County History', *Rural History* 22, no. 01 (March 2011): 59–87, doi:10.1017/S0956793310000142.

²⁶¹ Trant, 'Tagging, Folksonomy and Art Museums: Results of Steve.Museum's Research'.

York Public Library came to regard crowdsourcing as an extension of its core mission. Lyn Lewis Dafis, Lorna M. Hughes and Rhian James's translation of 'crowdsourcing' into Welsh ('cyfrannu torfol') links it to the 'social engagement' and 'collective contributions' traditionally important to libraries.²⁶² If an institutions' goal is to provide meaningful opportunities for the public to not only access, but connect with or learn from historical collections, then time spent on these additional tasks is time well spent. Crowdsourcing projects - and the publicity they garner - can help the public become aware of, and care more deeply about the institutions that preserve their heritage.

This review has shown the importance of good interface design in achieving these tangible and intangible goals. It has discussed the relationship between participant motivations and project marketing, niche and mini-projects, participant discussion and promoting participants to more complex tasks. It has examined the impact of platform choices on the participant experience of the source material, and the role of crowdsourcing ecosystems in providing a range of participant tasks. Future research that sets out to test these possible heuristics experimentally would be useful, because the factors discussed could help make the difference between projects that thrive and those that struggle to reach their goals. A potential limitation of this analysis is that it only involved one expert reviewer, using heuristic and expert evaluation and trace ethnography rather than formal user research. Allowing for that caveat, people designing heritage crowdsourcing projects might find the following suggestions useful. At a bare minimum, any barriers to participation should be reduced or removed. Usability testing can minimise dissatisfaction, but participatory projects must also convince potential contributors to take positive action. The most important factors are good publicity, polished task design and connecting the project to both a shared, significant goal and participant motivations. Project titles, straplines and

²⁶² Dafis, Hughes, and James, 'What's Welsh for "Crowdsourcing"?'

other messages should explain the altruistic and intrinsic value in contributing, while the front page should show the impact of previous contributions (providing social proof while demonstrating the difference each task makes) and ideally provide a preview of the main task. Factors such as a good onboarding process, a convincing call to action, designing the smallest possible tasks and helping participants find tasks suitable for their skills and interests can encourage initial participation. Designing rewards for participation carefully to match the project's goals, thanking participants appropriately, minimising the delay between task completion and the results appearing on the site, providing updates on progress and impact, and in many cases, providing opportunities for participants to discuss their task, can help motivate continued participation. This has implications for project management - organisations should allow time for community engagement and on-going marketing and outreach communication after a project launches. Ideally, further resources would be available to update interfaces and platforms after launch, as participants suggest improvements or the project develops in unexpected ways. Some projects may benefit from applying emerging design patterns such as crowdsourcing ecosystems, promoting participants, challenges, mini- and niche projects. Finally, projects should ensure that participants have access to the data they have helped create, and ideally also to the original source materials they worked on.

This chapter turned from the broad landscape of participatory digital history projects to the specific sub-category of heritage crowdsourcing projects. While new technologies like machine learning may change how participants encounter tasks like tagging and transcription, the content - and more importantly, the new relationships between the public and heritage institutions - created through crowdsourcing, have already created a significant legacy. The willingness to trust 'the crowd' to contribute to historical tasks online is a significant shift for institutions, and one likely to outlast any specific

participatory platform. The next chapter focuses on how some crowdsourcing projects encourage deeper engagement with their source material and related scholarship.

Chapter 3: History with the public: from microtasker to historian?

Some crowdsourcing projects discussed in previous chapters provide spaces in which participants can learn and develop historical skills. Understanding how they do so adds to our knowledge of how and why people contribute to digital history resources. The projects discussed in this chapter provide further evidence for the impact of digital projects on public participation and scholarly practices in historical research, and help us understand the implications of providing more complex tasks and responsibilities in crowdsourcing projects. This chapter seeks lessons from crowdsourcing and citizen science that can inform the development of citizen history. Specifically, it looks for project attributes that are important in supporting deeper engagement with historical materials and skills.

In this chapter I briefly discuss methods then define relevant terms and set out the context provided by the literature and previous projects, including topics such as engagement, serious leisure, and the role of flow and curiosity in learning and skills development. I then provide examples of crowdsourcing projects that make claims about citizen history and/or illustrate the development of deeper engagement and historical skills. I examine the role of project and task design, curiosity, access to data, communities of practice and access to expertise before arguing that some crowdsourcing projects have provided spaces in which participants can develop historical skills.

Methods

I have followed the methods outlined in the *Introduction* and *Chapter 1*, focusing

particularly on undertaking the task(s) that projects set for participants when reviewing interface and interaction design. While the skill and knowledge required to develop a historical argument or question may be acquired through participation in crowdsourcing projects, the process is generally likely to take place outside the project and may not necessarily leave visible traces except in community discussion or publications citing the project. Accordingly, I have used trace ethnography¹ to gain qualitative insight by examining documentary traces of the interactions of participants available in project documentation, participant forum posts and other social media. Forum participation rates vary, with those reviewed here ranging from fewer than ten to many hundreds of posts.² An important caveat is that forum posters are generally more deeply engaged with the project than the average participant, and that the issues they raise, and the motivations and interests they discuss may not represent all participants; nonetheless, they provide useful insights. I also looked to project publications for information on outcomes and their own analyses, supplemented with interviews or informal conversations with project stakeholders.³

¹ As discussed in the *Introduction*, this is the analysis of documents and other traces of online communication such as forum discussions. Geiger and Ribes, 'Trace Ethnography'.

² A recent *Zooniverse* blog post said that 'more than 10% of our volunteers make comments' in the Talk sections of their projects. The *Zooniverse*, 'It's Good to Talk!', *Daily Zooniverse*, 10 April 2014, <http://daily.zooniverse.org/2014/04/10/its-good-to-talk/>. Research on 10 *Zooniverse* projects 'Talk' and 'Task' participation rates found overall '40.5% of users had contributed to Talk, while 90.8% completed at least one Task' and the 'average' user contributed 'a median of 600 Task entries' and 'a median of 14 posts on Talk'. Luczak-Roesch et al., 'Why Won't Aliens Talk to Us?' *Zooniverse* forum participation rates may be particularly high as they have integrated their 'Talk' interface into object pages, and directed participants to forums for help and documentation.

³ Including three *Zooniverse* staff members, one *Founders and Survivors* staff member, four *Children of the Lodz Ghetto* staff members, and survey responses and other conversations with various other project staff.

Definitions and context

In this section I introduce some of the extensive literature, key concepts and significant projects in public participation in scientific research and 'citizen science' crowdsourcing. I have included projects and literature from citizen science as it has provided an influential model that has, in some cases, been adopted by historical projects. Citizen science contains a number of examples for participants in crowdsourcing projects moving from microtasks to deeper engagement with the related scientific research. These projects publish their methods and results widely (in peer-reviewed journals and on social media), providing insight into participant motivations and the role of participant forums in encouraging an interest in the science and history of their datasets.

Defining citizen science

Comparing definitions of citizen science can be difficult, as the context in which they were written ranges from local environmental stewardship and volunteer monitoring,⁴ to urban planning and participatory action research,⁵ and science education and outreach in subjects ranging from astronomy to zoology.⁶ However, common factors in definitions of

⁴ Candie C. Wilderman, 'Models of Community Science: Design Lessons from the Field', in *Citizen Science Toolkit Conference*, ed. C. McEver et al. (Citizen Science Toolkit Conference, Cornell Laboratory of Ornithology, Ithaca, New York: National Science Foundation, 2007), <http://birds.cornell.edu/citscitoolkit/conference/proceeding-pdfs/Wilderman%202007%20CS%20Conference.pdf>.

⁵ Here, some researchers draw the lines differently. For example, Cooper et al. consider citizen science to be distinct from participatory action research because it 'typically does not incorporate iterative or collaborative action', although participants may help develop methodologies in collaboration with professional researchers. Cooper et al., 'Ecology and Society'. For Bonney et al., both citizen science and participatory action research 'actively involve the public directly in the multifaceted and iterative processes of scientific investigation'. Bonney et al., 'Public Participation in Scientific Research'.

⁶ Additionally, some definitions are reacting against other work which considers 'participants' to include people who participate as subjects in surveys or medical trials. See for example Wiggins and

citizen science are: projects are initiated by organisations rather than self-organised by participants; core contributions generally involve collecting or processing data; and the projects must be 'authentic research projects'.⁷ Bonney et al. devised an influential model for participatory scientific research projects, grouping them according to the amount of control participants have over steps such as defining questions for study, collecting and analysing data, and drawing conclusions.⁸ Their model contains three categories for projects, each of which builds on the previous category: 'contributory', where the public contributes data to a project designed by the organisation; 'collaborative', where the public can additionally help refine project design and analyse data in a project led by the organisation; and 'co-creative', where the public can take part in all or nearly all processes and all parties design the project together.⁹

Definitions of citizen science tend to bifurcate over the question of whether citizen science projects necessarily allow participants to take part in data analysis and research design. In

Crowston, 'From Conservation to Crowdsourcing: A Typology of Citizen Science'. and Muki Haklay, 'Citizen Science and Volunteered Geographic Information – Overview and Typology of Participation', in *Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice*, ed. Daniel Z. Sui, Sarah Elwood, and Michael F. Goodchild (Dordrecht: Springer Netherlands, 2013), <http://www.springerlink.com/index/10.1007/978-94-007-4587-2>. Other work considers 'virtual citizen science' as a special category of hierarchical online project, contrasting it with self-organised movements like open source software and peer production. For example, see Reed, Rodriguez, and Rickhoff, 'A Framework for Defining and Describing Key Design Features of Virtual Citizen Science Projects'.

⁷ Reed et al., 'An Exploratory Factor Analysis of Motivations for Participating in Zooniverse'.

⁸ Public Participation in Scientific Research (PPSR) is in many ways a precursor to citizen science. Bonney et al., 'Public Participation in Scientific Research'.

⁹ Most heritage crowdsourcing projects discussed in the previous chapter are contributory, but some discussed in this chapter, such as *Old Weather* and *Herbaria@Home*, have evolved into collaborative and at least partly co-creative projects.

the 'contributory citizen science'¹⁰ camp, for example, researchers Andrea Wiggins and Kevin Crowston explain that participation in citizen science usually involves contribution made under an 'established protocol, or completing structured recognition, classification, or problem-solving tasks that depend on human competencies'.¹¹ In the 'collaborative citizen science' camp, projects are closer to partnerships.¹² For example, Muki Haklay's synthesis defines citizen science as 'scientific activities in which non-professional scientists volunteer to participate in data collection, analysis and dissemination of a scientific project'.¹³ For researchers on *Zooniverse* projects the authenticity in 'authentic research projects' comes from participants' involvement in parts of 'the process of scientific inquiry' including 'data gathering, data cleaning, data analysis and interpretation, and research question and hypothesis formation'.¹⁴ At the first AGM of the European Citizen Science Association (ECSA) in May 2014, its members defined '10 principles of citizen science', the first of which said: 'Citizens can act as contributors, collaborators, or as project leader', and are 'encouraged to participate in multiple stages of the scientific process'.¹⁵ This suggests an emerging consensus that citizen science projects are collaborative rather than (merely) contributory.¹⁶

¹⁰ Here I am drawing on Bonney et al.'s definitions of 'contributory' and 'collaborative'.

¹¹ Wiggins and Crowston, 'From Conservation to Crowdsourcing: A Typology of Citizen Science'.

¹² Wilderman, 'Models of Community Science'.

¹³ Haklay, 'Citizen Science and Volunteered Geographic Information'.

¹⁴ Reed et al., 'An Exploratory Factor Analysis of Motivations for Participating in Zooniverse'.

¹⁵ Principles quoted in Muki Haklay, 'European Citizen Science Association Suggestion for 10 Principles of Citizen Science', *Po Ve Sham - Muki Haklay's Personal Blog*, 14 May 2014, <http://povesham.wordpress.com/2014/05/14/european-citizen-science-association-suggestion-for-10-principles-of-citizen-science/>.

¹⁶ At the same time, there is a reaction to the increase in 'top-down' projects in which citizen science is 'defined as something that professional scientists, or "practitioners", do with the help of volunteers', leading to the coinage of new terms like 'upscience' to describe in contrast to 'self-initiated research'. Francois Grey, 'Upscience: Inverting the Pyramid of Scientific Enquiry', *Billion*

However, in practice many citizen science projects are focused on contributions. Internet researcher David Weinberger concluded that participants in citizen science projects 'are doing the work of science — gathering data certainly counts — but not the work of scientists'.¹⁷ While the *Zooniverse* team pointed out that the Galaxy Zoo project 'provides enough information for those who are interested to take control of the entire scientific process',¹⁸ the core goal of most *Zooniverse* projects does not require this level of involvement. Currently, it seems the label 'citizen science' perhaps describes the potential for deeper engagement rather than the necessity for it. That is, most projects rely on contributors to collect or process data, but very few are dependent on participants helping with analysis data or setting research questions.¹⁹

Following the model set by citizen science projects, the United States Holocaust Memorial Museum began using the term 'citizen history' in 2011 to describe their *Children of the Lodz Ghetto* project, an experiment in trusting their visitors with the research work of the museum.²⁰ Then-Community Manager Elissa Frankle explained: 'Citizen History opens up

Brain Blog, 11 May 2014, <http://www.billionbrainblog.com/?p=237>. This 'grassroots' activity may previously have been described as 'Participatory Action Research' e.g. Cooper et al., 'Ecology and Society'.

¹⁷ David Weinberger, '[2b2k] Citizen Scientists', *Joho the Blog*, 28 December 2010, <http://www.hyperorg.com/blogger/2010/12/28/2b2k-citizen-scientists/>. This statement was made after an exchange which began after Weinberger was quoted saying that participants 'are not doing the work of scientists. They are doing the work of scientific instruments' in Alex Wright, 'Managing Scientific Inquiry in a Laboratory the Size of the Web', *The New York Times*, 27 December 2010, sec. Science, <http://www.nytimes.com/2010/12/28/science/28citizen.html?pagewanted=all>.

¹⁸ Chris Lintott, "'Citizen' Science and 'Real' Science", *Zooniverse*, 29 December 2010, <http://blog.zooniverse.org/2010/12/29/citizen-science-and-real-science/>.

¹⁹ This may be because stakeholders are reluctant to rely on contributors undertaking more complex tasks, given the difficulties of predicting levels of engagement in advance.

²⁰ The earliest use of the term I could find is from the *Children of the Lodz Ghetto* project in 2011.

a museum's existing data to participants and, through scaffolded inquiry, invites participants to draw conclusions to answer big questions'.²¹ In a 2013 talk, Frankle clarified further: citizen history 'engages amateur scholars and enthusiasts in contemplating and answering authentic questions' based on authoritative research and resources held by the museum while 'being open to the new ideas, questions, and ways of thinking brought by these new collaborators'.²²

Historical skills

The various definitions of historians and historical thinking outlined in the *Introduction* present historians as people who can not only use and understand sources and the information they convey, but interpret them and make explicit historical arguments about them. This requires the development of specific skills, and here I turn again to the American Historical Association (AHA)'s articulation of historical skills. The AHA lists six 'core competencies' for students in history courses and degree programs: the ability to engage in historical inquiry, research, and analysis; to practice historical empathy; to understand the complex nature of the historical record; to generate significant, open-ended questions about the past and devise research strategies to answer them; to craft historical narrative and argument; and to practice historical thinking.²³ Jordanova posits

Romeo and Blaser discuss citizen science and historians but not 'citizen historians' as such. Frankle, 'More Crowdsourced Scholarship: Citizen History'. Romeo and Blaser, 'Bringing Citizen Scientists and Historians Together'.

²¹ Frankle, 'More Crowdsourced Scholarship: Citizen History'.

²² Elissa Frankle, 'Making History with the Masses: Citizen History and Radical Trust in Museums', 2013, <http://mith.umd.edu/podcasts/making-history-with-the-masses-citizen-history-and-radical-trust-in-museums/>. The second part of this statement suggests collaborative rather than contributory projects; further research into the impact of project structures on participants and institutions would be useful.

²³ American Historical Association, 'AHA History Tuning Project: History Discipline Core', *American Historical Association*, September 2013, <http://www.historians.org/teaching-and->

three inseparable groups of skills: technical skills (e.g. palaeography); source-based skills related to 'the finding and evaluating' of appropriate sources relevant to an identified historical problem; and interpretative skills, including the 'ability to construct a plausible argument'.²⁴

Engagement

In order to think about the ways in which crowdsourcing can support deeper engagement with historical materials or questions, some definitions of 'engagement' encountered within academia, such as 'the myriad of ways in which the activity and benefits of higher education and research can be shared with the public', must be set aside.²⁵ This conceptualisation of engagement as a unidirectional broadcast of information to the public is more limited in ambition than the inherently participatory, multidirectional definition used in this thesis. Similarly, many definitions of 'engagement' in the arts and cultural heritage merely mean physical attendance at events or venues. However, a more nuanced model comes from the United Kingdom's governmental agency, Department for Culture, Media, and Sport's *Culture and Sport Evidence* programme. This research defines four types of engagement, each building on the previous level: 'attending' to content (i.e. paying conscious, intentional attention); 'participating' through an interaction that contributes to the creation of content; 'deciding' by making decisions about the delivery of resources for content creation; and 'producing' through creating content 'which has a public economic impact'.²⁶ From a different angle, psychologist Stephen Bitgood and

learning/current-projects/tuning/history-discipline-core.

²⁴ Jordanova, *History in Practice*, 2006. p. 151.

²⁵ 'What Is Public Engagement?', *National Co-ordinating Centre for Public Engagement*, accessed 5 March 2013, <https://www.publicengagement.ac.uk/what>.

²⁶ The CASE programme, 'Evidence of What Works: Evaluated Projects to Drive up Engagement' (London: Culture and Sport Evidence (CASE) programme, January 2011), http://www.culture.gov.uk/images/research/evidence_of_what_works.pdf. The microtasks used in

colleagues undertook a comprehensive review of relevant psychology and museum research and defined engagement as involving 'deep sensory-perceptual, mental and/or affective involvement with exhibit content'.²⁷ At the very least, engagement through heritage crowdsourcing requires paying conscious, intentional attention to source materials and participating in some tasks related to historical sources or research.

In citizen science, 'Levels of Engagement' may refer to a model for increasing levels of participant agency. Raddick et al.'s 'Levels of Engagement' in citizen science references the concept of legitimate peripheral participation (LPP), the process of moving from 'being a newcomer and outsider to becoming an insider to a set of practices'²⁸ within a community of practice. In Raddick et al.'s model, based on their experience with the original *GalaxyZoo* blog, forum and 'volunteer-initiated projects', the first level of engagement is participating in the core microtask of a project (for example, classifying galaxies). The second involves joining in community discussion through forums,²⁹ while participants at the third level are 'working independently on self-identified research projects' related to the overall project.³⁰ Bonney et al.'s framework for Public Participation in Scientific Research (PPSR) can also be mapped to deepening levels of engagement with the skills, practices and concerns of a discipline.³¹ Another potential model for the transition from

some crowdsourcing projects mean that participants can 'produce' content without necessarily 'deciding' - for example, in a 'type what you see' transcription project - but most tasks involve some level of decision-making.

²⁷ Stephen Bitgood, 'An Attention-Value Model of Museum Visitors' (Center for the Advancement of Informal Science Education, 2010), http://caise.insci.org/uploads/docs/VSA_Bitgood.pdf.

²⁸ Mugar et al., 'Planet Hunters and Seafloor Explorers'.

²⁹ As in the *Introduction*, 'forum' stands in for all forms of online messageboard or social network.

³⁰ M. Jordan Raddick et al., 'Citizen Science: Status and Research Directions for the Coming Decade', in *Astro2010: The Astronomy and Astrophysics Decadal Survey*, vol. 2010, 2009, <http://www8.nationalacademies.org/astro2010/DetailFileDisplay.aspx?id=454>.

³¹ This research is otherwise best known for its categorisation of projects as 'contributory',

casual participation to deeper engagement comes from research into 'casual' and 'serious' leisure.³² Serious leisure, or the long-term 'systematic pursuit' of a voluntary activity, leads to the acquisition and expression of a combination of 'special skills, knowledge, and experience'.³³ Many items in Bonney et al.'s list of 'processes, steps, or activities' the public can do in scientific investigations have potential parallels in historical investigations, whether collecting data, information and resources, developing possible answers to questions, defining research questions, designing observational data collection methodologies, analysing and interpreting data, drawing and disseminating conclusions, or 'asking new questions'.³⁴

Learning and skills development

Literature on learning and skills development is drawn from both citizen science and history. Cooper et al. summarise evidence that citizen science provides informal learning experiences that improve science literacy, and that participants in some projects have been engaged in 'inquisitive thinking, and increased their ability to frame questions scientifically'.³⁵ Recent research on learning through participation in online citizen science found that participants learn in four areas directly related to the project: they learn how to do tasks, acquire pattern recognition skills, learn information about the topic, and improve

'collaborative' or 'co-creative' based on the tasks that participants are involved in and how much control that participants they have over different steps'. Bonney et al., 'Public Participation in Scientific Research'.

³² Serious leisure is 'sufficiently substantial and interesting for the participant to find a career there'. Conversely, casual leisure is 'immediately, intrinsically rewarding, relatively short-lived pleasurable activity requiring little or no special training to enjoy it'. Stebbins, 'Casual Leisure'.

³³ Stebbins, 'Casual Leisure'.

³⁴ Bonney et al., 'Public Participation in Scientific Research'. Experimental data collection methodologies and hypotheses development are less relevant to history, although hypotheses development could have parallels to developing historical arguments.

³⁵ Cooper et al., 'Ecology and Society'.

their scientific literacy.³⁶ Crowdsourcing projects enable learning by providing a context for participants to learn important 'keywords, concepts or references' and other knowledge and skills related to the discipline.³⁷ Providing a space outside participants' usual professional and social worlds might also be a factor: Csikszentmihalyi and Hermanson state that supportive environments with meaningful activities in which learners are free of anxiety and 'other negative mental states' can motivate learning.³⁸

Some research on intrinsically motivated learning relates it to Csikszentmihalyi's concept of 'flow'.³⁹ Flow is an inherently enjoyable state of deep focus or engagement reached when 'the challenges of the task meet the person's skills'.⁴⁰ If either skills or challenges increase, the other must also increase in order to maintain the flow state; the desire to achieve a state of flow could explain why people seek out more challenging tasks as they learn the skills required for simpler tasks. In their research on flow and learning in museums, Csikszentmihalyi and Hermanson posit that flow activities lead to 'personal growth' because people seeking to return to the inherently enjoyable flow state will seek greater challenges.⁴¹ Further evidence for the role of flow in motivating self-development comes from open source software. Lakhani and Wolf suggest that volunteer programmers on open source projects select projects that 'match their skill levels with task difficulty' in

³⁶ Laure Kloetzer et al., 'Learning by Volunteer Computing, Thinking and Gaming: What and How Are Volunteers Learning by Participating in Virtual Citizen Science?', in *Changing Configurations of Adult Education in Transitional Times*. (Triennial European Research Conference of the European Society for Research on the Education of Adults (ESREA), Berlin, Germany, 2013), 73–93.

³⁷ Kloetzer et al., 'Learning by Volunteer Computing, Thinking and Gaming'.

³⁸ Csikszentmihalyi and Hermanson, 'Intrinsic Motivation in Museums'.

³⁹ Csikszentmihalyi, *Flow: The Psychology of Optimal Experience*.

⁴⁰ Csikszentmihalyi and Hermanson, 'Intrinsic Motivation in Museums'.

⁴¹ Csikszentmihalyi and Hermanson, 'Intrinsic Motivation in Museums'.

order to seek a state of flow,⁴² and Oreg and Nov found that the 'desire for self-development' and 'enhancing one's abilities and skills'⁴³ was related to flow and 'internal needs for growth and self-actualization'.⁴⁴ Stebbins' research on casual and serious leisure also addresses the need for a worthy challenge to reach a state of flow.⁴⁵ While casual leisure requires 'virtually no skill and only minimal knowledge' (also a description of the ideal microtask), it also lacks a 'substantial challenge', without which the flow state will not be reached.⁴⁶

Scaffolding is a social process in which an expert helps a learner solve a problem, perform a task or achieve a goal by reducing the complexity or number of elements the learner has to manage.⁴⁷ Scaffolding methods include providing cues or hints, providing feedback while the task is being undertaken, asking the learner to reflect on their approach,⁴⁸ modelling desired behaviours, offering explanations, and generating questions and

⁴² Lakhani and Wolf, 'Why Hackers Do What They Do'.

⁴³ This self-development included 'learning from others in the field' and receiving feedback on work.

⁴⁴ Shaul Oreg and Oded Nov, 'Exploring Motivations for Contributing to Open Source Initiatives: The Roles of Contribution Context and Personal Values', *Computers in Human Behavior* 24, no. 5 (September 2008): 2055–2073, doi:10.1016/j.chb.2007.09.007.

⁴⁵ This draws on Csikszentmihalyi and LeFevre (1989). Stebbins, 'Casual Leisure'.

⁴⁶ Stebbins, 'Casual Leisure'.

⁴⁷ David Wood, Jerome S. Bruner, and Gail Ross, 'The Role of Tutoring in Problem Solving', *Journal of Child Psychology and Psychiatry, and Allied Disciplines* 17, no. 2 (1976): 89–100, <http://www.ncbi.nlm.nih.gov/pubmed/932126>.

⁴⁸ Koos Winnips and Catherine McLoughlin, 'Six WWW Based Learner Supports You Can Build.', in *Proceedings* (ED-Media 2001 World Conference on Educational Multimedia, Hypermedia & Telecommunications, Tampere, Finland, 2001), <http://eric.ed.gov/?id=ED466231>. as summarised in Catherine E. McLoughlin and Sultana Lubna Alam, 'A Case Study of Instructor Scaffolding Using Web 2.0 Tools to Teach Social Informatics', *Journal of Information Systems Education* 25, no. 2 (2014): 125, <https://www.questia.com/library/journal/1G1-406164672/a-case-study-of-instructor-scaffolding-using-web-2-0>.

comments.⁴⁹ Scaffolds are gradually 'faded', or reduced, as the learner becomes more competent, and are eventually removed.⁵⁰ Scaffolding is designed for face-to-face educational environments where teachers can monitor student performance and adjust lessons accordingly. However, some principles have been extrapolated to guide the design of task interfaces online,⁵¹ and communities of practice may be able to provide some scaffolding for participants learning new skills through crowdsourcing projects.

In my analysis of these projects, I focused on the role of participant communities in supporting learning and skills development.⁵² The educational theory of 'situated learning' is useful for explaining how these communities could encourage informal education. Situated learning provides a series of concepts useful for understanding how communities support participants in learning disciplinary skills. The concept of 'communities of practice', formed by people engaged in 'collective learning in a shared domain' of work,⁵³

⁴⁹ Catherine McLoughlin, 'Learner Support in Distance and Networked Learning Environments: Ten Dimensions for Successful Design', *Distance Education* 23, no. 2 (2002): 149–162, <http://www.tandfonline.com/doi/abs/10.1080/0158791022000009178>.

⁵⁰ Priya Sharma and Michael J. Hannafin, 'Scaffolding in Technology-Enhanced Learning Environments', *Interactive Learning Environments* 15, no. 1 (April 2007): 27–46, doi:10.1080/1049482060096972.

⁵¹ Nina Simon's influential work on the participatory museum also draws on scaffolding: the 'best participatory experiences are not wide open. They are scaffolded to help people feel comfortable engaging in the activity'. Nina K. Simon, 'Chapter 1: Principles of Participation', in *The Participatory Museum*, 2010, <http://www.participatorymuseum.org/chapter1/>.

⁵² For example, research investigating learning in citizen science projects including *Old Weather* found that learning outcomes appear to be most strongly related to participant engagement in 'the life of the community', or social aspects of the project. Kloetzer et al., 'Learning by Volunteer Computing, Thinking and Gaming'.

⁵³ The term 'communities of practice' was coined by Jean Lave and Etienne Wenger for their 1991 book, *Situated Learning. Legitimate peripheral participation*, Cambridge: University of Cambridge Press. The quotes come from Wenger-Trayner's (as he is now) homepage. Etienne Wenger-Trayner, 'Communities of Practice: A Brief Introduction', *Wenger-Trayner*, accessed 3 June 2014,

provides theoretical grounding for the behaviours observed in crowdsourcing participant communities. A community of practice can be viewed as 'a social learning system'⁵⁴ where newcomers 'learn and acquire knowledge' through everyday interactions with and observations of others.⁵⁵ Like scaffolding, situated learning emerged from analyses of in-person interactions, but can also be applied to online communities. Crowdsourcing project forums support many of the behaviours considered to be typical of communities of practice, including problem solving, requests for information, seeking the experience of past behaviours, coordinating actions, documenting shared knowledge and experiences, and discussing developments.⁵⁶ Text-based communities such as online forums tend to have searchable archives that provide examples of past experiences. These archived discussions also provide models for problem-solving and social interactions, contributing to a shared practice around a common 'domain of interest',⁵⁷ which additionally helps orient newcomers to the practices of the group.⁵⁸ Some forums actively compile documentation and maintain indexes to past discussions to help create a 'shared repertoire of resources'.⁵⁹ This use of forums to coordinate and manage the process of creating collaborative resources potentially provides models for other scholarly

<http://wenger-trayner.com/theory/>.

⁵⁴ Etienne Wenger, 'Communities of Practice and Social Learning Systems: The Career of a Concept', in *Social Learning Systems and Communities of Practice* (Springer Verlag and the Open University, 2010).

⁵⁵ Claire McInerney, 'Knowledge Management and the Dynamic Nature of Knowledge', *Journal of the American Society for Information Science and Technology* 53, no. 12 (2002): 1009–1018, <http://onlinelibrary.wiley.com/doi/10.1002/asi.10109/full>.

⁵⁶ Wenger-Trayner, 'Communities of Practice'.

⁵⁷ Wenger-Trayner, 'Communities of Practice'.

⁵⁸ Mugar et al. also found that participants benefit from 'finding questions left by other participants that are similar to their own questions'. Mugar et al., 'Planet Hunters and Seafloor Explorers'.

⁵⁹ Wenger-Trayner, 'Communities of Practice'. The *Old Weather* forums may be the most impressive examples of this, with hundreds of pages of documentation and related summaries.

collaborations. Appropriately for a field such as history that relies on an apprenticeship model to train researchers, 'cognitive apprenticeship' is another model for 'situated learning'. As seen in crowdsourcing forums, cognitive apprenticeship supports the acquisition of disciplinary skills, 'enabling students to acquire, develop, and use cognitive tools' while engaging in meaningful tasks.⁶⁰ In this theory, exposure to the practices of others through social interactions is vital for the 'social construction of knowledge'.⁶¹

The legitimate peripheral participation (LPP) evident in crowdsourcing forums is another form of situated learning. In LPP, novices begin by engaging in 'simple practices', a level of participation that makes them 'legitimate but peripheral members of the community' while offering opportunities for observing more experienced members of a community at work. This observation helps newcomers learn the 'tasks, vocabulary and organizational principles of the community'.⁶² The ability to observe the practices of experts is important for helping newcomers understand the context into which their work fits.⁶³ In crowdsourcing, simple microtasks provide an ideal opportunity for participants to legitimately observe and perhaps join communities of practice. Citizen science researchers Mugar et al. identified four modes of LPP in the citizen science projects *Sea Explorer* and *Planet Hunter*, from annotating data objects (e.g. categorising images), 'user-generated

⁶⁰ John Seely Brown, Allan Collins, and Paul Duguid, 'Situated Cognition and the Culture of Learning', *Educational Researcher* 18, no. 1 (1989): 32-42, <http://edr.sagepub.com/content/18/1/32.short>.

⁶¹ Brown, Collins, and Duguid, 'Situated Cognition and the Culture of Learning'. The term 'apprenticeship' was chosen in part to emphasise the 'inherently context-dependent, situated, and enculturating nature of learning', and the central role of active participation in learning.

⁶² Mugar et al., 'Planet Hunters and Seafloor Explorers'.

⁶³ Susan L. Bryant, Andrea Forte, and Amy Bruckman, 'Becoming Wikipedian: Transformation of Participation in a Collaborative Online Encyclopedia', in *Proceedings of the 2005 International ACM SIGGROUP Conference on Supporting Group Work* (ACM, 2005), 1-10, <http://dl.acm.org/citation.cfm?id=1099205>.

annotation' (e.g. adding optional free-text tags), 'asking questions about data objects', and finally participating in 'higher-level analysis', typically via 'talk' pages for objects or forums.⁶⁴ They found that each of these modes helps participants learn 'what characteristics in the data they should pay attention to'.⁶⁵

Mugar et al.'s work on 'practice proxies', (online traces of user interactions that provide insight into other people's work practices),⁶⁶ may be useful for disciplines such as history which tend to rely on tacit knowledge when assessing sources and arguments from others. This tacit knowledge was traditionally passed on through scholarly apprenticeship, but the articulation of expertise in online communities of practice may provide an alternative method for capturing and sharing this expertise. Again, the ability of online discussions and tasks to provide an archive of past practices helps newcomers benefit from the work of more expert participants.

Curiosity

Curiosity is the desire 'to know, to see, or to experience', which is intrinsically linked to learning.⁶⁷ Acquiring knowledge in response to curiosity 'is considered intrinsically rewarding and highly pleasurable', and curiosity motivates 'exploratory behaviour' aimed at acquiring new information.⁶⁸ Educators may attempt to induce curiosity by modelling

⁶⁴ Mugar et al., 'Planet Hunters and Seafloor Explorers'. These analytical discussions are 'often stimulated by a hypothesis or observations about data objects' made by participants then discussed on the forum.

⁶⁵ Mugar et al., 'Planet Hunters and Seafloor Explorers'.

⁶⁶ For example, looking at classification or tagging decisions made by other participants in a project. Mugar et al., 'Planet Hunters and Seafloor Explorers'.

⁶⁷ Jordan Litman, 'Curiosity and the Pleasures of Learning: Wanting and Liking New Information', *Cognition & Emotion* 19, no. 6 (September 2005): 793–814, doi:10.1080/02699930541000101.

⁶⁸ Litman, 'Curiosity and the Pleasures of Learning'.

it, by creating a hook with a 'thought-provoking question or surprising statement' or introducing 'curiosity-arousing elements' such as incongruity, contradictions, novelty, surprise, complexity or uncertainty.⁶⁹ I discuss below how selected projects have enabled an initial question or curiosity to grow into an interest in history. The role of curiosity about the next item in a task queue in motivating crowdsourcing participants has been discussed previously,⁷⁰ but a project's ability to nurture curiosity about the items shown to participants is also important. As the examples discussed in this chapter show, a diverse community can provide a range of skills and perspectives. A critical mass of friendly, constructive (i.e. invested in the success of the project) community members helps amplify curiosity.

Citizen science projects provide several examples of moments of curiosity leading to serendipitous but significant discoveries. One prominent example is *Galaxy Zoo's* 'green peas', a new class of galaxy discovered by a group of posters in the Galaxy Zoo forum.⁷¹ As described by one forum participant, one member 'found a little green galaxy and she started a thread called "Give peas a chance" with peace spelled p- e-a-s and everyone thought it was funny and we started collecting these green blobs'.⁷² In the spirit of collective enquiry, some members of the community formed a 'Peas Corp',⁷³ collecting over

⁶⁹ Marilyn P. Arnone, 'Using Instructional Design Strategies To Foster Curiosity' (Syracuse, NY: ERIC Clearinghouse on Information and Technology, August 2003), <http://www.ericdigests.org/2004-3/foster.html>.

⁷⁰ Ridge, 'Playing with Difficult Objects: Game Designs for Crowdsourcing Museum Metadata'.

⁷¹ Arfon Smith, 'Zooniverse, GitHub and the Future', *Zooniverse*, 13 August 2013, <http://blog.zooniverse.org/2013/08/13/zooniverse-github-and-the-future/>.

⁷² Galaxy Zoo volunteer quoted in Charlene Jennett et al., 'Creativity in Citizen Cyber-Science: All for One and One for All' (WebSci'13, Paris, France, 2013).

⁷³ Forum posters, 'Give Peas a Chance!', *Galaxy Zoo Forum*, 2007, <http://www.galaxyzooforum.org/index.php?topic=3638.o>.

100 'peas' that were later analysed by scientists.⁷⁴ Similar objects had been discussed previously, so the community was already primed to think about the technical characteristics of these 'green blobs', and some used computational techniques on the data published by the project to collect further examples.⁷⁵ For the *Zooniverse's* Arfon Smith, this marked the moment that 'citizen-led science' began in the *Zooniverse*.⁷⁶

Examples of citizen history projects

Examples of some significant projects will provide valuable context for later discussions. The examples were chosen following the review of participatory history projects discussed in *Chapter 1*. I was looking for projects with one or more of the following factors: participant behaviours that indicated deeper engagement, skills development or the emergence of new research questions; substantial, accessible histories (for example, archived forums or interfaces) and sufficient productivity and levels of participation to provide material for analysis; and the ability for me to undertake the site tasks and review contributed content. An additional factor was the availability of supplementary material such as informal communications (blog or social media posts, newsletters) or articles providing further background and insight into communication strategies. In some cases I have also had access to project stakeholders to clarify questions that arose during my

⁷⁴ Carolin Cardamone et al., 'Galaxy Zoo Green Peas: Discovery of a Class of Compact Extremely Star-Forming Galaxies', *Monthly Notices of the Royal Astronomical Society* 399, no. 3 (November 2009): 1191–1205, doi:10.1111/j.1365-2966.2009.15383.x.

⁷⁵ These posts give a sense of the role of the wider community investigating the 'peas': Forum poster, 'Re: Peas Project', *Galaxy Zoo Forum*, 5 November 2008, <http://www.galaxyzooforum.org/index.php?topic=270633.msg218401#msg218401>. Alice Sheppard, 'Peas in the Universe, Goodwill and a History of Zooite Collaboration on the Peas Project', *Galaxy Zoo*, 7 July 2009, <http://blog.galaxyzoo.org/2009/07/07/peas-in-the-universe-goodwill-and-a-history-of-zooite-collaboration-on-the-peas-project/>.

⁷⁶ Smith, 'Zooniverse, GitHub and the Future'.

analysis or for discussion that informed my analysis.⁷⁷ I chose projects with a range of approaches in order to consider the attributes that they had in common, despite their outward differences in project age, platforms or disciplinary ties.

Long-established or hugely popular crowdsourcing projects like *Herbaria@Home* (established in 2006), *FamilySearch Indexing* (2007), *Galaxy Zoo* (2007) and other science projects from the Zooniverse group including *Old Weather* (2010) have had enough time and participants to demonstrate ways in which project participants can develop new skills and knowledge as a result of their growing interest in the project source material, or graduate to more complex tasks or bigger responsibilities. The United States Holocaust Memorial Museum's *Children of the Lodz Ghetto* and *Marine Lives* have been intentionally designed to 'encourage more people to become historians',⁷⁸ or at least practice historical thinking and skills. *Operation War Diary*'s front page proclaims it to be a citizen history project. *Quench* represents an experiment in scientists inviting participants to 'take part in the ENTIRE scientific process', including writing scientific articles.⁷⁹

Herbaria@Home participants began transcribing historical herbarium sheets in 2006. By 2008 their work had inspired an 'ongoing project to collect biographies of historical botanists'⁸⁰ and participants were posting short biographies of collectors on a wiki.⁸¹ By

⁷⁷ Here I owe particular thanks to project stakeholders for responding to emailed or in-person questions, and to Stuart Dunn for inviting me to two workshops on 'humanities crowdsourcing' he ran in May and October 2012 with Mark Hedges as part of the AHRC-funded research project.

⁷⁸ Frankle, 'More Crowdsourced Scholarship: Citizen History'.

⁷⁹ Laura Trouille, 'Galaxy Zoo Quench – Experience the Full Scientific Process', *Galaxy Zoo*, 10 July 2013, <http://blog.galaxyzoo.org/2013/07/10/galaxy-zoo-quench-experience-the-full-scientific-process/>.

⁸⁰ Botanical Society of Britain & Ireland, 'Herbaria', *Botanical Society of Britain & Ireland*, accessed 22 June 2014, <http://www.bsbi.org.uk/herbaria.html>.

2010, the site had been developed so that structured information stored in wiki 'info boxes' could be linked to forum posts.⁸² Familiarity with handwriting, dates and places collected helps the project by providing useful context for deciphering text⁸³ and identifying collectors.⁸⁴ Working on collective biographies meant finding, debating and comparing evidence from the herbarium sheets and historical documents.⁸⁵ Participants reported making some significant discoveries⁸⁶ while developing new interests and skills. Unlike more recent projects, it does not have a publicity machine, slick interface and targeted calls to action, but having a programmer available to modify systems seems to have enabled the project to respond to participants' interests. In this case, a wiki that supports both structured data and free text has helped maximise the impact of data recorded across the whole site.

Old Weather also demonstrates the value of a participant forum and of close, active exposure to historic documents. While indexing the ships' logs, participants became interested in daily life on-board, the voyages of the ships and the things they encountered along the way, from sighting other ships to animals. They have documented their

⁸¹ Forum posters, 'Biographies', accessed 22 June 2014,

<http://herbariaunited.org/core/forum/viewtopic.php?t=917>.

⁸² Tom Humphrey, 'Linking Collectors with Message Board Discussions', *Herbariaunited.Org Forum Index -> Botanists*, 10 March 2010, <http://herbariaunited.org/core/forum/viewtopic.php?t=3441>.

Tom Humphrey, 'Updated Version of the Doc. Form, -- Please Help Test It!', *Herbariaunited.Org*, May 2008, <http://herbariaunited.org/core/forum/viewtopic.php?t=1040>.

⁸³ 'Feedback Request: Cardamine Flexuosa (21386)', accessed 22 June 2014, <http://herbariaunited.org/forum/viewtopic.php?t=2613>.

⁸⁴ Tom Humphrey, 'Request for Examples of Handwriting', *Herbariaunited.Org Forum Index -> Botanists*, 10 March 2010, <http://herbariaunited.org/core/forum/viewtopic.php?t=3442>.

⁸⁵ See for example Forum posters, 'Doncaster & Huntingdon Collector', *Herbaria@home*, 2010, <http://herbariaunited.org/core/forum/viewtopic.php?t=5627>.

⁸⁶ Various, 'AHRC Crowd Sourcing Study: Scoping Seminar'. By May 2015, 70 biographies had been posted on the related site meiosis.org.uk.

discoveries on forum threads with titles like 'Worse things happen at sea' or 'Riveting Log Entries' that run for (so far) over five years and 150 pages of posts. One participant mentioned one advantage of a well-populated forum: 'when someone posts a question, it's almost a competition to answer it first'.⁸⁷ Participants have used the forum to assemble detailed documentation for newer participants, including guidance for individual ships, and lists of places and naval jargon; this also led to the creation of a Compendium of Maritime and Weather Information.⁸⁸ They have compiled ships' histories for the website *Naval-History.net* from transcribed logs and other sources.⁸⁹ Participants have also followed their initiative in investigating the relationship 'between the 'Number on Sick List' section of the log and the well-known 'Spanish flu' outbreak in 1918', suggesting the forum community supports the development of new research questions.⁹⁰ For at least one participant, having the ships' histories project to work on provided an alternative to working on a different set of logs added at a later stage of the project.⁹¹ This last point

⁸⁷ Statement made at Various, 'AHRC Crowd Sourcing Study: Scoping Seminar'.

⁸⁸ Old Weather Moderators, Editors and Volunteers, 'Compendium of Maritime and Weather Information', accessed 25 June 2014, <http://www.naval-history.net/OWShips-ForumMat.htm>. Participants have also compiled a 'web based searchable database of ship names, places, and people mentioned in the log books of our ships', complete with data dictionary. Forum posters, 'Searchable Database', *Old Weather Forum*, June 2012, <http://forum.oldweather.org/index.php?topic=2872.0>. Kathy, 'Database Data Dictionary', *Old Weather Forum*, 11 June 2012, <http://forum.oldweather.org/index.php?topic=2873.0>.

⁸⁹ Philip Brohan, 'New Crew Members at Naval-History.Net', *Old Weather Blog*, 23 May 2014, <http://blog.oldweather.org/2014/05/23/new-crew-members-at-naval-history-net/>.

⁹⁰ Philip Brohan, 'HMS Africa in Action against Orthomyxoviridae', *Old Weather Blog*, 12 January 2011, <http://blog.oldweather.org/2011/01/12/hms-africa-in-action-against-orthomyxoviridae/>.and Romeo and Blaser, 'Bringing Citizen Scientists and Historians Together'. Operation War Diary participants may also have found evidence for the first outbreaks in Forum posters, 'Epidemic', *Talk Operation War Diary*, 2014, http://talk.operationwardiary.org/#/boards/BWDoooooob/discussions/DWDooooozef?page=1&comment_id=52e8obaef11e0413040001dc.

⁹¹ Statement made at Various, 'AHRC Crowd Sourcing Study: Scoping Seminar'.

illustrates a tension apparent in other *Zooniverse* projects: some participants will spend more time on the forum or on side projects than on the original task, meaning that some of the most skilled participants are no longer contributing to the core goal of the project.

Operation War Diary (OWD) was launched in January 2014 to transcribe military unit diaries from the First World War⁹² and is included because of its ambitions to be a citizen history project.⁹³ One of the most prominent lines on the front page says 'Working together with Citizen Historians during the First World War Centenary' (a claim that will be examined in this chapter). The site and its 'Talk' forum are based on *Zooniverse* software. The *Zooniverse* project as a whole has a large pool of over 1 million volunteers who can be notified of new projects,⁹⁴ but it is unclear how many of those are interested in historical projects. The (at the time) upcoming centenary of the First World War undoubtedly helped with publicity, and the *OWD* blog reported that 11 days after launch, 135,000 pages had been classified by thousands of volunteers.⁹⁵ By June 2014 there were 9,200 registered users, with 86,000 completed pages overall and 200 active users

⁹² The analysis presented here dates from the pre-launch beta in December 2013 and the public site from January to June 2014, and has tried to take account of changes over that time. As the forum has been restructured since I began collecting data, some links will no longer work but Zotero snapshots are available upon request.

⁹³ An additional reason for inclusion is that it, and the related *Lives of the First World War*, is an example of the difficulties in multi-stakeholder projects, and of copyright and commercialisation issues with commercially-valuable family history sources. Partners for *OWD* are *Zooniverse*, the Imperial War Museum (IWM) and The National Archives for *OWD*, and IWM and *Findmypast/DC Thomson Family History* for *Lives of the First World War*. *Lives of the First World War* is a historical, prosopographical project rather than an educational one.

⁹⁴ Robert Simpson, 'One Million Volunteers', *Zooniverse*, 14 February 2014, <http://blog.zooniverse.org/2014/02/14/one-million-volunteers/>.

⁹⁵ Grant Miller, 'Saturday Status – Operation War Diary', *Daily Zooniverse*, 25 January 2014, <http://daily.zooniverse.org/2014/01/25/saturday-status-operation-war-diary/>.

transcribing 1000 pages per week.⁹⁶

The skills required to participate in *Children of the Lodz Ghetto* go beyond the 'type what you see' model of most transcription projects. Unlike projects concerned with re-mediating data into digital formats as efficiently as possible, the project is designed to 'encourage more people to become historians, or at least make history and historical thinking more accessible to participants'.⁹⁷ The project has a specially designed workspace to support participants in the task of researching the fates of over 14,000 children from the Ghetto.⁹⁸ Participants are prompted to decide the relevance of historical materials to the question at hand, and to reflect on their decision-making process. It requires participants to develop source-based skills and to link evidence to arguments. The workspace interface provides a form of scaffolding by breaking the process down into smaller tasks with data entry fields and prompts, divided into five chronological sections and tailored to the task and information sought at each stage. Scaffolding in the form of personal feedback on specific tasks is also provided by the Community Manager, who is responsible for both checking records for accuracy,⁹⁹ and encouraging citizen historians as they iteratively learn the methods for each stage in the process of 'moving from a question to a data point to a narrative'.¹⁰⁰ Selected participants can also vet others' work,¹⁰¹ but the Community Manager

⁹⁶ Jim O'Donnell, Operation War Diary stats, interview by Mia Ridge, 29 June 2014.

⁹⁷ Frankle, 'More Crowdsourced Scholarship: Citizen History'.

⁹⁸ Frankle, 'More Crowdsourced Scholarship: Citizen History'.

⁹⁹ Accuracy is a particular concern for moderators 'since inaccuracies are often used to fuel the fires of Holocaust denial'. Frankle, 'More Crowdsourced Scholarship: Citizen History'.

¹⁰⁰ Frankle, 'More Crowdsourced Scholarship: Citizen History'.

¹⁰¹ Nina K. Simon, 'Participatory Design and the Future of Museums', in *Letting Go? Sharing Historical Authority in a User-Generated World*, ed. Bill Adair, Benjamin Filene, and Laura Koloski (Left Coast Press, 2012).

role is still immensely time-consuming.¹⁰² Participating in *Children of the Lodz Ghetto* does not offer any instant rewards - the research process requires learning about the transliteration of Yiddish and Polish names, assessing the relative locations of homes and schools, devising and applying source search strategies, and evaluating possible matches against the evidentiary requirements of the project. However, it is clearly a valuable task, and the call to remember or find out the fate of a child is emotionally powerful. The moderator feedback visible in forum posts shows both the patience required to convey and convince newcomers to act on the project's concern for accuracy, and the educational value of being gently challenged to practice and reflect on the research skills that may help create new citizen historians.

Marine Lives aims to create a public, academic edition of records from the High Court of Admiralty, London, 1650-1669, by transcribing, linking and enhancing original manuscript documents. They run mini-projects like the twelve-week 'Summer Programme', which aims to digitise one archival volume while developing the research skills of programme participants by 'developing short biographies, thematic analyses, and geographical profiles' from Admiralty Court material.¹⁰³ Rather than developing a highly specialised software interface, the project uses a range of off-the-shelf technologies and provides the scaffolding necessary to learn historical research skills through personal interactions. Participants work in 'small online groups of three to four people' with 'experienced team facilitators'.¹⁰⁴ The project also provides activities for those who do not wish to learn historical skills but might want to help visualise the data already transcribed to 'identify

¹⁰² Elissa Frankle et al., *Children of the Lodz Ghetto* project staff members interview, interview by Mia Ridge, 18 September 2014.

¹⁰³ Colin Greenstreet, 'Summer Programme 2014', *The Shipping News*, 7 May 2014, <http://marinelives-theshippingnews.org/blog/2014/05/07/summer-programme-2014/>.

¹⁰⁴ Greenstreet, 'Summer Programme 2014'.

and confirm trends and patterns in seventeenth century commercial life'.¹⁰⁵

The *Civil War Pathways* project had a quite complex and open-ended 'distributed reading' task in which volunteer research assistants were assigned large amounts of primary source material to read for information on specific topics. Project staff trained 300 participants to use *Omeka* and provided some background on 'the topics and themes of the Civil War' through day-long workshops; about 200 participants went on to contribute 2800 scanned and described records to the project.¹⁰⁶ The project also created a network of people who could answer specialist historical questions based on their reading.¹⁰⁷

How do crowdsourcing projects provide spaces for learning historical skills?

The literature discussed earlier suggests that crowdsourcing projects can provide a space in which participants can learn new skills and apply them to real challenges. Each of the projects mentioned have forums that show signs of participants' curiosity, developing new research questions and learning the historical skills necessary to help answer those questions.

How is citizen history different from citizen science?

While useful lessons can be drawn from citizen science, there are important differences between history and science to consider. Data 'collection' in citizen science often refers to observations of flora or fauna, which may be similar to 'type what you see' tasks like text transcription. Historical research has a lower barrier to entry than most sciences - a

¹⁰⁵ Colin Greenstreet, 'Patterns in Time', *The Shipping News*, 14 May 2014, <http://marinelives-theshippingnews.org/blog/2014/05/15/patterns-in-time/>.

¹⁰⁶ Civil War Pathways, Survey: Crowdsourcing and public participation in digital history.

¹⁰⁷ Civil War Pathways, Survey: Crowdsourcing and public participation in digital history.

newfound interest in history can be developed with access to a public library and digitised archives or locally sourced primary materials, and requires no special equipment or computational abilities.

Platforms developed for science projects and later re-purposed for historical projects may encounter subtle differences in participant motivations and relationship to the source material as well as more obvious differences related to the form of the input items. As mentioned previously, *Operation War Diary (OWD)* illustrates many of the differences between historical and scientific crowdsourcing. The underlying *Zooniverse* citizen science platform began with image classifications for *Galaxy Zoo*, and the validation models developed for these projects work better with discrete data points than with longer passages of text which have potentially wide but valid variations in interpretive markup.¹⁰⁸ Scientific crowdsourcing software is also designed for self-contained items like images of celestial objects. However, individual pages of a diary often have a relationship to preceding or subsequent pages, and these pages may also contain a narrative that spans several entries. As discussed in the previous chapter, interpreting handwriting can be a complex task and a learnt skill that benefits from more exposure to a particular hand. As diary entries and similar documents were designed to be read as a whole rather than as isolated pages, information on an individual page may rely on previous pages for context. Posters on *OWD* threads provide examples of subsequent pages lacking information provided on the first, including references to 'the village', presumably named on an earlier

¹⁰⁸ *OWD* diary pages are indexed through a series of structured tags linked to specific locations on the page (for example, the diary entry date, or locations mentioned within the text). This discrete data, based on pre-determined categories of tags, is more easily validated than free-text transcriptions of the diary - if five participants say a particular section of the page should be recorded as April 6th and one participant says it should be recorded as April 5th, the system can record April 6th with a high level of confidence; similarly if most participants record the relevant 'Unit Activity' as 'resting' rather than being 'under fire'. Unclear tags can be manually reviewed.

page, and dates with no month or year.¹⁰⁹

As part of the historical record, historical documents also represent moments within a broader chronological narrative. This can have a powerful effect on people transcribing pages of military diaries with some knowledge of the battles or deaths facing those units later in the war. While the data generated from source images in citizen science may be more important than the images, this is not true for history. Indexed or transcribed text is immensely useful for finding sources, but access to images of the original document is vital in cases where the transcription is not guaranteed to be complete.¹¹⁰ And unlike scientific data, historical material directly represents at least one real person in the past. This sense of connection to the trace of an individual's voice or handwriting may be why many participants want to tag other activities or items of potential interest, or expect to be able to transcribe the entire page (or at least larger sections of it)¹¹¹ as text rather than record it as individual tags, as discussion on the *OWD* forum shows.¹¹² Forum posts that use terms like "'my' diary"¹¹³ suggest an intimate relationship with the source material and the people it

¹⁰⁹ Forum posters, 'Why Is There No Back Button'.

¹¹⁰ For example, the interests of genealogy companies that index census records might not entirely overlap with the interests of other historians, necessitating access to the original to retrieve additional information.

¹¹¹ For example, one poster said: "'my' diary was written in really personal style [...] it is a pity there isn't a small box to add some 'quotes' thus showing how the soldiers were feeling?" Forum posters, 'What Could We Improve?' This may also be a reaction against the sense that highly structured partial transcriptions reduce the diaries to 'damn statistics' Forum posters, 'Report Transcription', *Talk Operation War Diary The Mess Hall Board / The Diary Pages*, January 2014, <http://talk.operationwardiary.org/#/boards/BWDoooooo6/discussions/DWDoooo2by>.

¹¹² Some earlier discussion threads seem to have been archived or deleted but one thread still available in May 2015 is Forum posters, 'Report Transcription'.

¹¹³ Forum posters, '[FEATURE REQUEST] Ability to Go Back and Change Tags'. The use of 'my object' occurs in many projects. A search on the *Old Weather* forum turns up many posts using the phrase 'my ship'; e.g. some in Forum posters, 'Signs of OW Addiction ...'

represents.

The sequential nature of pages has emotional as well as practical effects. Csikszentmihalyi and Hermanson point out that 'our wish to know about peoples in faraway places' is related to the desire for emotional connection as well as intellectual understanding, and that we are drawn to diaries and personal letters 'because they connect us with another's feelings'.¹¹⁴ Some *OWD* forum posters expressed frustration about their inability to transcribe consecutive pages of diaries in terms of their desire to read the stories contained within those pages. In the words of one *OWD* forum participant, 'Why take away the continuity like that by re-allocating the story to someone else?'.¹¹⁵ 'I spent ages doing the [specific military unit] and got to the last few pages...only to find that someone else has completed it and so I don't know the end of the story...aargh!!'.¹¹⁶ Some participants wanted to follow the diary of particular units¹¹⁷ and were frustrated by the treatment of diary pages as 'isolated packets of data'.¹¹⁸ Other projects have found workarounds - the *Old Weather* forum contains instructions for viewing images for subsequent pages¹¹⁹ - but the license *OWD* negotiated with The National Archives restricts

¹¹⁴ Csikszentmihalyi and Hermanson, 'Intrinsic Motivation in Museums'.

¹¹⁵ Forum posters, '[FEATURE REQUEST] Ability to Go Back and Change Tags'.

¹¹⁶ Forum posters, '60,000 Pages Tagged in 2 Days! You Are Making History – THANK YOU.', *Talk Operation War Diary History Board*, January 2014, <http://talk.operationwardiary.org/#/boards/BWDooooooo3/discussions/DWDoooooopi>.

¹¹⁷ Forum posters, 'Why Is There No Back Button'.

¹¹⁸ Forum posters, '[FEATURE REQUEST] Ability to Go Back and Change Tags'. The official response was that the *Zooniverse* platform is designed to optimise the classification and tagging process, and that the full diaries can be read at The National Archives at Kew or downloaded for a fee. Poster lukesmith on Forum posters, 'Why Is There No Back Button'.

¹¹⁹ Randi, 'How to Look at Log Pages before and after the One You Are Working On', *Old Weather Forum* » Questions and Answers » The Logs and FAQ, 12 February 2013, <http://forum.oldweather.org/index.php?topic=3598.o>.

the availability of the diary images.¹²⁰ Others suggest the participant feels they have earned the right to transcribe the more eventful or concluding pages by slogging through boring pages: 'It's also annoying that you can spend some time tagging repetitive entries of them digging trenches and then the next day [...] someone else has tagged the 'exciting' page where the unit moved!!!'.¹²¹

Participants in other historic crowdsourcing projects also have a preference for transcribing handwritten sources¹²² - this may relate to a sense that their contribution is more meaningful or vital,¹²³ or to a desire for emotional connection. The intimacy of handwritten accounts may also create a greater sense of responsibility in participants.¹²⁴ The personal accounts and artefacts common in heritage data might mean participants are more likely to find something they can relate to in the material they are transcribing. The intimacy of reading handwritten texts, viewing handmade or worn objects or photographs of people is unlike, but possibly as powerful as the sense of awe of nature reported in some scientific projects. The significance of humanities data may also be more immediately

¹²⁰ Forum poster, 'Image AWDooo2keb', *Operation War Diary Talk*, May 2014, <http://talk.operationwardiary.org/#/subjects/AWDooo2keb>.

¹²¹ Forum posters, '[FEATURE REQUEST] Ability to Go Back and Change Tags'.

¹²² For example, some FreeBMD syndicates 'are keen to transcribe handwritten scans' FreeBMD Newsletter, 'RootsWeb: FREEBMD-SYNDICATES-L FreeBMD Newsletter - February 2014', *FREEBMD-SYNDICATES-L Archives*, 8 March 2014, <http://archiver.rootsweb.ancestry.com/th/read/FREEBMD-SYNDICATES/2014-03/1394314857>. Holley commented on transcriber attention to name-rich and handwritten sources seen on the National Archives of Australia's Hive project in Rose Holley, 'National Archives of Australia Embraces Crowdsourcing and Releases "The Hive"', *Rose Holley's Blog*, 10 November 2012, <http://rose-holley.blogspot.co.uk/2012/11/national-archives-of-australia-embraces.html>.

¹²³ For those who know that typed records are more amenable to OCR transcription.

¹²⁴ One OWD forum poster said, 'Oh God. I spent 30 pages misspelling someone's name. [...] I really REALLY wish I could go back and correct it. I feel horrible.' Forum posters, '[FEATURE REQUEST] Ability to Go Back and Change Tags'.

accessible. As one *Old Weather* forum poster said, 'I've tried some of the astronomical sites and they don't "grab" me, I think that the problem is that I don't know enough to build a picture of the data'.¹²⁵ These examples demonstrate a range of reasons why transcribing historical documents might capture participants' interest and encourage them to engage more deeply with the material.

The role of project and task design

Project design has an impact on participants' ability to learn and practice skills at two levels. At the macro level, my analysis suggests that projects are more likely to support skills development if they: provide forums; allow participants to collect and share items; provide access to research data; support the development of curiosity; and provide stepping stones to more complex tasks or greater responsibilities. At the micro level, task design can support learning by reducing anxiety and decreasing cognitive load. Cognitive load theory¹²⁶ may offer some insight into how participants benefit from their experience on a project. The cognitive load of a task decreases as a participant learns the skills required,¹²⁷ freeing up working memory to permit more problem solving.¹²⁸ As experience increases, the newly learnt material is incorporated and more working memory is freed,

¹²⁵ Forum posters, 'Zooniverse News', *Home Port: Welcome to Old Weather => Old Weather News*, to present 2013, <http://forum.oldweather.org/index.php?action=printpage;topic=4013.0>.

¹²⁶ As discussed in the previous chapter, cognitive load is the amount of mental effort required to operate a system or learn new information.

¹²⁷ As skills are learnt, they are grouped into cognitive constructs called 'schemas' that can be processed as one element. An example is the relationship between learning to read individual letters of the alphabet, reading words and understanding entire sentences. Paas, Renkl, and Sweller, 'Cognitive Load Theory and Instructional Design'.

¹²⁸ Fred Paas et al., 'Cognitive Load Measurement as a Means to Advance Cognitive Load Theory', *Educational Psychologist* 38, no. 1 (2003): 63–71, http://www.tandfonline.com/doi/full/10.1207/S15326985EP3801_8.

allowing advanced knowledge and skills to be acquired over many cycles.¹²⁹ Microtasks that reduce complexity by simplifying the number of choices required (and the number of items in working memory) can also help reduce the cognitive load.¹³⁰ For example, the *FamilySearch Indexing* interface is tightly focused on the indexing task, leaving more working memory free for the participant to become engaged with the sources. As participants learn the skills for smaller tasks, they develop more capacity for understanding and can move on to less tightly-defined tasks such as discussion. However, the range of tasks available in the examples I have presented suggests that task design is not the most significant factor in learning historical skills through crowdsourcing. That does not mean task design can be ignored. Projects with tightly-focused, polished interfaces and entry-level microtasks are more likely to have a larger audience, which in turn means a critical mass of participants who could contribute to forums and encounter opportunities for learning through practice and discussion.

The role of curiosity

'Computers don't have curiosity. People often find things in the data that computers can't.' (*Zooniverse* team member Robert Simpson)¹³¹

When combined with access to data, curiosity can be a powerful motivator for undertaking complex or time-consuming tasks. The ability to mark (or collect) records in some way is an important enabler of curiosity and the learning it inspires. My awareness of the role of curiosity dates to my earlier research on crowdsourcing games for museums,¹³²

¹²⁹ Paas, Renkl, and Sweller, 'Cognitive Load Theory and Instructional Design'.

¹³⁰ Paas, Renkl, and Sweller, 'Cognitive Load Theory and Instructional Design'.

¹³¹ 'Nasa Seeks Coders to Hunt Asteroids', *BBC News*, 11 March 2014, <http://www.bbc.co.uk/news/technology-26528516>.

¹³² Ridge, 'Playing with Difficult Objects: Game Designs for Crowdsourcing Museum Metadata'.

and is further supported by two other researchers in crowdsourcing games for heritage data. Mary Flanagan and Peter Carini found that '[c]uriosity and doubt are key design opportunities', reporting that some players 'became so curious about the images they were tagging that they would tag images with inquiry phrases, such as "want to know more about this culture"'.¹³³ Projects that provide links to return to interesting items enable participants to pursue that curiosity.

One way of allowing participants to mark items of interest is with free-text tags. The *Zooniverse* Talk system uses hashtags (a string of letters beginning with the hash character '#'), which are recognised as a special structure by the system and function as ad hoc indexes to records without the need to define a special data structure or pre-determine a vocabulary. For example, *Operation War Diary* staff actively encourage their participants to use hashtags: 'If you find something of interest to you and others – consider using a tag in the comments– e.g. #gas / #prisoners. This will enable others to find it on the discussion boards'.¹³⁴ The *OWD* Talk platform has a module that highlights 'popular hashtags', helping participants learn how others have tagged items, and making the value of tagging items more evident. Participants on *OWD* have used hashtags to mark records they think might be of interest to historians or that they want to discuss with the community or moderators, including potentially 'new and untested' technologies or 'particularly poignant' actions mentioned.¹³⁵ Design features like free-text tags help projects support unexpected uses of their sources. For example, they allow people with specific knowledge to add more detail about their areas of expertise. The *Snapshot Serengeti* project had not

¹³³ Mary Flanagan and Peter Carini, 'How Games Can Help Us Access and Understand Archival Images', *American Archivist* 75, no. 2 (2012): 514–537, <http://archivists.metapress.com/content/b424537w2797ogu4/>.

¹³⁴ Forum posters, '60,000 Pages Tagged in 2 Days!'

¹³⁵ Forum posters, 'Casulties Tag - Room For Improvment'. (sic)

aimed to identify birds, but free-text tags allowed participants who were knowledgeable about birds to add more specific labels systematically, and as a result the project has now 'catalogued all the birds as well, for free'.¹³⁶

Tagging and other methods for marking or saving records for participants' own purposes may also be seen as part of the contract between 'citizen' projects and participants.¹³⁷

Other *Zooniverse* projects have reported on potential discoveries enabled by the combination of free-text tagging and forum communities,¹³⁸ and *Zooniverse* staff have commented that they regard them as important and use them extensively.¹³⁹ Perhaps influenced by this discourse, posters on *OWD* forums have forthright opinions on the need for free-text tags: 'you could (and should) widen your horizon and enable tagging for needs that other researchers might have, now or in the future. (Yes this is a collaborative project, it's supposed to be give-give, not give-take)'.¹⁴⁰

The relationship between close, active attention and deeper engagement

'I've read several of these war diaries in the past but nothing quite brings it alive in

¹³⁶ Karen Eng, 'You Found a Planet!: Robert Simpson Crowdfunders Scientific Research', *TED Blog*, 18 April 2014, <http://blog.ted.com/2014/04/18/you-found-a-planet-robert-simpson-on-zooniverse/>.

¹³⁷ However, unless carefully designed, adding optional free-text tagging instructions to indexing tasks may make other participants feel that the task is more complex, potentially reducing uptake.

¹³⁸ Amber York, 'Possible New Species? "Convict Worm"', *Seafloor Explorer*, 14 September 2012, <http://blog.seafloorexplorer.org/2012/09/14/possible-new-species-convict-worm/>.

¹³⁹ The Zooniverse and Commenters, '1 Million Classifications', *Seafloor Explorer*, (10 December 2012), <http://blog.seafloorexplorer.org/2012/12/10/1-million-classifications/>. The comment thread also reveals that at least one participant would like the work they have put into tagging images to count towards their profile.

¹⁴⁰ Forum posters, 'Why Can't I Tag Places / Activities / Other Units / Etc in Orders Pages? [Because We Are Interested in What Actually Happened]', *Talk Operation War Diary History Board / Help Board / FAQ - Tags and Tagging*, accessed 6 May 2014, <http://talk.operationwardiary.org/#/boards/BWDoooooof/discussions/DWDoooo2qd>.

the same way as tagging.' (OWD forum poster)¹⁴¹

Crowdsourcing projects provide a rare opportunity for members of the public to spend time actively and closely examining specific historic items. Historian Arlette Farge called the 'exact recopying of words' an 'exclusive and privileged way of entering into the world of the document';¹⁴² it seems there is something in the active, intimate encounters with historical materials that encourages people to engage with them beyond the immediate task of transcription. During interviews for previous research on crowdsourcing games in museums,¹⁴³ I noticed that the act of paying attention to 'boring' technical objects long enough to find tags to describe them made some participants interested in objects they would normally ignore. Possible reasons include a notion of leaving a 'trace' of one's self when tagging a historic item, a sense of connection to an artefact created by another person, or immersion in the 'small things and discrete particulars'¹⁴⁴ of the material. It could be that close reading creates the ideal conditions for Barthes' punctum, the detail in an image or objects 'which attracts or distresses' personal attention.¹⁴⁵ Another possibility is that participants' experiences with different examples of historical materials over time give them opportunities for 'mindfulness', an 'open, creative, probabilistic state of mind' which results from 'drawing novel distinctions, examining information from new perspectives, and being sensitive to context'.¹⁴⁶ Participation in crowdsourcing tasks and exposure to a range of source materials may also provide opportunities to iteratively build

¹⁴¹ Forum posters, '60,000 Pages Tagged in 2 Days!'

¹⁴² Arlette Farge, *The Allure of the Archives*, trans. Thomas Scott-Railton (New Haven, Connecticut: Yale University Press, 2013). p. 16.

¹⁴³ Ridge, 'Playing with Difficult Objects: Game Designs for Crowdsourcing Museum Metadata'.

¹⁴⁴ John Brewer, 'Microhistory and the Histories of Everyday Life', *Cultural and Social History* 7, no. 1 (2010): 87–109, doi:10.2752/147800410X477359.

¹⁴⁵ Roland Barthes, *Camera Lucida: Reflections on Photography* (London: Vintage, 2000). p. 40.

¹⁴⁶ Langer, 1989, cited in Csikszentmihalyi and Hermanson, 'Intrinsic Motivation in Museums'.

'cognitive schemas' that 'allow for multiple use of the same general knowledge for performing different tasks' and thereby enable more expert work.¹⁴⁷

I have turned to research on museum visitors to understand more about attention and engagement, drawing on Bitgood's definition of engagement as 'deep sensory-perceptual, mental and/or affective involvement'.¹⁴⁸ It generally requires 'some type of exertion or concentration' plus 'more than a few seconds' attention; the final outcome includes 'meaning making' and personal interpretation of content.¹⁴⁹ Bitgood's attention-value model based on studies of museum visitors posits that 'attention is a three-level continuum' of capturing attention, focusing it then engaging it, 'a progression from broad, unfocused attention' to 'highly focused examination' of the object and 'narrow, deep processing' of information about it.¹⁵⁰ His theory is that museum visitors are motivated to pay attention to exhibition objects when they have 'perceived value (a ratio of utility/satisfaction divided by costs such as time and effort)'.¹⁵¹ Participatory projects that present people with random objects may bypass this 'perceived value' calculation and encourage people to pay attention to items they may otherwise overlook. As one crowdsourcing participant put it, 'that's the thing. Originally it's a puzzle, work out the writing; then you get interested in what it says'.¹⁵² Engagement may lead to 'inquiry, critical thinking, and/or scientific reasoning', which may explain why participants are primed to become more involved with disciplines underlying projects that engage them with objects.

¹⁴⁷ Van Merriënboer, Kirschner, and Kester, 'Taking the Load off a Learner's Mind'.

¹⁴⁸ Bitgood, 'An Attention-Value Model of Museum Visitors'.

¹⁴⁹ Bitgood, 'An Attention-Value Model of Museum Visitors'.

¹⁵⁰ Bitgood, 'An Attention-Value Model of Museum Visitors'.

¹⁵¹ Bitgood, 'An Attention-Value Model of Museum Visitors'.

¹⁵² Various, 'AHRC Crowd Sourcing Study: Scoping Seminar'.

Providing stepping stones to more complex tasks and responsibilities

Research suggests that providing meaningful and varied roles within crowdsourcing projects will lead to continued participation in more complex tasks, which in turn leads to the development of further skills. It is also an excellent way to keep participants motivated and excited about new challenges and responsibilities while learning and practicing disciplinary skills and knowledge.¹⁵³ Human-computer interaction researchers Zhang and Von Dran discuss intrinsic motivational factors for website use including 'more task variety, less task routine, and provision for capabilities to move task performance to higher levels',¹⁵⁴ supplying evidence for the importance of providing a range of tasks, particularly more complex tasks. They also quote research that found jobs that score high on 'skill variety, task identity, task significance, autonomy, and feedback' lead to high performance and satisfaction.¹⁵⁵ Some cultural heritage crowdsourcing projects allow participants to graduate from tightly scaffolded microtasks to higher cognitive level processes¹⁵⁶ such as explaining, relating and theorising, particularly through community forums. The genealogy site *FamilySearch Indexing* use the phrase 'stepping stones' when they describe how the site aims to get people to try a simple indexing task, knowing that as transcribers are exposed to other people's histories, they may get interested in finding out more about their own families, and subsequently take on more complex research tasks.¹⁵⁷ *FamilySearch* point out that transcribing historical documents 'provides some much-needed,

¹⁵³ Providing different roles within a project, such as transcriber, data validator or moderator also helps deal with challenges specific to crowdsourcing, including the need to validate or approve contributions despite limited resources for community outreach and content moderation.

¹⁵⁴ Zhang and von Dran, 'Satisfiers and Dissatisfiers'.

¹⁵⁵ Hackman & Oldham, 1975, cited in Zhang and von Dran, 'Satisfiers and Dissatisfiers'.

¹⁵⁶ John Biggs, 'What the Student Does: Teaching for Enhanced Learning', *Higher Education Research & Development* 18, no. 1 (April 1999): 57–75, doi:10.1080/0729436990180105.

¹⁵⁷ Jessie Davis, 'Stepping Stones of Genealogy', *FamilySearch Blog*, 20 November 2012, <https://familysearch.org/blog/en/stepping-stones-genealogy/>.

introductory, family history education¹⁵⁸ while providing practice in palaeography.¹⁵⁹

Citizen science projects tend to be hierarchical with clearly defined roles,¹⁶⁰ including higher-level activities like data analysis and research dissemination. One advantage of this is the visibility of roles and responsibilities that participants may be able to take on as their skills and interest grow.¹⁶¹ Recent research on learning through participation in citizen science found that more advanced participant roles could be provided in community management,¹⁶² data quality control and helping devise project improvements.¹⁶³ Appointing participants to specific roles can free up project resources while helping experienced participants stay engaged with a project. Those with experience as participants may also be better placed to provide advice to others.¹⁶⁴ Informal roles and responsibilities can also emerge over time, and need not be formally labelled. As one *Old Weather* participant said, people 'assume roles based on what they can contribute'.¹⁶⁵

Participants are usually promoted based on the extent and quality of their contributions to a project. *FreeBMD* participants can become 'Syndicate Co-ordinators' after six months in

¹⁵⁸ Including knowledge about the range of record types and typical genealogical information.

¹⁵⁹ Davis, 'Stepping Stones of Genealogy'.

¹⁶⁰ Reed, Rodriguez, and Rickhoff, 'A Framework for Defining and Describing Key Design Features of Virtual Citizen Science Projects'.

¹⁶¹ Reed, Rodriguez, and Rickhoff, 'A Framework for Defining and Describing Key Design Features of Virtual Citizen Science Projects'.

¹⁶² For example, in participant forums and behind-the-scenes.

¹⁶³ Kloetzer et al., 'Learning by Volunteer Computing, Thinking and Gaming'.

¹⁶⁴ An example of this is seen in Forum poster, 'Civil War Pathways - 3-Way Brain Drain', *Civil War Pathways*, 24 April 2013, <http://civil-war-pathways.12348.n7.nabble.com/3-Way-Brain-Drain-td60.html>.

¹⁶⁵ Various, 'AHRC Crowd Sourcing Study: Scoping Seminar'.

an existing Syndicate and 'a reasonable number of files' uploaded to the site.¹⁶⁶ *Zooniverse* scientists get to know the more 'dedicated users' through their forums and can ask them for additional help with tasks such as tagging images with extra information.¹⁶⁷ *Zooniverse* projects also allow volunteers to self-nominate as moderators, and project staff will choose moderators at the start of each project.¹⁶⁸ *FamilySearch Indexing* recruits indexers to act as 'arbitrators' who review submitted transcriptions that need manual resolution. Their proxy for experience is having 'indexed at least 100 batches or 2,000 names with high accuracy'.¹⁶⁹ In their words, this suggests that 'the indexer is committed and serious about indexing', has enough experience reading handwritten records and has been able to transcribe with few errors.¹⁷⁰ Some projects have less quantitative criteria and may rely more on personal interactions than on formal measurements of accuracy or commitment. *Founders and Survivors* staff identified more reliable volunteers for 'tasks where we need people we can trust' from their interactions with participants.¹⁷¹ The manager of *Herbaria@Home* added 'experienced users' to the pool of those able to 'modify the collector and locality term-lists' and review others' tasks.¹⁷² He also explained that the process had not been 'systematic' and it was not intended to set up an 'exclusive group of privileged users', suggesting some

¹⁶⁶ 'FreeBMD Help: Welcome to the FreeBMD Signup Process', *FreeBMD*, accessed 23 March 2014, <http://freebmd.org.uk/Signup.html>.

¹⁶⁷ Eng, 'You Found a Planet!: Robert Simpson Crowdsources Scientific Research'.

¹⁶⁸ Eng, 'You Found a Planet!: Robert Simpson Crowdsources Scientific Research'.

¹⁶⁹ This blog post is also an interesting example of the use of emotive narratives and 'recovery' or 'remembrance' triggers: individuals 'waiting to be found and remembered'. Anderson, 'The Key to Unlocking Their Stories'.

¹⁷⁰ Jennifer Anderson, 'Arbitration Results Ruining Your Day?', *FamilySearch Blog*, 23 April 2012, <https://familysearch.org/blog/en/arbitration-results-ruining-day/>.

¹⁷¹ Claudine Chionh, *Founders and Survivors* staff member interview, interview by Mia Ridge, 13 April 2012.

¹⁷² Tom Humphrey, 'Reviewing Sheets, Editing Collectors and Localities Etc.', *Herbariaunited.Org Forum Index -> Bugs and Suggestions*, 10 January 2010, <http://herbariaunited.org/core/forum/viewtopic.php?t=2916>.

of the risks in elevating some users above others without clear criteria for doing so.¹⁷³

Sometimes credentials gained outside the project are considered important. Discussion on the *Civil War Pathways* forums suggest that a level of investment (such as actively participating in discussion, or sharing tips) or mentioning GLAM¹⁷⁴ credentials could prompt an invitation to take on validation tasks: 'we'd be happy to have a former cataloger!'.¹⁷⁵ Another approach to 'credentialing' is used on the *AskHistorians* forum (technically a 'sub-reddit') on Reddit.¹⁷⁶ Posters with 'detailed knowledge of their historical speciality and a proven record of excellent contributions'¹⁷⁷ are accorded 'flairs' (coloured labels listing their area of expertise) by moderators after nomination or application and listed on the 'flaired users' page.¹⁷⁸ The process is documented publically for scrutiny by other posters and moderators.¹⁷⁹

¹⁷³ He also made the future process more transparent by inviting self-nominations. Humphrey, 'Reviewing Sheets, Editing Collectors and Localities Etc.'

¹⁷⁴ Gallery, library, archive and museum

¹⁷⁵ Forum posters, 'Civil War Pathways - Re: A Plug for the Metadata Thesaurus!', accessed 27 March 2014, <http://civil-war-pathways.12348.n7.nabble.com/Re-A-plug-for-the-Metadata-Thesaurus-td84.html>.

¹⁷⁶ While this might seem an odd example given Reddit's reputation, *AskHistorians* could be considered one of the largest popular history sites, with over 406,000 readers in June 2015 (including 2,100 users concurrently browsing the site when I checked it on a random Sunday afternoon). <http://www.reddit.com/r/AskHistorians>. Last accessed 7 June 2015.

¹⁷⁷ Further information about the 'requirements for a flair' is available at Various, 'The Panel of Historians VIII: AskHistorians', *Reddit AskHistorians*, March 2014, http://www.reddit.com/r/AskHistorians/comments/1zifm6/the_panel_of_historians_viii/.

¹⁷⁸ The page only includes those who have posted in the last six months - while this may suggest that reputation relates to recency it may also be a way of encouraging people to post often enough to retain their listing on this page, or simply a result of the technical architecture of Reddit itself. Various, 'Flairedusers', *Reddit AskHistorians*, accessed 7 April 2014, <http://www.reddit.com/r/AskHistorians/wiki/flairedusers>.

¹⁷⁹ e.g. CrossyNZ, 'CrossyNZ Comments on The Panel of Historians Thread VII. Apply for Flair Here!', *Reddit AskHistorians*, March 2014, http://www.reddit.com/r/AskHistorians/comments/1yhjnh/the_panel_of_historians_thread_vii_ap

Some projects do not share responsibilities with volunteers. This may be because they are unwilling or unable to devolve authority or expertise, do not have an active community of trusted participants, or need to focus on their core task. Much like traditional volunteering, 'there is a trade-off between efficiency on the one hand and democracy and sustainability on the other hand', particularly for collaborative or co-creative projects.¹⁸⁰ Sharing responsibility for a project with participants may reduce the institutional workload for community engagement or data validation, but it may also require a greater commitment to community discussion than some projects can support.

The role of participant communities in scaffolding learning

The dynamic and personalised scaffolding that teachers provide in formal education as learners develop historical skills is not present in task-based crowdsourcing interfaces.¹⁸¹ However, crowdsourcing projects that encourage community participation can provide scaffolding through 'continuous and constructive interactions' between experts and novices.¹⁸² My observations of participant communities in crowdsourcing projects - including *Ancestry*, *Children of the Lodz Ghetto*, *Civil War Pathways*, *FamilySearch*, *Find A Grave*, *Founders and Survivors*, *FreeBMD*, *Galaxy Zoo*, the *Great War Forum*, *Herbaria@Home*, *iSpot*, *Lives of the First World War*, *Micropasts*, *Old Weather*, *Online*

ply_for/cfkwurh.

¹⁸⁰ That is, discussing and making decisions with a large group takes time. Wilderman, 'Models of Community Science'.

¹⁸¹ While it is impossible to adequately anticipate and programme scaffolded interactions in advance, machine learning and other technologies may make personalised scaffolding interfaces for learners possible in future. For now, the scaffolding required as participants begin to self-initiate questions and projects is more difficult to predict. Effective scaffolded feedback requires both a sense of what is required to complete the task, and of how the participant is currently performing. Wood, Bruner, and Ross, 'The Role of Tutoring in Problem Solving'.

¹⁸² Sharma and Hannafin, 'Scaffolding in Technology-Enhanced Learning Environments'.

Parish Clerks, Operation War Diary, Planet Hunters, Quench, Reddit AskHistorians, the Smithsonian Transcription Center, WikiTree and various local history societies and family history forums - showed the importance of participant communities in shaping the learning experiences of participants.

For example, participant communities on *Old Weather*¹⁸³ and *Herbaria@Home*¹⁸⁴ offer feedback on initial tasks for beginners. These conversations provide a welcome to the community in addition to the personalised feedback and cognitive apprenticeship that helps newcomers improve their skills. As these discussions are public, any reader can learn from the disciplinary skills being modelled by more experienced posters. A post requesting help transcribing a record might result in discussions that explain historical phrases, contextualise the record, or suggest other relevant primary or secondary sources. This can be more effective than leaving individuals to try and locate relevant resources on their own.¹⁸⁵ Through conversations on forums, participants with more expertise are able to respond to the changing skills, knowledge and motivations of newcomers, providing the personalised scaffolding required for learning or mastering skills and knowledge. The reflection required to document processes and respond to questions is in itself a valuable form of learning. *Wikipedia's* reliance on secondary rather than primary sources has made it less relevant to this thesis, but, in providing a space in which people actively engage in the process of writing history, it demonstrates the value of learning through discussion. Historian Roy Rosenzweig said that engaging in historiographical debate on *Wikipedia*

¹⁸³ Forum posters, 'Ask for Expert Advice'.

¹⁸⁴ Humphrey, 'About Feedback Requests'.

¹⁸⁵ From his decades' experience teaching adult local history classes, Riden says beginners in local history 'want to be told how they can make worthwhile discoveries quickly from what is available to hand, not given a long list of widely scattered archival sources or a vast bibliography'. Philip Riden, *Local History: A Handbook for Beginners* (London, U.K.: Batsford Academic and Educational Ltd, 1983).

'Talk' pages fosters an appreciation of 'the very skills that historians try to teach'.¹⁸⁶ The *iSpot* project¹⁸⁷ explicitly ties forums to their goal of encouraging 'as much learning as possible': 'Reading other people's queries, replies and discussion can be a great way of learning things'.¹⁸⁸ Participants are rewarded with points for social activities like 'making observations and posting comments in the forums' as well as for identifying species.¹⁸⁹ The project has developed a sophisticated 'Reputation' system designed to 'recognise and develop people's expertise in identification' of flora and fauna¹⁹⁰ and thereby encourage expert participation.¹⁹¹ This supports the learning goals of the project by 'providing a form of external feedback, recognizing and rewarding the activities' that the project wanted to encourage.¹⁹² Their model embodies the principles of situated, social learning.

Galaxy Zoo's 'green peas' provide an example of the ways in which group conversations contribute to situated learning by accumulating the 'individual knowledge of their members'¹⁹³ and generating insights and solutions that would not have arisen without

¹⁸⁶ Roy Rosenzweig, 'Can History Be Open Source? Wikipedia and the Future of the Past', *The Journal of American History* 93, no. 1 (June 2006): 117–46, <https://chnm.gmu.edu/essays-on-history-new-media/essays/?essayid=42>. Talk pages are used to discuss potential changes to *Wikipedia* articles.

¹⁸⁷ *iSpot* is a citizen science and 'participatory learning' project that encourages people to record and identify observations of flora and fauna, learning more about wildlife in the process. Clow and Makriyannis, 'ISpot Analysed: Participatory Learning and Reputation'.

¹⁸⁸ 'FAQ: Why No Personal Messaging or Closed Groups?', *ISpot*, accessed 13 June 2014, <http://www.ispotnature.org/faq-personal>.

¹⁸⁹ 'Reputation on ISpot', *ISpot*, accessed 13 June 2014, <http://www.ispotnature.org/help-reputation>.

¹⁹⁰ 'Reputation on ISpot'.

¹⁹¹ Experts' 'badges' also link back to their organisation's webpage, giving experts additional incentive 'to contribute constructively to the community'. Clow and Makriyannis, 'ISpot Analysed: Participatory Learning and Reputation'.

¹⁹² Clow and Makriyannis, 'ISpot Analysed: Participatory Learning and Reputation'.

¹⁹³ Schoenfeld, in preparation, cited in Brown, Collins, and Duguid, 'Situated Cognition and the

communal discussion. While the original Galaxy Zoo forum was an afterthought 'created to deal with the fact that we couldn't possibly deal with the volume of mail'¹⁹⁴ the project received when it launched, it serendipitously provided an ideal platform for situated learning. Reception of the original 'green peas' post was aided by the fact that the *Galaxy Zoo* forum was well-populated, and busy chat threads in sections with titles like 'Cafe at the end of the Universes' had helped get participants accustomed to posting. Importantly, the community also encouraged posts about unusual, beautiful or amusing items (or as they called them, 'Weird and wonderful' objects).¹⁹⁵ While the forum provided a platform to post those initial moments of curiosity, the community's collective curiosity and ability to collate similar examples provided the momentum to turn curiosity into discoveries. To paraphrase Louis Pasteur, chance favours the prepared community.¹⁹⁶

The role of access to expertise and project data

The discourse in various *Zooniverse* forums shows that the ability to share an image or other data about source materials is vital for encouraging discussion with other participants. Access to datasets created by the project is also important as it enables people who developed questions during a project to explore them with real data. However, as the following examples illustrate, access to data alone is not enough to turn non-experts into confident interpreters of that data.

Culture of Learning'.

¹⁹⁴ Chris Lintott, 'Thanks to the Forum – and Farewell', *Galaxy Zoo*, 9 July 2014, <http://blog.galaxyzoo.org/2014/07/09/thanks-to-the-forum-and-farewell/>.

¹⁹⁵ The title of a thread on the GZ forum. Another astronomical object discovered in *Galaxy Zoo* via persistent curiosity supported by forum discussion is 'Hanny's voorwerp', noticed by a 25 year old school teacher with no prior experience in astronomy. Paul Rincon, 'Teacher Finds New Cosmic Object', *BBC NEWS*, sec. Science/Nature, accessed 25 June 2014, <http://news.bbc.co.uk/1/hi/sci/tech/7543776.stm>.

¹⁹⁶ Puns also help.

In 2010 the *Zooniverse* team began building tools for data analysis,¹⁹⁷ and in 2013 they launched *Quench* as an opportunity for volunteers to take part in 'everything from classifying galaxies to analyzing results to collaborating with astronomers to writing a scientific article!'.¹⁹⁸ After a classification stage, volunteers were to 'use the tools available within *Zooniverse* to plot the data and look for trends'.¹⁹⁹ They were to 'gain background information' by reading summarised versions of 'seminal articles' about relevant scientific research²⁰⁰ and through 'interaction in Talk'.²⁰¹ Finally, after volunteers discussed their interpretations of the results with the science team, both groups were to 'collaboratively write a 4-page *Astrophysical Journal* article'.²⁰² About 1,200 participants participated in the classification stage, and 'around 250 participants then started to analyze them' with *Zooniverse Tools*.²⁰³

However, by November 2013, only 4 or 5 non-scientist participants were still posting on the forum. They posted to ask the scientists for the data created to date so they could continue their research independently,²⁰⁴ noting that two other *Zooniverse* projects that had made

¹⁹⁷ Mentioned as part of a discussion of 'citizen science' vs. 'real science' in Lintott, "Citizen" Science and "Real" Science'.

¹⁹⁸ Trouille, 'Galaxy Zoo Quench – Experience the Full Scientific Process'.

¹⁹⁹ Trouille, 'Galaxy Zoo Quench – Experience the Full Scientific Process'.

²⁰⁰ PDFs of the articles were to be provided 'Post-Quench Galaxies: Literature Review', accessed 23 June 2014, <http://postquench.blogspot.co.uk/>.

²⁰¹ Trouille, 'Galaxy Zoo Quench – Experience the Full Scientific Process'.

²⁰² Trouille, 'Galaxy Zoo Quench – Experience the Full Scientific Process'.

²⁰³ Ciara Byrne, 'Armchair Astronomers Help Discover Why Galaxies Stop Producing Stars', *Fast Company*, 3 January 2014, <http://www.fastcolabs.com/3024378/open-company/armchair-astronomers-help-discover-why-galaxies-stop-producing-stars>.

²⁰⁴ Forum posters, 'Quench Project: A Proposal Aimed at Reviving and Completing It', *Talk Galaxy Zoo Quench*, 2014, <http://quenchtalk.galaxyzoo.org/#/boards/BGSooooooe/discussions/DGSoooo22f>.

significant discoveries with participants, *SpaceWarps* and *Planet Hunters*,²⁰⁵ had enabled data analysis by providing access to the project data.²⁰⁶ They also pointed out that in those projects, unlike their experience of *Quench* at that time, 'professional astronomers are very actively engaged'.²⁰⁷ This was particularly important as they found it difficult to continue learning the science and analysing the data without that expertise: 'feedback and guidance are key here'.²⁰⁸ Meanwhile, the original *Galaxy Zoo* forum was commenting on the end of an 'era of dynamical exchange between the membership and the researchers',²⁰⁹ and linking it to a lack of 'serendipitous discoveries'. One poster pointed out that participants have 'continued to make serendipitous discoveries, and by now have possibly noted dozens, even hundreds' but that 'it takes a professional astronomer to read those posts, containing such discoveries, to recognize that there's something odd/really new/cool/etc, and then to take the time to look into them a bit more'.²¹⁰

²⁰⁵ Megan E. Schwamb, 'Planet Hunters's First Circumbinary Planet- A True Team Effort', *Planet Hunters*, 15 October 2012, <http://blog.planethunters.org/2012/10/15/planet-hunterss-first-circumbinary-planet-a-true-team-effort/>. Chris Lintott, 'PH1: A Planet in a Four-Star System', *Planet Hunters*, 15 October 2012, <http://blog.planethunters.org/2012/10/15/ph1-a-planet-in-a-four-star-system/>.

²⁰⁶ Forum posters, 'Reviving and Completing the Quench Project: A Proposal', *Talk Galaxy Zoo Quench*, January 2014, http://quenchtalk.galaxyzoo.org/#/boards/BGSooooooe/discussions/DGSoooo22g?page=1&comment_id=52e6e579f1e0404a9000876.

²⁰⁷ Forum posters, 'Reviving and Completing the Quench Project: A Proposal'.

²⁰⁸ As one poster said, 'I could happily waste huge numbers of hours on scientifically marginal (or worse) questions. The upshot of this is that very few of analyses ever got very far'. Forum posters, 'Quench Project: A Proposal Aimed at Reviving and Completing It'.

Forum posters, 'Where Have All The ...?', *Talk Galaxy Zoo Quench*, 2013, <http://quenchtalk.galaxyzoo.org/#/boards/BGSoooooo2/discussions/DGSoooo1yp>.

²⁰⁹ 'Has Galaxy Zoo Research Become Manufacturing? (Why Astronomers Are so Rare Here?)', accessed 24 June 2014, <http://www.galaxyzooforum.org/index.php?topic=281847.0>.

²¹⁰ Forum posters, 'The Unbearable Arbitrariness of Serendipitous Galaxy Zoo Discoveries?', *Galaxy Zoo Forum*, 2014, <http://www.galaxyzooforum.org/index.php?topic=281590.0>.

These forum posts echo statements from staff involved with *Solar Stormwatch* and the first stages of *Old Weather* at the National Maritime Museum: 'professional researchers' must be involved in the community discussion not only to 'set specific challenges and provide feedback', but also 'to respond to the questions and interests that emerge from the community'.²¹¹ Earlier research by environmental scientist Candie C. Wilderman found that getting participants to design the study and interpret the data involves 'intensive mentoring by the service provider and a high level of commitment by the volunteers', and that this investment of time is required for moving participants 'from a sort of worker mentality into a scientist mentality'.²¹² These *Quench* and *Galaxy Zoo* posts also highlight the role of expertise in spotting interesting oddities among the un-interesting oddities. One scientist's post encouraged participants to investigate 'anything odd or weird';²¹³ but, as an expert in the field, the scientist may not be able to appreciate the role their own tacit knowledge plays in knowing which oddities are more likely to lead to discoveries.²¹⁴ Through their participation in forum conversations, experts create an environment that supports cognitive apprenticeship and the passing on of tacit disciplinary knowledge. Expectations are also created through participants' experience with previous projects.²¹⁵ This may have led to participants in *Operation War Diary* expecting a greater level of expert input in community discussion than the project was able to provide. In a blog post explaining why 'we have not been as active as we would like', *OWD* staff hope that by

²¹¹ Romeo and Blaser, 'Bringing Citizen Scientists and Historians Together'.

²¹² Wilderman, 'Models of Community Science'.

²¹³ Forum posters, 'Quench Project: A Proposal Aimed at Reviving and Completing It'.

²¹⁴ This does not discount the possibility that the scientist lacked time to trawl through forum posts.

²¹⁵ For example, one *OWD* forum poster mentioned 'the expected tapestry of a zooniverse project' in Forum posters, 'Casulties Tag - Room For Improvement'.

'reading the diaries, volunteers are gaining expertise themselves', and implore participants to 'please take the initiative and conduct conversations with other Citizen Historians'.²¹⁶ However, novices in conversation with each other have fewer opportunities to engage in situated learning and cognitive apprenticeship, compared to situations where they are able to observe and converse with experts at work.²¹⁷ The scaffolding required for learning disciplinary skills requires some level of initial expert support before it is gradually removed. In a departure from the apparent *Zooniverse* model of regular participation from scientists at the start of each project, the same *OWD* post says that the project is keen for 'professional historians from academia and elsewhere' to join discussions but that they 'do not wish to create the expectation that we will ever guarantee any level of response from us to your history enquiries'.²¹⁸ This post may be an attempt to manage expectations about their ability to respond to queries about specific family histories²¹⁹ rather a statement of their attitude to historiographic questions. However, the inability to support expert input appears to have disappointed many early participants whose expectations were excited by publicity material around 'citizen historians'.

Can crowdsourcing projects help participants become historians?

It seems clear that crowdsourcing projects can encourage participants to develop disciplinary skills and knowledge. But can they help participants become historians? And to what extent does participation in the core tasks of historical crowdsourcing (such as transcription or description of historical material) and situated learning encourage the

²¹⁶ Luke Smith, 'Inside Team War Diary...', *Operation War Diary*, 8 April 2014, <http://blog.operationwardiary.org/2014/04/08/inside-team-war-diary/>.

²¹⁷ This is not intended to downplay the value of the posts made and questions raised by novices, nor the levels of expertise present in posters not officially listed as project historians.

²¹⁸ Smith, 'Inside Team War Diary...'.

²¹⁹ An issue reported by other projects with records relevant to family historians. Chionh, Founders and Survivors staff member interview.

development of those skills? Briefly revisiting the literature on historical thinking and core competencies for historians provides some answers.

The American Historical Association's core competencies for students in history courses and degree programmes include practising historical empathy, understanding the complex nature of the historical record, and generating significant, open-ended questions about the past,²²⁰ all of which are seen in participants' discussions. Tasks like transcription provide opportunities for becoming familiar with a variety of sources while practicing the 'technical skills' of palaeography.²²¹ Participants undertaking meaningful crowdsourcing tasks as part of an online community of practice may also be 'scaffolded' through the process of devising research strategies and practicing historical thinking. Project forums contain examples of source-based and interpretive skills,²²² as participants search for and debate the value of sources related to their discussion. Some community discussions help participants notice and understand the 'mediations' at work on historical sources,²²³ and others show evidence of Tally and Goldenberg's historical thinking in action as participants post questions and discuss sourcing, inferences and evidence.²²⁴ Forum discussions also provide evidence for people corroborating information between sources (for example, looking at other sources to check personal and place names). Previous research suggests that working with primary sources assists historical thinking as it encourages active engagement 'in the construction and interpretation of history'.²²⁵

²²⁰ American Historical Association, 'AHA History Tuning Project: History Discipline Core'.

²²¹ Jordanova, *History in Practice*, 2006. p. 151.

²²² Jordanova, *History in Practice*, 2006. p. 151.

²²³ Jordanova, *History in Practice*, 2006. p. 160.

²²⁴ Tally and Goldenberg, 'Fostering Historical Thinking With Digitized Primary Sources'.

²²⁵ Thea Lindquist and Holley Long, 'How Can Educational Technology Facilitate Student Engagement with Online Primary Sources?: A User Needs Assessment', *Library Hi Tech* 29, no. 2 (2011): 224–41, doi:10.1108/07378831111138152.

To return to Andrews and Burke's 'five C's',²²⁶ the intimate exposure to cumulative tiny details in historical materials can provide participants with a sense of contingency and change over time (whether changing technologies, social mores, legal structures, life circumstances or other information mentioned in source documents). While individual documents are potentially rich in detail, the provision of further context depends on the priorities of the project and/or the community around it. However, for crowdsourcing participants, the application of these more technical skills typically precedes the development of interpretative skills. With enough time working with source texts and images, and the opportunity to follow-up questions that arise, it is possible that crowdsourcing does encourage participants to understand causality and develop 'persuasive explanations of historical events and processes based on logical interpretations of evidence'.²²⁷

Crowdsourcing projects and claims about citizen history

While crowdsourcing can develop some historical competencies, it is less clear that crowdsourcing projects provide opportunities to craft historical narratives²²⁸ or plausible arguments.²²⁹ If making explicit historiographic arguments in narrative form is considered a core historical skill, then some claims about citizen history projects are called into question. It also seems that the development of historical narratives is more likely to take

²²⁶ That is, understanding 'change over time, causality, context, complexity, and contingency'. Andrews and Burke, 'What Does It Mean to Think Historically?'

²²⁷ Andrews and Burke, 'What Does It Mean to Think Historically?'

²²⁸ American Historical Association, 'AHA History Tuning Project: History Discipline Core'.

²²⁹ Jordanova, *History in Practice*, 2006. p. 151. However, as discussed in relation to family and local historians in the next chapter, it is not clear that developing arguments is always a goal for non-faculty historians.

place in side projects,²³⁰ like those focusing on the compilation of biographies or ships' histories, and that the quality of the historical arguments produced is variable. For example, the *Old Weather* ships' histories tend to be compilations and visualisations of data,²³¹ while the meiosis.org.uk biographies produced by *Herbaria@Home* participants are closer to traditional narrative biographies.²³² These side projects, created by participants to support their interest in historical research projects, effectively add a collaborative aspect to contributory projects.²³³

While *Operation War Diary* claimed to be 'a new way of doing history - citizen history',²³⁴ my analysis of their forums has revealed a tension between the desire to label a project as 'citizen history' (implying a goal of going beyond crowdsourcing microtasks to encourage curiosity and deeper engagement, and/or develop the skills of participants) and the desire for quickly and efficiently achieving data processing goals. The absence of experts throws into relief an inherent tension in the project. *OWD*'s stated aims are to create material for use in the *Lives of the First World War* project, for academics and for the National

²³⁰ As these new questions and narratives are more likely to relate to the source materials than to the goals of the official project, the status of these as related-but-separate projects may well be appropriate.

²³¹ However, as Edward H. Carr, said, facts from the past are not history. Edward Hallett Carr, *What Is History?* (New York: Knopf, 1962). quoted in Kelly, 'Thinking: How Students Learn About the Past'.

²³² That is, the material is descriptive, organised chronologically and may contain 'sub-plots', but is focused to tell a 'single coherent story'. Lawrence Stone, 'The Revival of Narrative: Reflections on a New Old History', *Past & Present*, no. 85 (1 November 1979): 3–24, <http://www.jstor.org/stable/650677>.

²³³ Here I am referencing Bonney et al.'s contributory projects 'designed by scientists' and collaborative projects where the public can help refine the project design and analyse data. In some ways these side projects are similar to the grassroots or community history projects organised by local histories societies and other specialist groups.

²³⁴ Forum posters, '60,000 Pages Tagged in 2 Days!'

Archives.²³⁵ The project uses the term 'citizen historians' in promotional material, but - if supporting the development of participants as historians is not an explicit aim, then the project might not have allocated sufficient resources for supporting the emergence of a community of practice. Other projects may discover that they cannot find the resources to match their ambitions, or they may fail to produce enough community engagement from their own experts.²³⁶ Some resistance may also stem from the additional workload and requirement for professionals involved in projects to communicate, facilitate, explain complex reasoning, and express their tacit knowledge in prose. It is also important to remember that it took time for *Old Weather* participants to discover an interest in history related to the ships' logs, so projects such as *OWD* may still yet nurture future historians.

This analysis is important and timely because grand claims are being made about crowdsourcing and citizen history. One NEH²³⁷ funding document contains the statement: 'By developing a transcription tool for the Coptic language, the team will engage citizen scholars to help transcribe thousands of Egyptian papyrus documents that help tell the story of early Christian life';²³⁸ yet this analysis has made clear that providing access to tasks does not automatically support the development of citizen scholars. Claims like this seem to be either the result of a misunderstanding of the requirements for supporting 'citizen historians',²³⁹ or of a desire to follow a trend and make it easier to sell a project to

²³⁵ Operation War Diary, 'About Operation War Diary'.

²³⁶ It is also possible that some projects unintentionally undermine their own claims because of a sense of discomfort with the process of breaking down the distinction between 'non-professionals' and 'non-experts'.

²³⁷ The American National Endowment for the Humanities

²³⁸ 'NEH Grant Details: Resurrecting Early Christian Lives: Digging in Papyri in a Digital Age', *National Endowment for the Humanities*, 12 February 2014, <https://securegrants.neh.gov/publicquery/main.aspx?f=1&gn=HJ-50178-14>.

²³⁹ Itself a problematic term, discussed further in the next chapter. See also Kate Theimer, 'Why We

those who would be less excited by the idea of a volunteer digitisation (or 'traditional' crowdsourcing) project. But is there any real harm in this mislabelling?

To understand the implications of over-stating the participatory or collaborative qualities of a project, I have drawn on earlier research from the museum sector. The Paul Hamlyn Foundation commissioned Bernadette Lynch to investigate 'the real nature and effectiveness of the engagement practices of museums and galleries'. Lynch's subsequent report used the term 'empowerment-lite' to describe the 'disillusioning experience' of community members in participatory projects when the actual experience involved 'a level of control, risk aversion and "management" by the organisations'.²⁴⁰ Lynch noted that this experience undermined the impact and value of the project for participants, and may have prevented them from realising their capabilities.²⁴¹ Similarly, promising transcribers that they can become 'citizen historians' without supporting the process could reduce the ability of real citizen history projects to attract participants. The emerging field of citizen scholarship could suffer consequently. It also raises ethical issues if participants join projects in part to learn new skills but are not given opportunities to do so. The experiences of *Lodz*,²⁴² *Quench* and *OWD* show that providing personalised feedback for participants is time-consuming and can be difficult to resource adequately.²⁴³ However, as

Need to Find a Term to Replace "Citizen Archivist", *ArchivesNext*, 9 April 2010, <http://www.archivesnext.com/?p=1214>.

²⁴⁰ Bernadette Lynch, 'Whose Cake Is It Anyway? A Collaborative Investigation into Engagement and Participation in 12 Museums and Galleries in the UK' (Paul Hamlyn Foundation, 2011), <http://www.phf.org.uk/page.asp?id=1417>.

²⁴¹ Lynch, 'Whose Cake Is It Anyway? A Collaborative Investigation into Engagement and Participation in 12 Museums and Galleries in the UK'.

²⁴² Frankle et al., Children of the Lodz Ghetto project staff members interview.

²⁴³ Particularly as community participation and guidance work mostly seems to be added on top of existing job descriptions rather than resourced as a specialist role. Earlier *Zooniverse* project forums may have benefitted from the participation of on-staff early career scholars, who tend to have fewer

expert participation in community discussion seems to be important for 'citizen' projects, finding ways to provide it might be a necessary part of citizen history projects.²⁴⁴ Future research could usefully determine the level of expert participation needed to seed expertise within a community, and whether those communities could subsequently provide most of their own scaffolding as early participants pass on their own acquired knowledge and skills.

Invisible or unstated participant hierarchies may also cause issues. *OWD* calls its participants citizen historians, but applies the label 'historian' to posts by 'official' project historians on its forum. In this context, there is a risk that the label 'citizen historian' could be read as 'faux historian'.²⁴⁵ Other projects take a different approach and directly address the existence of externally recognised experts. *iSpot's* Reputation page explicitly states that they have 'two levels of "pre-set" reputation scores' for participants who have been labelled as Expert or Knowledgeable, based on 'evidence of their experience in wildlife identification and recording', whether through academic training or through membership of a 'recording scheme or natural history society'.²⁴⁶

Conclusion

This chapter demonstrates the value in theories of legitimate peripheral participation, cognitive apprenticeship and informal learning in understanding how participation in microtasks and communities of practice can provide a form of disciplinary training. The analysis presented in this chapter has shown that crowdsourcing projects can help

administrative calls on their time and may be more excited about discussing their research topic.

²⁴⁴ Perhaps in addition to the plentiful research into motivations for public participation in citizen scholarship, we should be looking to understand and increase expert motivations for participation in community discussion.

²⁴⁵ This is discussed further in the final chapter.

²⁴⁶ 'Reputation on iSpot'.

participants gain new skills by providing them with opportunities to learn technical skills. The close and active engagement with historical sources that heritage crowdsourcing requires can create opportunities for curiosity about the source material and the lives they represent. In some cases, this spark of curiosity is nurtured through participant discussion, which may also provide communities of practice and situated learning that help participants develop historical thinking and some historical skills.

Overall, my review of crowdsourcing and citizen science/history projects has revealed that there are three types of crowdsourcing/citizen history projects.²⁴⁷ The first is the large group of crowdsourcing projects that have the potential to support emergent communities of citizen scientists or historians gathered around contributory tasks. Included in this category are projects that wish to encourage citizen history but do not make enough provision for expert participation and scaffolding. That is, any deeper engagement or learning is a beneficial outcome but not a core goal (and is thereby unlikely to be resourced). Many excellent crowdsourcing projects with hopes of supporting citizen history are included in this category. The second group encompasses projects that serendipitously provided enough of the attributes important for learning historical skills; these could perhaps be called 'accidental citizen history' projects. Accidental citizen history projects, such as *Old Weather* and *Herbaria@Home*, did not include history among their original goals. However, the combination of interesting material, an active forum that provided opportunities to discuss or ask questions about historical records, and in the case of *Old Weather*, encouragement from maritime historians, eventually lead to historical research projects. Finally, based on my analysis of the field, and in the interests of clarity, I would define 'citizen history' projects as those that require or teach some

²⁴⁷ Here I have excluded grassroots, or self-organised, projects created by community historians. Examples include *FreeBMD*, *Online Parish Clerks* and other local and history society projects.

historical skills beyond the technical, palaeographic skill of transcribing text. This chapter has presented two citizen history projects - *Marine Lives* and *Children of the Lodz Ghetto* - that can only succeed if participants are able to learn or bring some disciplinary skills to the more complex tasks that contribute to the projects' goals.

If, as Jordanova states, '[h]istory is indeed about what historians do',²⁴⁸ then it seems crowdsourcing can provide excellent opportunities for exposure to the things historians do. At a practical level, crowdsourcing provides opportunities to learn technical skills like palaeography, and gain the familiarity with sources that leads to source-based skills. Intimate encounters with a range of historical sources during transcription, classification and indexing tasks provide opportunities for reflecting on the concepts of change over time, complexity and contingency related to historical thinking.²⁴⁹

Crowdsourcing projects may provide a space in which citizen historians can develop in a number of ways. As for the recommendations in the conclusion of the previous chapter, the analysis presented in this chapter would need testing and validation to have the weight of formal recommendations. That said, historians and institutions looking to develop citizen history projects could usefully consider: giving people an opportunity for closely and actively engaging with items they might not otherwise have viewed; giving them meaningful tasks to do with those items; supporting any moments of curiosity by allowing people to mark, share and re-visit interesting items; providing access to a community similarly engaged with the meaningful goals of the wider project; and providing a visible and accessible expert presence.

²⁴⁸ Ludmilla Jordanova, *History in Practice* (London: Arnold, 2000).

²⁴⁹ Andrews and Burke, 'What Does It Mean to Think Historically?'

Chapter 4: Historians' working practices and digital tools, resources and methods

This chapter shifts our focus from public participation (through crowdsourcing and citizen history) to scholarly practices in historical research. How has the availability and use of digital platforms, tools and methodologies affected the everyday practices of faculty and community historians?¹ The ability to compare two broad groups of historians helps nuance our understanding of the impact of digital technologies on historical research. It also reveals the impact of academic reward structures on faculty historians' attitudes to the sharing and collaboration practices that have been enabled by technologies and models for participative digital history. This chapter presents the results of interviews conducted in 2012 with 29 historians, who discussed the digital tools, resources and methods they use in their research. Data gained through these interviews reveals how faculty, family and local historians evaluate, use and contribute to 'traditional'² and participative digital resources. I have focused on historians whose research includes early modern England in order to analyse a group who face similar issues with the availability and evaluation of primary

¹ As elsewhere in this thesis, 'faculty historians' refers to those employed in academic research or teaching posts, in contrast to 'academically-trained historians', who might work as public historians, in 'alt-ac' jobs within universities, or have an entirely different job unrelated to their academic training as historians. 'Community historians' refers to family, local and other historians who undertake historical research as a voluntary activity.

² Inevitably, print-based resources were discussed in these interviews but they are not the focus of this research. In this context, 'traditional' digital resources tend to be based on print-based models that pre-date participatory websites and social media. For further background, see Tim O'Reilly's discussion of 'Web 1.0' and 2.0 models. Tim O'Reilly, 'What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software', *O'Reilly*, 30 September 2005, <http://www.oreilly.com/pub/a/web2/archive/what-is-web-2.0.html>.

sources.³ The data presented here are synthesised and contextualised within a wider multi-disciplinary discourse on the changes in scholarly practices prompted by digital technologies. The chapter discusses the qualities that researchers consider important when evaluating participatory digital resources, including resources created by or with contributions from unknown members of the public. It provides an overview of the extent and circumstances in which historians share data and collaborate with others. It concludes that the most experienced faculty and community historians have similar approaches to discovering, evaluating and gathering resources. However, their different motivations and goals are apparent in their attitudes to sharing and the importance of credit and attribution. My analysis of these interviews contributes to the overall thesis by deepening our understanding of the impact of digitality on scholarly practices in historical research.

Context and research questions

I draw upon definitions of 'data' and 'information' from information science to define more precisely what is being exchanged in my discussion of sharing and collaboration.

Accordingly, data is defined as unorganised, unprocessed 'discrete, objective facts or observations' which do not yet convey any specific meaning.⁴ For historians, this would typically be the content of historical materials. Conversely, information is data that has been processed or organised through processes including 'classification, rearranging / sorting, aggregating, performing calculations, and selection'.⁵ To further complicate

³ While I intended to focus on early modern England, very few family historians were able to extend their research as far as the 18th century, and much of their research focused on the 19th and 20th centuries; this is reflected in the interview data. These limitations reveal gaps in the historical record and the ability to reliably identify earlier individuals in the sources available.

⁴ J. Rowley, 'The Wisdom Hierarchy: Representations of the DIKW Hierarchy', *Journal of Information Science* 33, no. 2 (15 February 2007): 163–80, doi:10.1177/0165551506070706.

⁵ Rowley points out that the boundary between information and knowledge is not clearly defined so I have used the term 'information' rather than focus on definitional issues. Rowley, 'The Wisdom

matters, information is often processed with a scholar's tacit knowledge. By definition, tacit knowledge is 'unspoken and hidden'⁶ - taken for granted, difficult to articulate, acquired by 'sustained involvement in the relevant cultural milieu'⁷ rather than taught.

Various disciplines have contributed literature on the research practices of historians. Librarians, archivists, information science and professional support organisations have sought to understand the behaviours and preferences of historians (and those in related humanities disciplines). The scope of, and scripts for, my interviews were informed by literature on 'scholarly primitives', the 'basic activities common to research across humanities disciplines',⁸ to ensure that the interviews covered the most relevant activities within various stages of the historian's research process. Delineating the stages of research was useful for thinking about how sharing and collaboration practices might change as the researcher moved between different stages. After an initial review,⁹ I determined that the

Hierarchy'.

⁶ McInerney, 'Knowledge Management and the Dynamic Nature of Knowledge'.

⁷ Becher further states that tacit knowledge in history takes the form of 'folklore and gossip' rather than 'implicit rules of conduct or explicit concerns with epistemological status'. Becher, 'Disciplinary Discourse'.

⁸ John Unsworth, 'Scholarly Primitives: What Methods Do Humanities Researchers Have in Common, and How Might Our Tools Reflect This?' (symposium on 'Humanities Computing: formal methods, experimental practice', King's College, London, 13 May 2000), <http://people.lis.illinois.edu/~unsworth/Kings.5-00/primitives.html>.

⁹ Other discussions of scholarly primitives and research activities or methods I reviewed include the discussion of the *DARIAH* project in Sheila Anderson, Tobias Blanke, and Stuart Dunn, 'Methodological Commons: Arts and Humanities e-Science Fundamentals', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 368, no. 1925 (18 July 2010): 3779–96, doi:10.1098/rsta.2010.0156. and Benardou et al., 'A Conceptual Model for Scholarly Research Activity', 'TaDiRAH - Taxonomy of Digital Research Activities in the Humanities (v. 0.5, 02/2014)', *TaDiRAH - Taxonomy of Digital Research Activities in the Humanities*, February 2014, <https://github.com/dhtaxonomy/TaDiRAH>. and the 'stages of inquiry' in Wiggins and Crowston, 'From Conservation to Crowdsourcing: A Typology of Citizen Science'. While my

most broadly useful categories were University of Minnesota Libraries' (UML) 'discover', 'gather', 'create' and 'share',¹⁰ devised to represent the activities that scholars 'regularly and consistently' engage in throughout the research process. Meho and Tibbo's work on the four stages of information-seeking behaviour by social scientists (including historians) - searching, accessing, processing and ending¹¹ - provided useful background material, as did Palmer et al.'s detailed review of scholarly information practices.¹² The breadth of UML's categories meant they were flexible enough to accommodate interviews with both faculty and non-faculty historians as well as non-traditional scholarly activities and forms of knowledge or resource creation. For instance, UML's 'discover' was a more flexible heading than Palmer et al.'s 'searching'¹³ because it allowed for the fact that not all information discovery is active.¹⁴ UML's 'create' was a useful catch-all for the various processes for turning data into knowledge through comparing, annotating, relating, referring, writing, illustrating, contextualising, analysing, enriching and interpreting materials, while allowing for the variations in outputs I expected to find between different types of

research was carried out in 2011-2014, forthcoming work such as the *NeDiMAH* project's Methods Ontology 'NeMO' (<http://nedimah.eu>, due mid-2015) could be useful for future research in this area.

¹⁰ Their research was based on over 50 interviews with faculty and graduate students designed to inform a process of designing, prototyping and evaluating a 'research support environment', an approach similar to my original proposal. The library also holds special collections including rare books and manuscripts, and their interviews discussed the challenges of archival research, making their results particularly applicable to my research. University of Minnesota Libraries, 'A Multi-Dimensional Framework for Academic Support'.

¹¹ Lokman I. Meho and Helen R. Tibbo, 'Modeling the Information-Seeking Behavior of Social Scientists: Ellis's Study Revisited', *Journal of the American Society for Information Science and Technology* 54, no. 6 (2003): 570-587, <http://onlinelibrary.wiley.com/doi/10.1002/asi.10244/full>.

¹² Palmer, Teffeau, and Pirmann, 'Scholarly Information Practices in the Online Environment'.

¹³ Palmer, Teffeau, and Pirmann, 'Scholarly Information Practices in the Online Environment'.

¹⁴ For example, some information discovery is the result of recommendations from archivists, librarians or other academics, serendipitous discovery in physical locations, emailed journal content alerts, etc, as well as catalogue or web searches.

historians. I was also interested in some cross-category activities that appeared in the interview data, particularly processes such as annotation that leave a textual trace or link. Under the broad UML headings, primitives related to gathering, creating and sharing were the most relevant for my research questions. My questions were also influenced by Meho and Tibbo's category of verifying, 'checking the accuracy of the information found', and their concept of 'differentiating', in which information sources are evaluated 'according to their nature, quality, relative importance, and usefulness'.¹⁵

Online local and family history forums and related literature provided valuable sources for understanding their research practices and anticipating some of the difficulties faced by historians of early modern England. This suggested that disambiguation would be a key task for historians. Identifying individual people across different databases or differentiating between people with the same name is described by archivists Wendy Duff and Catherine Johnson as an 'extremely time-consuming but an essential task',¹⁶ while historian Ralph W. Mathisen also points out that looking for the same person across different databases is also extremely difficult.¹⁷ During my interviews it also became apparent that while faculty historians are free to choose their area of specialism based on their interests and the sources available, local and family historians are limited to the sources available for the periods and places relevant to their location and/or ancestors.¹⁸

While a small number of university- or faculty-led projects, such as Tanya Evan's work on

¹⁵ Meho and Tibbo, 'Modeling the Information-Seeking Behavior of Social Scientists'.

¹⁶ Wendy M. Duff and Catherine A. Johnson, 'Where Is the List with All the Names? Information-Seeking Behavior of Genealogists', *American Archivist* 66, no. Spring/Summer (2003): 79–95.

¹⁷ Ralph W. Mathisen, 'Where Are All the PDBs?: The Creation of Prosopographical Databases for the Ancient and Medieval Worlds', *Chronicle of Higher Education*, 1989, A7, <http://prosopography.modhist.ox.ac.uk/images/04%20Mathisen%20pdf.pdf>.

¹⁸ It seems that most, but not all, local historians research the areas close to their location.

the Australian Benevolent Society¹⁹ or the Cambridge Group for the History of Population and Social Structure's population reconstitution²⁰ rely on the labour of family and local historians, other faculty historians do not value the contributions of 'amateur' historians. Statements by faculty historians, such as 'sorting out who's who [in a prosopographical database] is not a job for amateurs',²¹ seemingly ignore the fact that amateur family and local historians regularly undertake similar tasks. The range of responses to the work of 'amateur' historians I encountered informed my questions about how historians assess participative resources.

Yochai Benkler, who coined the term 'commons-based peer production', pointed out that online publishing has broken the link between the distribution of information and its credibility.²² Previously, the cost of print publication and distribution was a barrier to self-publishing that afforded a printed book some respectability. Physical items like books or journals contained further visible signifiers of authority and status such as the name and location of the publishing house (and by association, their reputation and standards), distribution channels, tactile aspects like the quality and format of the printed object, and conventions like author biographies. While the advent of desktop publishing and What-You-See-Is-What-You-Get (WYSIWYG) editing software gave anyone with access to comparatively low-priced consumer-level equipment 'the power to produce documents that appear "published"' as if on a printing press,²³ there were still visible differences between professionally produced and desktop published items. The web has lowered the

¹⁹ Evans and Curthoys, 'Family History, Identity, and Public History'.

²⁰ Wrigley, 'Small-Scale but Not Parochial'.

²¹ Mathisen, 'Where Are All the PDBs?'

²² Benkler, 'Coase's Penguin'.

²³ Joel P. Bowman and Debbie A. Renshaw, 'Desktop Publishing: Things Gutenberg Never Taught You', *Journal of Business Communication* 26, no. 1 (1989): 57-77, <http://job.sagepub.com/content/26/1/57.short>.

costs of production and distribution even further, and a skilled amateur can produce a website that looks as good, if not better, than those produced by academic projects.²⁴ Meanwhile, the pace of change means that websites start to look outdated within a year or two, regardless of who created them. Given these challenges, which combination of attributes makes a site seem authoritative and trustworthy, and which lead to suspicion? How have scholars adapted their methods for assessing resources for these challenges of digital history? People contributing data or information to participatory projects range from world-class experts in highly specialist topics to novice historians or ordinary members of the public with no background in history. How, then, do historians judge which online resources they should and should not use?

Digital, networked technologies have enabled many new tools and practices, only some of which have been adopted by historians and institutions holding historical materials. What determines whether a technology is adopted or ignored by historians, and how are those technologies transformed through the 'mangle of practice'?²⁵ Where are the gaps between the affordances and functions of software and the practices of the researchers who use them? How have the practices of sharing and collaborating on research been affected by digital tools for finding, using and publishing resources? Are researchers more likely to

²⁴ Naturally, amateur sites will not have the imprimatur of university presses, but this may concern some researchers less than others.

²⁵ Here I draw on Andrew Pickering's call to consider the 'open-ended, reciprocally structured interplay of human and nonhuman agency', where nonhuman actors include material instruments and machines (Pickering, p. 3), further described by Brian Marick as the 'prolonged interplay of resistance, accommodation, and chance' (p. 185). With thanks to Dr Brunton for the suggestion. Andrew Pickering, 'New Ontologies', in *The Mangle in Practice: Science, Society, and Becoming*, ed. Andrew Pickering and Keith Guzik (Durham: Duke University Press, 2008). Brian Marick, 'A Manglish Way of Working: Agile Software Development', in *The Mangle in Practice: Science, Society, and Becoming*, ed. Andrew Pickering and Keith Guzik (Durham: Duke University Press, 2008).

share information or data as they amass largely digital personal research collections? Do professional and non-professional historians have different attitudes to the sharing and collaboration practices enabled by digital technologies, and if so, why?

Research methods

As outlined in the *Introduction*, I used semi-structured interviews to understand historians' current research practices, how they evaluate resources, their attitudes to sharing and their current data sharing behaviours. The semi-structured script was designed to take a conversational tone and allow for digressions and the discussion of emergent themes, while providing specific prompts in each section to ensure consistency between interviews.²⁶ Using the same interview script for different types of historians militated against my making unintentional assumptions about the values or research processes of different historians. The interviews had a five-part structure. The main questions are summarised here and the full interview script is provided in *Appendix C: Interview and survey questions*. The interview began with some simple demographic questions, including the extent of the interviewee's experience with historical research and comfort level with computers, designed to put the interviewee at their ease. Further questions asked them to describe their research, then outline their research processes (i.e. the stages, resources and tools used). The language about tools and resources was deliberately vague so as not to lead their responses, but specific examples were available if required. In each section, general questions were followed with increasingly specific questions.²⁷ The first main section asked how they evaluated different types of resources

²⁶ For example, I varied some questions for participants in *Founders and Survivors* and FACHRS projects in order to explore their involvement with specific projects.

²⁷ This is described as a 'funnel approach'. Marcia J Bates, Deborah N Wilde, and Susan Siegfried, 'Research Practices of Humanities Scholars in an Online Environment: The Getty Online Searching Project Report No. 3', *Art History* 540, no. 3 (1995).

(e.g. books, databases, resources created by members of the public or amateur historians and blog/social media posts); the next section asked about their use of place, mapping and geospatial tools;²⁸ and the final section asked when and how they shared research data. The concluding section simply provided an opportunity to revisit any previous questions or make additional comments on any of the issues raised during the interview.

I used non-probabilistic sampling techniques to recruit interviewees, with a combination of purposive, snowball and self-selection sampling.²⁹ In selecting interviewees, I was looking for a wide variety of respondents within the wider frame of people researching lives in early modern England. I also aimed for roughly equal numbers of faculty and family/local historians, and a reasonable number of international interviewees.³⁰ I initially contacted or requested contacts through personal connections, posts to specialist mailing lists and on social media.³¹ Some later interviewees were recruited through the 'snowball' method, where early interviewees recommended others known to them who might meet my interview criteria. Other interviewees were referred by academic colleagues. I found

²⁸ This question reflects my original focus on spatially-indexed content and participatory resources.

²⁹ A probabilistic sampling technique is not possible when the overall frame size (i.e. the number of faculty and community historians) is unknown. Briony J. Oates, *Researching Information Systems and Computing* (London: Sage, 2006). pp. 95-8.

³⁰ My final sample included 17 UK-based and 12 overseas interviewees. Allowing for overlap between categories, it included 12 faculty historians, 12 family historians and 7 local historians.

³¹ For example, a Facebook post in February 2012 update requesting interview participants generated several potential leads from contacts who suggested friends and family members. A tweet sent on February 16, 2012 saying 'Are you a family or local historian in London or Oxford, or know someone who is? I'll be looking for interviewees for [<http://www.miaridge.com/my-phd-research/information-for-potential-research-participants/>]' also led to some interviews. The link referred to a web page made to provide further information for potential participants, written in non-specialist-friendly language with headings such as 'What I hope to learn', 'Who I want to talk to' and 'What happens during an interview, and afterwards'. I also posted requests for research participants to other sites, including academia.edu and the womenshistorynetwork.org mailing list.

that data saturation occurred within the final sample, suggesting that it was of an appropriate size.³²

Some interviews were conducted in-person, with locations including interviewees' homes or university offices, and public spaces such as cafes. Other interviews were conducted through Skype.³³ Two interviews were conducted over email, and two Skype interviews were supplemented with questions answered over email. The interview locations were chosen according to the preferences of the interviewee and the logistics of travel between our mutual locations. Those interviewed in a location where they do some of their research were able to show examples of work-in-progress to the interviewer, but there were no other significant differences linked to the interview location. With the consent of participants, interviews were recorded with a personal digital recorder or via software from Skype calls. One in-person interview was not recorded, as the interviewee was not comfortable with technology. Most interviews took 1 to 2 hours, with an average length of 1 hour 20 minutes and approximately 40 hours of recordings in total.

Sixteen interviews were transcribed in full (up to 13,600 words per interview, average 8,200 words per interview); notes were taken during the other interviews and supplemented with exact quotes upon re-listening to the recordings (average 4,400 words per interview), leading to a total corpus of 188,600 words. As discussed in the *Introduction*, the transcripts were coded in NVivo software using thematic analysis in which a process of

³² Data saturation is the point in data collection and analysis when 'new information produces little or no change' to the codes applied. Greg Guest, Arwen Bunce, and Laura Johnson, 'How Many Interviews Are Enough?: An Experiment with Data Saturation and Variability', *Field Methods* 18, no. 1 (1 February 2006): 59–82, doi:10.1177/1525822X05279903.

³³ A Skype number was used so that participants could call from their phones if they preferred.

open coding was followed by axial coding.³⁴ I anonymised the interview data by assigning the code names used throughout this document, and in some instances I have changed small details of places, names or research questions to further anonymise the data.

About the interviewees

Most interviewees had a long interest in history, often from childhood, while a few developed an interest in history later in life. Motivations for the family and local historians interviewed were similar to those for participants in voluntary and crowdsourcing projects, and included the intrinsic rewards of learning, discovery and enjoyment. Some additionally sought to keep active in retirement. The inherent variability of historical research provides a constant intellectual challenge that helps keep some interviewees interested. Social interactions around their research sustained some interviewees, whether with family discovered through their research, online communities of researchers, or other members of specialist societies. Tracing their family and understanding the times and places in which they lived was important for family historians, and some older interviewees were also motivated by a desire to finish their research and/or record their family history while they were still able. For faculty historians, historical research is directly related to their employment. One senior faculty historian summarised the motivation for professional historians in stating that publishing research has a direct impact on 'your promotion prospects and your salary'.

Responses to the questions 'how did you learn to do historical research?' demonstrate the varied paths to historical practice. Some historians working in academia were trained in related disciplines and learnt historical skills 'in the field', as interviewee Sarah put it. For some academically-trained historians like Claire, Anne and John, their degree was an

³⁴ Braun and Clarke, 'Using Thematic Analysis in Psychology'.

opportunity to undertake historical research and gain access to guidance from more experienced historians, but a feeling that 'I wasn't taught, I learned on the job' was more common than one might expect from academically-trained historians.³⁵ Some self-taught historians related their methods to the disciplines in which they trained - one uses 'electronic troubleshooting techniques' and another says his science degree taught him to be logical and methodical. Family historians like Martin, Nell and Peter received some training through tutorials, conferences and online material provided by genealogical societies and library-based family history research groups. Learning through discussion with others was a key theme in my interviews. Emily had previously studied for a related heritage degree but says she learnt to do historical research through reading and online discussions; similarly, Martin read and asked 'a lot of questions' through a Genealogical Society and local history societies. Doris learnt by observing more experienced colleagues, Ed and Oliver learnt from other family members, and Nell has consciously learnt all she can from others doing family history research, whether distant relatives or members of her genealogy club.³⁶

³⁵ For example, John says 'I did a little piece of research off my own bat and then pestered people, one individual in particular, about what I was doing'. This refrain is also present in external interviews: Terence Ranger is quoted as saying 'I wouldn't say I was "trained" as an archival historian; I just got an awful lot of practice at being an archival historian'. In Diana Jeater, 'Terence Ranger: Life as Historiography', *History Workshop*, 16 July 2011, <http://www.historyworkshop.org.uk/terence-ranger-life-as-historiography/>. One university archivist who surveyed his colleagues on their role in educating students corroborates this. He reported that '[s]tudents get a range of instruction from their professors but when encountering original documents for the first time, they are rarely fully prepared'. Marcus C. Robyns, 'The Archivist as Educator: Integrating Critical Thinking Skills into Historical Research Methods Instruction', *American Archivist* 64, no. 2 (2001): 363-384, <http://archivists.metapress.com/index/Q4742X2324J10457.pdf>.

³⁶ The club is organised by a nearby public library.

The family and local historians interviewed had a range of skills and experience, from novices, to the academically-trained, to the autodidact with decades of experience. Their approach could be described as 'serious leisure', using Stebbins' label for the systematic, long-term pursuit of an activity to the point where they had acquired 'its special skills, knowledge, and experience'.³⁷ For example, regardless of their formal training, it appears that all of the more experienced researchers had gained 'archival intelligence', an understanding of the logic behind archival rules and procedures, and the ability to develop search strategies appropriate for their research question.³⁸ While they work outside GLAMs and academia, they are not what public historian Benjamin Filene called 'outsider history-makers', '[f]reed from scholarly and professional conventions' and mostly concerned with 'establishing emotional connections to the past'.³⁹ Most interviewees were careful to observe scholarly conventions. Some, though not all, prioritised emotional connections to their family or local area's past, but not at the expense of good practice.⁴⁰

Historians' current practices

Some activities, tasks and sources discussed were relevant to multiple stages of the research process. Specific stages relevant to the overall research question discussed below are discovering, evaluating, gathering, creating, and sharing. I have used UML's definitions for discovering, gathering, creating and sharing,⁴¹ supplemented with an additional

³⁷ Stebbins, 'Casual Leisure'.

³⁸ Elizabeth Yakel and Deborah A. Torres, 'AI: Archival Intelligence and User Expertise', *The American Archivist* 66, no. Spring/Summer (2003): 51–78.

³⁹ Benjamin Filene, 'Passionate Histories: "Outsider" History-Makers and What They Teach Us', *The Public Historian* 34, no. 1 (February 2012): 11–33, doi:10.1525/tph.2012.34.1.11.

⁴⁰ In fact, a common refrain in the interviews described 'bad' historians who let their desire to find connections override the requirement for multiple sources of reliable evidence.

⁴¹ Therefore, 'discovering' refers to any serendipitous or structured searches that lead to finding and identifying resources; 'gathering' refers to acquiring and organising resources; 'creating' includes activities undertaken with those resources, including analysing, synthesising, annotating,

heading that reflects the focus of my research - 'evaluating', which occurs between the discovering and gathering stages, as sources are assessed before being ignored, read and/or gathered.

The beginning stage of a research project tends to be different for family and faculty historians. Family historians usually start with an individual or family, looking for documentation for commonly recorded events including births, deaths, marriages and census records. Once they have confirmed a connection to their family, they expand their searches to other sources including historical newspapers and other written sources. Many of the historians I interviewed subsequently seek to go beyond 'names and dates', researching related occupations, broader historical events or locations in order to contextualise their lives.⁴² Faculty historians tend to have a topic or archive in mind, and start with a general survey of existing scholarship in the field before investigating promising sources. Like professional historians,⁴³ they are often working to a particular deadline or goal, whether a conference paper, thesis, article or book.

The interviewees described several forms of 'cross-cutting'⁴⁴ or iterative activities that

collaborating, teaching; and 'sharing' includes all aspects of dissemination and sharing data, ideas, draft or completed works. University of Minnesota Libraries, 'A Multi-Dimensional Framework for Academic Support'.

⁴² This echoes an earlier study where family history researchers said they would develop new research questions to replace those they had solved. Richard Butterworth, 'Information Seeking and Retrieval as a Leisure Activity', *DL-CUBA* '06, 2006, 29–32, http://web.archive.org/web/*/http://www.ucl.ac.uk/events/dl-cuba2006/papers/Butterworth.pdf.

⁴³ Typically academics and authors who are paid for historical research or writing.

⁴⁴ 'Cross-cutting primitives' are those associated with more than one activity. Palmer, Teffeu, and Pirmann, 'Scholarly Information Practices in the Online Environment'.

occurred in multiple stages.⁴⁵ At a micro level, researchers revisit handwritten documents to help them make out words that were indistinct on first viewing, particularly as their experience with a particular hand and the language of the document grows. Kathryn and Martin said that re-reading documents can help them spot information or references they missed the first time. Many move iteratively between primary sources and secondary literature as questions are raised through their encounters with sources; depending on the subject or research questions, some may seek further primary sources or talk to archivists or collections staff. Writing is often also an iterative process that might require finding and collecting more sources. Professional historians tend to read more broadly in order to situate their analysis in relation to other scholarship and key debates in their field as they write. Finally, creating non-written digital outputs often requires much iteration as the scholar negotiates technical and source-related constraints. For example, Charles tried several different resolutions for a map he was creating to find one that supported zoom but was not so big that his computer could not manipulate the file. He also experimented with readable colour schemes and layouts that fit names neatly over the map.

Note taking, transcription and annotation had different roles in different contexts and stages. Note-taking practices seem to depend in part on the interviewee's past experience,⁴⁶ on the amount of time a researcher has with an archive and how difficult it would be to return to that archive. For example, Helen prefers to take notes while at the archive, but has also photographed pages from more detailed sources when short of time on-site.

⁴⁵ These activities used a combination of digital and physical tools, depending on the source format and the goals of the researcher.

⁴⁶ For example, Doris' experience in the workforce means she creates handwritten summaries of texts, recording the full text if it's 'an essential quote' while in the archive.

Transcription has a practical purpose in allowing researchers to record specific details and quotes from documentary sources,⁴⁷ but it also has an intellectual role closely related to the meaning-making process of annotation. Despite the fact that she cannot touch-type, Kathryn will transcribe entire wills if they are not available online, as 'the process of transcribing something like that does make you confront precisely what it says'. Doris also transcribed all letters and diaries written by the subject of her research; while it was time-consuming, 'it meant that I am now familiar with all that she did and with her feelings about various situations'. She also created an index of subjects, personal names, places and themes by reviewing the documents; the process was 'quite laborious' but means she can work more quickly as she has 'absorbed' the material. These interviewees linked the close attention they paid to the text with the process of transcribing it.

The interviewees used a variety of sources, from repositories of birth, death and marriage certificates and census enumerator records to the Statutes of the Realm. Source formats ranged from gravestones to hobbyist websites and academic books. Some interviewees particularly valued sources, such as military records and historical newspapers, that provided additional social context or physical description. Other sources mentioned include pension records, court reports, unpublished theses, maps, employers' records, wills, and street and trade directories.⁴⁸ Online resources included academic projects, 'synthetic'⁴⁹ commercial documentary collections, *Access to Archives*,⁵⁰ free and

⁴⁷ For example, John, Kathryn and Anne transcribe text into *Word* tables.

⁴⁸ The variety of sources used is not itself a new phenomenon, see for example Donald Owen Case, 'The Collection and Use of Information by Some American Historians: A Study of Motives and Methods', *The Library Quarterly* 61, no. 1 (1991): 61–82.

⁴⁹ The term one interviewee used to describe sites similar to print anthologies.

⁵⁰ Now part of The National Archives Discovery service, *Access to Archives* was a catalogue of records held by local record offices, GLAMs and other institutions. Jonathan Cates, 'Discovery - Finding More Archives', *The National Archives Blog*, 18 September 2014,

commercial genealogical aggregators and personal websites. Resource topics were varied, including newspaper archives, websites on specialist occupations from railwaymen to clergy, and collections of manuscripts and early print books. Printed books are still important, but seem mostly to be used for general, long-form background information on a specific topic or for sources not available online.

Historical newspapers were mentioned in several interviews. Claire, John and Martin had spent time immersed reading newspapers relevant to their areas of research. Other interviewees found newspaper articles through searches for named individuals. The Australian repository *Trove* was mentioned with particular affection by several interviewees, in part because its design suits the need of typical users.⁵¹ *Trove* provides access to images of the original documents, and it is freely available and well-indexed by search engines so appears in internet search results.⁵² Newspaper articles can provide information not easily found in other sources, including physical descriptions of people, their clothing and houses; the names of extended family members (Larry discovered a previously unknown branch of one family), and other contextual information including social events, participation in community life, obituaries and appearances in court. Claire values the glimpses newspapers provide of the historical figure, and seems amused to

<http://blog.nationalarchives.gov.uk/blog/discovery-finding-archives/>. Barbara used *Access to Archives* to focus her discovery efforts on specific archives, and in one instance was able to use the information provided to request a specific file from a distant archive.

⁵¹ Peter compared it favourably to New Zealand's *Papers Past* site, which does not include the whole of the original page when presenting individual articles and is 'not quite as easy to navigate around'.

⁵² A small caveat: most of these interviewees were Australian or researching Australian individuals, but *Trove* was also mentioned by others, including a senior faculty historian who referred to 'the splendiddness of Trove'. Only one interviewee mentioned paying for newspaper access (a subscription to the *British Newspaper Archive*). Future research investigating whether freely available historical newspapers online have encouraged a wider range of people to undertake contextual research would be interesting.

discover that 'people who have been held up to me as pillars of society' were in fact slightly less respectable.

Genealogy sites like *Ancestry* and *Findmypast* are not only repositories of aggregated records from many sources, but they also provide software for creating family trees and managing references to sources. *Ancestry* evokes strong feelings, including comments from nine interviewees on subjects including the expense of subscribing,⁵³ the numbers of incorrect family trees hosted on the site, and the behaviour of novice researchers and/or 'bad historians' on the site. Participants working on prosopographical or community history projects also use records from *Ancestry*. Census enumerator data held by *Ancestry* was used in The Family and Community Historical Research Society (FAHCRS)'s *Station Masters* and *School Mistresses* occupational studies, and a range of records are used by participants in the *Founders and Survivors* projects.⁵⁴ *Findmypast* is *Ancestry*'s chief competitor in the UK and was mentioned by several family history researchers. Other family history sites and tools mentioned include *FamilySearch*, *Cyndi's List*,⁵⁵ *Curious Fox*,⁵⁶ *JewishGen*, *FreeBMD* and *Origins Network*,⁵⁷ *PHPGedView* (self-hosted online family

⁵³ Most people subscribe to either *Findmypast* or *Ancestry* depending on the types of records they most use; Claire is unusual in subscribing to more than one service and only did so after making a conscious decision to spend money on her family history. Ursula subscribes to two services as genealogy sites are useful for her academic research. A few other interviewees make their research more affordable by accessing commercial genealogy sites at their local libraries.

⁵⁴ *Founders and Survivors* aims to find records for over 70,000 people transported to Tasmania, and ideally to trace any descendents who enlisted in the First World War. *F&S* aims to collect the results of existing family history research and to recruit already-experienced genealogists as volunteers. Founders and Survivors, 'Ships Project News'. McCalman et al., 'Building a Life Course Dataset from Australian Convict Records'.

⁵⁵ Effectively a portal for family history researchers.

⁵⁶ Used by those looking to connect with other researchers interested in particular names or places.

⁵⁷ Bought by *Findmypast* in June 2014.

tree software, now *WebTrees*), *GenesReunited* (which also hosts members' family trees), *Reunion*, and *Legacy*.

Discovering

Finding relevant records is an important stage in the research process, closely linked to the discoverability (that is, the ability to be found) of primary and secondary sources. The discoverability of records is limited by many factors, including the types of information originally recorded in written form, the survival of physical objects containing that information, the cataloguing systems that represent those artefacts, and finally the digitisation processes that make catalogues, images, indexes or full transcriptions available online. Some family historians had a good working knowledge of the impact of time on the survival of records, including records lost during the Irish war of independence and the Blitz. Others related stories of family records or photographs destroyed before they could record them. The discoverability of records is also influenced by the interests of previous generations of historians; for example, the records compiled by contributors to the original *Victoria County History* had helped some of my interviewees, and some credit the interest (and wallets) of family historians with helping save local records offices. Finally, commercial subscriptions and the need to travel to specific sites affect the discoverability of sources. For example, historians outside well-funded universities do not often have access to commercial digitisation projects and the full range of journals.

Interviewees tend to accumulate mental lists of library catalogues, journal sites and primary source repositories or aggregators they will turn to for particular types of sources or information. They also review information or sources they had previously collected or look through the bibliographies of published works. Historians additionally learn about resources by word of mouth (in conversation or through social media). Some interviewees email institutions with specific enquiries about physical or online resources, either in

search of material not online or as part of their collection evaluation process. Faculty historians may also receive information about new databases from their university librarians, and researchers registered with online sites may subscribe to journal and publisher content alerts, as well as notices about newly available sources. Some participants reported benefitting from 'serendipitous altruism', when peers who knew of their research topics passed on relevant references or information.⁵⁸

Interviewees are willing to travel to specific libraries or local archives centres but also appreciate the convenience of online access because, as Claire said, 'it doesn't involve trekking around the country'. However, online access may have disadvantages. For example, it may be easier for researchers to forget to look for undigitised resources to supplement those available online.⁵⁹ Theresa and others felt that fewer visits to physical archives might provide fewer opportunities for serendipitous discoveries through physical adjacency or conversations with archivists, other staff and researchers.⁶⁰ Novice researchers may not have the same opportunities to gain confidence in handling historical materials or to observe more experienced researchers at work. One faculty historian also worried about students producing 'two dimensional work' as a result of relying entirely on digital repositories rather than supplementing them with other material found during archival visits.⁶¹ Another mentioned the 'EEBO effect',⁶² in which researchers without a

⁵⁸ Michael B. Twidale, David M. Nichols, and Chris D. Paice, 'Browsing Is a Collaborative Process', *Information Processing & Management* 33, no. 6 (November 1997): 761–83, doi:10.1016/S0306-4573(97)00040-X.

⁵⁹ A fear mostly expressed by historians on behalf of other researchers (particularly students), but one said that he was also at risk of forgetting to review physical resources.

⁶⁰ However, online resources provide opportunities for other forms of serendipitous discovery.

⁶¹ In other conversations, faculty historians worried that relying on keyword searches would lead to researchers only reading snippets of text rather than the entire paragraph or chapter.

⁶² EEBO is the *Early English Books Online* repository.

broad knowledge of the field assume that repositories are comprehensive and that every relevant source has been digitised.

Interviewees gave many reasons why online access to digitised historical materials is more convenient than physically visiting an archive. Anne works near print journals in the library but still prefers to look at journal articles online where possible: 'it's just quicker and easier if the journal is available as a PDF. And I can do word searches...'. Another faculty historian will look at digitised sources rather than go to the library, partly because 'you can look at it briefly rather than having to order it up'. The activities of moving from primary to secondary sources then returning to the archival sources can all be conducted from one's desk.⁶³ While there was an undercurrent of worry that researchers will not realise that undigitised resources exist or will forget to consult them, one interviewee's comment that she used local archives over more distant ones when she had young children, and others' experience of planning trips around archives abroad, are reminders that access has always affected the use of archives.

One of the most interesting aspects of full-text digitisation is that it turns every phrase in a document into a potential search result. When these texts are available for indexing by search engines, very specific queries may return obscure phrases previously buried in the pages of books, newspapers and other documents. While Tibbo's 2003 study of historians found that only 44% used search engines,⁶⁴ the use of search engines now seems ubiquitous. Most interviewees will try searching for a name or topic of interest in the

⁶³ However, working entirely from a screen has disadvantages. Doris described 'a certain amount of backwards and forwards' when trying to take notes from a source into a document on the screen.

⁶⁴ Helen R. Tibbo, 'Primarily History in America: How US Historians Search for Primary Materials at the Dawn of the Digital Age', *American Archivist* 6 (2003): 9–50, <http://archivists.metapress.com/index/b1203701hg718n74.pdf>.

expectation of finding some results.⁶⁵ As one faculty historian said, when researching particular individuals, 'one needs to be rather eclectic ... you've just got to try everything'. Doris has also been impressed by the availability of useful information about almost any obscure topic online. The range of resources now discoverable makes the task of assessing the quality of a given resource even more important; this is discussed further below.

The increased use of search engines by historians means that discoverability in search engines is vital for content holders.⁶⁶ For example, Helen has found it difficult to find information about small or regional museum collections. In 2003, Tibbo's study found that newspapers were the 'most often used' and 'most important' materials for historians, but the prevalence of name searches and digitisation might have made them even more prominent.⁶⁷ A 2009 synthesis of user studies on archives and special collections found that archive users do not understand 'archival standards for description and cataloging' or want to search for collections by institutional provenance.⁶⁸ Better indexing of historical repositories by search engines may help these users.

Generally the interviewees were aware of the large amount of material that has not been

⁶⁵ Claire reports finding a specific name within a passage of text used in an exam paper, which eventually lead to her finding a cottage where people she was researching had lived.

⁶⁶ The dominance of Google may also account for the popularity of Google Books over other repositories of historical texts among the interviewees.

⁶⁷ 'Unpublished correspondence' was next on both headings. As it is currently harder to create automatic full-text transcriptions of handwritten documents, future research into the impact of digitisation on the most commonly used materials may help quantify the impact of greater discoverability. Emerging technologies such as handwritten text recognition will also have an impact.

⁶⁸ Jennifer Schaffner, 'The Metadata Is the Interface: Better Description for Better Discovery of Archives and Special Collections' (OCLC Research, May 2009), <http://www.oclc.org/programs/publications/reports/2009-06.pdf>.

digitised, and of the variation in the levels and forms of digitised material from different places. The financial incentives that lead GLAMs⁶⁹ to prioritise the digitisation of materials used by genealogists can cause resentment, as can the subsequent locking of these materials behind pay walls and interfaces that prioritise searches for named individuals.⁷⁰ Some felt that search engines were not locating all the resources that exist online.⁷¹ Martin reported that 'you have to be a little bit creative and try and add in separate words' to get more specific results, particularly as he feels Google favours American resources.

Some interviewees also discussed the impact of search functionality within sites. In enabling or impeding different searches, search boxes mediate the researcher's experience of a repository. Peter notes that *Ancestry* requires three characters in a search string before a wildcard search character can be used,⁷² which makes it harder to search for Victorian abbreviations for names, and thereby trace individuals. Martin values advanced search functions that help him make search terms more precise with Boolean operators ('those and/or's') to combine, include or exclude certain words. The provision of algorithms like

⁶⁹ Galleries, libraries, archives and museums.

⁷⁰ Andrew Prescott, 'Dennis the Paywall Menace Stalks the Archives', *Digital Riffs*, 2 February 2014, <http://digitalriffs.blogspot.co.uk/2014/02/dennis-paywall-menace-stalks-archives.html>. In reviewing projects I also came across an instruction to indexers transcribing documents related to relief funds for the Irish Famine on *Ancestry's World Archives Project*: 'Do not key companies'.

'File:IrelandFamineRelief List Key 1.JPG', *Ancestry.Com Wiki*, 3 December 2014, http://www.ancestry.com/wiki/index.php?title=File:IrelandFamineRelief_List_Key_1.JPG. This clearly reduces the usefulness of the index for any historians searching for particular companies.

⁷¹ This may be because sites that require search queries to generate lists of resources are often 'hidden' from search engines. Alex Wright, 'New Search Technologies Mine the Web More Deeply', *The New York Times*, 22 February 2009, sec. Technology / Internet, <http://www.nytimes.com/2009/02/23/technology/internet/23search.html>.

Scholarly articles may also not be discoverable in search engines unless the researcher knows about specialist versions such as *Google Scholar* and *Microsoft Academic Search*.

⁷² On other sites he could search for 'J*' to find all variations of John, Jno, etc.

Soundex or *NameX* that look for phonetic matches for names can help find more possible matches for a name.⁷³ Many genealogy sites have the same sets of records, so researchers will subscribe to one commercial site but use others to supplement that site's search.⁷⁴ For example, Bob and Ed subscribe to *Ancestry* but prefer *Findmypast's* search, so they use it to locate records then view the original documents on *Ancestry* rather than paying to view the document on *Findmypast*.

Evaluating

Understanding the qualities that a historian looks for when evaluating a newly discovered primary or secondary source helps those publishing digital resources understand what information should be included and how they should be presented. Here I summarise the interviewees' responses to questions about how they evaluate resources, and present a synthesis of the characteristics of authoritative and/or useable resources. My interview questions asked which qualities were important when assessing a newly discovered resource, whether they varied their methods when evaluating digital resources or social media content, whether they would use resources created by the public or amateur historians, and whether they would use some resources without relying on them. Responses discussed both books and sites as a whole, and individual sources within larger works.

Many initial responses drew on the interviewee's tacit knowledge: Doris said you can tell 'from the way the material is presented', and Ed offered 'an intuitive kind of internal check

⁷³ The U.S. National Archives and Records Administration, 'The Soundex Indexing System', *The U.S. National Archives and Records Administration*, 30 May 2007, <http://www.archives.gov/research/census/soundex.html>. 'What Is NameX?', *NameX*, accessed 2 February 2015, <http://namevariants.co.uk/introduction.aspx>.

⁷⁴ This requires an investment of time as it can take time to learn the peculiarities of different sites' searches, and sometimes the searches for different repositories within a site.

of things'. Many interviewees, including Ed, Peter and John, used variations of the phrase 'feels right' to describe resources that fit their past experience. Design and usability expert Don Norman describes these unconscious 'snap' judgements as 'visceral' responses - automatic evaluations of 'the perceptual properties of objects, and a quick classification of them as safe or dangerous, good or bad'.⁷⁵ These judgements are driven by 'pattern recognition mechanisms' driven entirely by 'the here-and-now of perceivable features'. In other words, first impressions count. On prompting, most interviewees expanded on the features that contributed to their judgement.

The correct citation of sources (i.e. the presence and quality of footnotes) and the quality of the sources used are both important. A lack of evidence is an immediate cause for suspicion as it both suggests a lack of skill and means any claims cannot easily be checked. Footnotes help contextualise an argument within the field, and, in Anne's words, indicate whether is it based on 'reputable sources or primary research work, or whether it's just supposition, speculation'. Helen needs to feel confident that any material she uses has itself been properly referenced so that she can find and view their sources. Doris is reluctant to use resources that do not reference any scholarly sources, and looks for those referencing 'original documents or original printed sources'.

Resources were sometimes evaluated by verifying information they contained. Most interviewees use multiple sources to corroborate facts found in online or offline sources.⁷⁶

⁷⁵ Donald A. Norman and Andrew Ortony, 'Designers and Users: Two Perspectives on Emotion and Design', in *Proc. of the Symposium on Foundations of Interaction Design at the Interaction Design Institute* (Ivrea, Italy, 2003).

⁷⁶ Similarly, educational psychologist Michael Wineburg found that corroboration was one of the three key heuristics that academic historians applied to unknown resources. Sam Wineburg, 'Historical Problem Solving: A Study of the Cognitive Processes Used in the Evaluation of

Daniel finds that confirming facts 'by seeking them elsewhere' is often easier and faster with online resources. Peter and Gina test websites or databases by searching for information already known 'to see if I get that information back'. Ed's experience with alternate spellings and name variations means he can get a sense of 'how good the transcription is on a particular site' based on the number of results returned by a name search.

Similarly, resources are judged against a researcher's knowledge of other primary materials and/or scholarship on the topic, with resources that conflict with accepted accounts or material found elsewhere viewed with suspicion. Daniel becomes suspicious when resources repeat phrases found elsewhere without attribution. This evaluation is not always a conscious process. In Kathryn's words, 'you just look at it and think about whether it fits with other things you know', and Anne explained, 'I've got enough knowledge of the area that I work in to be able to see whether something is correct or not'. The value of this sense of what is normal or expected, developed after exposure to many similar records, emerged from many interviews.

Many interviewees consider the presence of information about an author and their credentials to be important, while resources with unclear authorship are treated with suspicion. For example, Helen wants to know what credentials the author has in terms of their experience and academic background, and Anne would take a blog written by someone who presented information about their academic credentials more seriously.⁷⁷ Doris will look first for books written by 'a proper historian', a categorisation that may be

Documentary and Pictorial Evidence.', *Journal of Educational Psychology* 83, no. 1 (1991): 73, <http://psycnet.apa.org/journals/edu/83/1/73/>.

⁷⁷ Along with some other interviewees, she is generally sceptical of social media, believing that researchers in her field do not post about their research.

based on academic credentials or the judgement of other trusted researchers. Evidence of institutional origin is also seen as a positive. Interviewees mentioned research councils, national archives, university libraries, JSTOR, Project Muse, and what one described as 'the more recognised publishers of the big databases'.⁷⁸ It should also be noted that several participants, including Theresa, Tom, Wendy and Yvonne, reported that they did not have access to commercial repositories or journal sites, and some described strategies including travelling to other research libraries or asking friends to look up material. Conversely, content found on social media and self-publishing platforms (such as blogs or personal websites) is regarded as potentially interesting but entirely subject to requirements of evidence, including corroboration with other sources, checking citations or viewing images of original documents.

Relevance is a key factor when deciding whether or not to use a resource: as one senior academic said, 'the main thing is, does it actually contain anything on your immediate concern'. Interviewees had various techniques for assessing relevance. Martin does 'a little bit of a reconnaissance', skimming the site for relevant material. Ed uses site search to look for relevant records in genealogical sites, then evaluates the quality of the sources listed as results.⁷⁹ This bears out earlier research on 'information seeking' behaviour which found that scholars quickly assess documents for their 'relevance and utility'.⁸⁰ The depth and accuracy of information is also a factor for sites with interpretive information or scholarly arguments. Finally, the languages in which records were written are a factor in determining their use. Charles was grateful to find that the local legal records he wanted were in English rather than Latin, but language is not always a barrier - Sarah brought a

⁷⁸ Such as Gale Cengage or ProQuest.

⁷⁹ When the results are behind pay walls, this might mean evaluating the overall reputation of the repository listed rather than individual records.

⁸⁰ Palmer, Teffeau, and Pirmann, 'Scholarly Information Practices in the Online Environment'.

friend who could read Latin with her when she accessed some archival records.

Many interviewees had come across resources created by amateur historians. Personal family history websites are discoverable through name searches; narrative information and transcribed or indexed records posted on local history sites are another common source. Emily applies the same scepticism to those as other sources. Helen will use well-referenced and well-written non-academic material, being 'less concerned about the format of the information than the quality of it and the person/people who created it'. Peter reports that some family history sites are 'excellent' while others are 'cobbled together'. Kathryn might look at sites listed in search engine results but tends not to use them. The correct use and citation of sources may encourage some interviewees to consider an amateur site as potentially reliable.

Resources are also evaluated differently depending on the purposes for which they might be used. Helen limits herself to 'traditional' sources when writing for a journal but will otherwise consider a wide range of sources, and in some circumstances will cite 'poor quality information' from websites as long as she can cite the author and date. There also seems to be a key difference between the resources a researcher will use for factual information or to inform their sense of a field, and those they will rely upon when making an argument or presenting evidence for a claim. Doris will reluctantly quote some contemporaneous sources she feels are 'rather impressionistic' when the absence of other sources means 'it is the closest I will get to a clear account'. Secondary sources are often read to provide background information, or for a sense of the key debates about a topic, but these are not always cited in the final publication. Some use *Wikipedia* to quickly look up factual information like a year of death, but some are more trusting than others - Anne would 'still like to confirm it in some sort of printed form' while Charles feels he can rely

on *Wikipedia* for a date (even when he knows that the entry for his own village contains errors). Helen considers well-referenced *Wikipedia* articles as 'encyclopaedic background reading that would lead me to consult the sources it was composed from'. Interviewees differed on whether they judged sites as a whole or evaluated individual contributors or authors on participative sites, although this difference may be partly due to the types of sites they typically encounter.⁸¹

Family trees published online are a good example of the distinction between resources used or relied upon. For the interviewees, family trees are not all created equal. The family trees on sites like *Ancestry* or *MyHeritage*⁸² could be the results of decades of painstaking research or may have been created by a novice accepting every 'hinted' possible match during a free trial period.⁸³ As Peter described some novice genealogists, 'they think that because they've found a name, it's got to relate to them'. While some issues with family trees can be attributed to novice genealogists making their mistakes in public, others may be the results of wishful thinking; in either case the ease of copying information from a public tree means that bad data can spread rapidly. Consequently, most experienced family historians have learnt not to trust these trees or the research they represent. Ed

⁸¹ Daniel pointed out that 'information on a particular poster is likely to consist mostly of their other contributions to a site', but Theresa and Ursula felt that forum posts were useful in assessing a contributors' skills and knowledge.

⁸² *MyHeritage* claims to have 28 million family trees. 'Free Family Tree, Genealogy and Family History', *MyHeritage*, accessed 7 April 2015, <http://www.myheritage.com/>.

⁸³ *Ancestry* describes hints as 'potential matches' based on comparisons of a family tree with 'billions of records, photos, trees, and more'. Hints are signified by a 'shaky leaf'. 'Hints', *Ancestry*, accessed 7 April 2015, <http://www.ancestry.com/hints>. *MyHeritage* has a similar function called 'Smart Matches'. The practice of accepting 'hints' and/or attaching someone else's family tree to yours without verifying the correctness of the match are notorious to the extent that a blog called 'Barking Up the Wrong Tree' collects examples under the tag 'clickophile': <http://buwt.blogspot.co.uk/search/label/Clickophile>. Last accessed 8 June 2015.

described his feelings about a family tree created by another genealogist which goes back 300 years further than his: 'I've got no reason to necessarily say that it's wrong', but he would need convincing that information beyond a certain point in time was based on reliable sources. However, Emily echoed Peter, Bob, Claire and others when she said family trees are 'sometimes a good starting point if you are stuck' with a brick wall;⁸⁴ all would look for further documentation to corroborate information found rather than rely on an online tree.⁸⁵ This also illustrates a related theme in which using pointers to primary sources is regarded as less problematic than relying on an unknown person's interpretation of those sources.

Transcribed sources introduce another layer of assessment, as errors could have been introduced when the data was transcribed. Bob assesses genealogical resources by checking for transcription errors that might have been copied from one database to another. Sites that allow transcriptions to be changed add further complexity, as the quality of corrections relies on the skills and intent of the corrector. Martin finds it 'very, very suspicious' when sites allow corrections to transcriptions but do not provide an 'audit

⁸⁴ 'Brick wall' is a term used when family historians cannot find information to take their tree further; as their research is led by the lives of their ancestors rather than their own interests, they are more than usually dependent on the fortuitous survival of, and access to, relevant sources. See for example the extensive resources listed at <http://www.cyndislist.com/brickwall/articles/>. Last accessed 8 June 2015.

⁸⁵ Larry generally values 'the connections and the history' over 'substantiating whether it was the 1st or 2nd July'. He is the only interviewee who takes online family trees on trust: 'if it's a whole branch that I don't already have, I just plug it in. I'm really not very discerning'. He acknowledges that it is 'a dreadful thing to say', but beyond cross-referencing it with other available sources, he says he does not have the financial resources to purchase certificates and check the research. He warns others that his tree may contain mistakes, and hopes that in future, 'when the information becomes all online, and you swipe your credit card, and you get 300 for the price of \$2, I'm sure someone will trek through all of that'.

trail' for those changes.⁸⁶ However, the most important factor in determining whether the sources provided on a site are regarded as reliable is the provision of access to images of the original documents. As Martin said, people 'have different interpretations of As and Es in a word, or numbers, so you've got to be very careful of that sort of stuff'.⁸⁷ If no transcription is provided, most interviewees will use the information provided to track down the original; if it cannot be found then the information is regarded as unproven. Images of documents not only help verify their content but in some instances they also provide information not included on the transcription or index. John appreciates the access to images of the original document provided by *Ancestry* and some archive sites, considering them no less valuable 'than actually going to the archives and seeing the original'.

When evaluating websites and digital resources, interviewees were particularly interested in the context in which a resource was created, and the credentials and aims of the individual or institutional publisher.⁸⁸ Martin and Daniel mentioned their need to know 'who had produced [a digital resource] and why'. This is partly because for-profit genealogy sites are judged differently from grassroots and GLAM sites. Martin wants to know whether some pages have been left out, about quality control during digitisation, and the relationship between an online repository and the order of documents in the source archive. Kathryn prefers repositories to 'synthetic' databases organised around a particular topic, as she wants access to everything rather than 'somebody else's idea of a

⁸⁶ He says comments on transcriptions (as on *FreeBMD*) are ok as both the original and suggested transcriptions are visible.

⁸⁷ However, he would make an exception for transcriptions certified by the institution that is providing it.

⁸⁸ Kathryn was the only interviewee who mentioned similar questions about the purpose of those who originally put traditional archives together.

collection of things that might be interesting'. Martin is interested in what has been excluded by sites making a particular argument, and listed various places where he would look for that information in different types of resources. Finally, Helen likes to cite a resource 'so that others can find it' but is concerned that some digital resources will not be available when a future scholar looks for them.

While 'enthusiasts' may be viewed with more suspicion than credentialed historians, individual researchers are judged on a range of attributes related to the material they have shared: that is, they are judged on what they have done rather than who they are. Interviewees provided a range of instances of this judgement in action, and in some cases provided explanations for it. For example, one family historian discussed another's failure to realise that sources themselves can be wrong: 'people don't think that they've changed spellings. They don't realise that somebody turns up on the door and says, "who lives here" at census, and they say "well I think it's Mrs This and This", you know'. Others were censured for relying on only one source for a statement, for failing to interpret gaps in the archival record correctly, and in one instance, for amassing a large collection of copies of uncommon historical documents without recording where they were found; others were praised for using a range of sources.⁸⁹ One interviewee admitted that she could not help contacting people who had posted incorrect information, but most simply ignored the material.

⁸⁹ Many of these issues reported in others can be linked to a lack of historical training or experience. These attitudes may reflect the fact that most family historians I interviewed were quite experienced and had themselves learnt to use a range of sources and developed their own 'complex search strategies', in Yakel's words. Elizabeth Yakel, 'Seeking Information, Seeking Connections, Seeking Meaning: Genealogists and Family Historians.', *Information Research* 10, no. 1 (October 2004), <http://informationr.net/ir/10-1/paper205.html>.

Other attributes used when judging sites included graphic design and visual appeal. Martin contrasted the visual style of 'cheap and cheerful' sites with those government or other official sources he would consider 'reputable'. He has different expectations for grassroots local or family history sites but still appreciates 'professional', 'visually appealing' sites. Basic usability is important. Martin admits he will not persist when relevant information is buried within a site unless he thought it contained information he really needed - but older sites from long-established projects may continue to be used despite feeling 'clunky' (as Kathryn put it) if they contain useful content. However, as standards change, the content of older sites may seem less usable. For example, Martin has stopped using sites with poor-quality, hard-to-read images; Kathryn becomes suspicious when sites contain 'a lot of advertisements' and extensive misspellings; and Daniel prefers sites that provide information about when they were created. Finally, the reputation of a resource gathered via word of mouth recommendations or criticisms has some weight. Martin says he will hear over time whether other people think a resource is 'reliable, accurate, authentic', while Nell relies on librarians' assessment of resources, reasoning that they have been trained to do so.

Gathering

Responses to questions about acquiring and storing physical and digital resources showed that personal research collections are often a mixture of physical and digital resources. Other key findings are the range of techniques used to mark material that was not yet verified, and an overview of the winnowing process linked to the transfer of data or information between formats.

Most interviewees had information and notes in a mixture of physical and digital formats, including digital images, *Word* documents, family history software, physical documents and objects, notebooks and physical folders and files. For example, Bob has 'boxes and

boxes of bits of paper' but he also maintains records on a genealogy site, and has collected microfiche copies of 'all documents' from his ancestral town. Martin's files include notes on relevant locations, correspondence with family members and information that does not fit into family history software. *Word* documents are variously organised by research output, topic or source. Barbara, Karen, Peter, Kathryn, Tom and Yvonne use *Excel* spreadsheets to help organise their activities while Peter uses the notes function in *Founders and Survivors* to share detailed comments on his research progress and list possible records for an individual. Other tools used for gathering resources include *DevonThink* for managing documents and *del.icio.us* for bookmarks. Many interviewees like to make copies of documents where possible.⁹⁰ Digital images of documents may also be printed out and stored in physical folders for particular families or research subjects. When Helen takes photos of archival documents she stores them in *iPhoto* and attaches brief descriptive notes to them. Finally, interviewees reported two forms of illicit access to sources - using institutional contacts to obtain copies of documents without paying access fees, and photographing documents in contravention of the rules of an archive.

The importance of personal research collections that emerged in these interviews is not surprising, as it has been a theme of other studies of historians' practices. In 1989 Mathisen's informal survey found that nearly all prosopographers questioned at an event had created personal collections and concluded that they were 'in an age of small, personal databases created to satisfy specific research agendas'.⁹¹ Case's 1991 study noted collections of index cards, photocopies and computerised records.⁹² Palmer, Teffeau, and Pirmann discussed 'personal scholarly collections' consisting 'primarily of documents rather than

⁹⁰ Some have found methods for recording information outside the functionality provided by sites. For example, Ed takes screenshots on websites that do not provide an option to download images.

⁹¹ Mathisen, 'Where Are All the PDBs?'

⁹² Case, 'The Collection and Use of Information by Some American Historians'.

raw data'⁹³ while DeRidder and Matheny reported that many of their interviewees 'used a lengthy Microsoft Word document to compile information, aggregating anything from citations and metadata to copied content and research notes'.⁹⁴ Meho and Tibbo found that some of the value in personal record collections is in their ability to be organised or classified in ways that make sense to the researcher,⁹⁵ although the advent of full-text searching may have reduced the importance of this.

Interviewees had a range of approaches to documenting the sources they collected, with the more experienced and/or academically-trained scholars having the most consistent habits for recording references.⁹⁶ A few used reference managers such as *Zotero* or *EndNote*. Some noted their sense of responsibility as a scholar to record and cite references. Most family history software supports good referencing but Larry reported that the *Reunion* software package makes it particularly easy to link news articles from *Trove* to individual people. Some family historians update the evidence they reference for particular facts as they find better sources⁹⁷ or more precise information.

Gathering resources often involves transcribing text from source documents. Some interviewees take their laptops to archives so they can transcribe notes directly into digital formats, but others transcribe their handwritten notes when they return from the archive.

⁹³ Palmer, Teffeau, and Pirmann, 'Scholarly Information Practices in the Online Environment'.

⁹⁴ Jody L. DeRidder and Kathryn G. Matheny, 'What Do Researchers Need? Feedback On Use of Online Primary Source Materials', *D-Lib Magazine* 20, no. 7/8 (2014), doi:doi:10.1045/july2014-deridder.

⁹⁵ Meho and Tibbo, 'Modeling the Information-Seeking Behavior of Social Scientists'.

⁹⁶ As one said, the advantage of experience is knowing in advance that you are 'in serious trouble' if you fail to record archival references at the time. However, another senior academic confessed that he occasionally forgets to note references, so training and experience is no panacea.

⁹⁷ For example, a reference to a published index may be replaced with a full certificate reference.

For example, Ed will transcribe information found in a record office into his family history software from his handwritten notes or photocopied documents, while Charles types his notes from microfiche newspapers and archival documents into documents to create a searchable index of names linked to relevant material. Helen finds that processing information from books or documents 'flows better with notebook and pencil'; she will then try to capture the gist of her notes in a *Word* document, which is then filed in a subject-based directory on her computer. The benefits of being able to search digital documents were mentioned several times. Creating a digital version of something, even as a simple text document, means it can be emailed, shared and backed up online.

The process of transferring material from one format to another also seems to provide an important opportunity to review material. For example, Nell fills in printed forms (supplied by her genealogy group) by hand before transferring information to family history software when she is satisfied that she has the correct name and dates and has completed research for that individual. Printing out digital records can also be useful. Martin prints out his *Founders and Survivors (F&S)* life histories so he can review them and annotate points he thinks need further evidence. Some material is inevitably winnowed out during the transfer process, but generally the original notes are retained for reference.⁹⁸

The process of gathering resources also involves discarding some resources. Some interviewees spent a lot of time disambiguating personal names in order to be certain that they were gathering records about the right person. For example, Martin spent years tracking one individual, eliminating anyone else with his name by working backwards

⁹⁸ Rosemary demonstrates the value of these personal archives by pointing out that 'nothing is ever finished' as new opportunities may result from previous research.

through the records until he found proof that excluded them. Kathryn's research includes mothers and daughters with the same given name in the same family, and one of Peter's *F&S* ships had four 'James Smiths'. Peter's research for *F&S* includes instances where people deliberately used different variations of their name to mask their convict past. Confirming research by ordering the relevant documents that will help disambiguate historical individuals can also be expensive. The need to learn the skill of disambiguation is also one reason for 'bad' family trees, as described by Peter whose tree was erroneously linked to another family: 'they've just got a similar name and were living in the same town at the same time'. Most of the experienced family historians had stories about novice researchers assuming that any instance of a name must be the one they were interested in before they had discovered the need to disambiguate records.

Ensuring that information recorded is accurate is another key task. Anne reiterated the responses of many other interviewees when she said 'you never cite something unless you've seen it yourself'. Checking the original (or an image of it) is 'just good practice' that recognises that others can make mistakes, a sentiment echoed by Ursula. Bob only regards a family history fact as proven 'when I have two independent sources'.⁹⁹ As this is not always possible, it 'can often mean you don't say anything'. This need to keep as-yet unproven information separate from evidenced statements is one reason for keeping resources in multiple formats. Participants used a range of methods to record the status of potential evidence (information in the process of being confirmed). Some used the location of documents in physical or computer folders to mark their progression through various stages.¹⁰⁰ While Bob uses the words 'I believe but have not yet proven' in his notes,

⁹⁹ As did most interviewees.

¹⁰⁰ For example, in Martin's case, the stages are 'what's proven, or what's still to be researched, what's still of interest' and what's not to be used.

Claire 'flatly refuses' to put down unproven information. To an extent the method used depends on the format. Historians are used to qualifying or hedging statements in text, but, as family history software generally only allows for simple assertions of a relationship, unproven information may be stored in notes attached to personal records, general 'shoebox' (*Ancestry*)¹⁰¹ or 'source box' (*FamilySearch*) spaces, or remain in physical notes until it can be proven. Written notes also allow for question marks or further qualifications, while database fields generally exclude these markers of ambiguity. Most family history software excludes the attached notes that contain information on the certainty of records when exporting a version for sharing, so online family trees may have lost that vital qualifying information.

Creating

The most common products created by interviewees were scholarly publications (such as journal articles, theses and monographs), community histories and family trees. Some specialist outputs are discussed in the section on historians' use of place, space and geospatial tools. Several interviewees used 'office suite' software like *Word*, *Excel*, *Access* and *FileMaker Pro* to create tables that met their specialised data recording requirements. This structured data might be used in analysis, visualised on a map, or published as a diagram or table. The structured format also helps them record information consistently. Creating structured data formats can be an iterative process as it requires a good understanding of both the source material and the intended uses of the data. Anne added some fields retrospectively, and Kathryn spoke from experience when she said 'you've got to understand the archive well enough to know what data you want to collect'. Barbara stores the 20,000 names related to her research in *Ancestry's Family Tree Maker*, partly because *Ancestry* is a key source for basic records, and partly because she can specify

¹⁰¹ Not to be confused with their *Shoebox* mobile application for photographing documents.

parameters for reports (for example, occupations), then import the report into a spreadsheet for further interrogation.

Most interviewees seem to begin with the software they have to hand, only using more specialist software when their needs grew beyond simple documents. Family historians have a range of software packages designed to meet their needs; some interviewees also used this software to record other life histories.¹⁰² Graduate students and early career academics seem more likely to take up new software, perhaps because they have slightly more time than other faculty historians to explore their options, or because they do not already have significant amounts of material in legacy formats.¹⁰³

The amateur/professional divide shows itself most clearly when it comes to publications, the main output for professional historians but relatively uncommon for family historians. Two non-faculty interviewees have written historical books for general readers, and publishing scholarly monographs and peer-reviewed journal articles is crucial for career progression within academia.¹⁰⁴ Many of the family historians I interviewed were interested in the context of their ancestors' lives but did not feel the need to synthesise their research for scholarly publication. Family historians might have eventual publication in mind, and some might be, as Martin put it, 'hammered by my extended family to publish', but they generally feel that recording the information they have discovered is a

¹⁰² For example, Peter uses *Legacy* to manage data before copying it over to his collaborative project; he appreciates that it can deal with 'convoluted' histories he encounters.

¹⁰³ One example dates back to the 1970s when a now-senior historian used the process of turning their thesis into a book as 'an opportunity to learn how to use a database'.

¹⁰⁴ However, one academic at a private university pointed out that staff at non-publicly funded universities do not face the same pressures to publish.

more important task.¹⁰⁵ For those family and local historians who did want to write about their research, scholarly publication is rarely a goal. Claire's research is informed by her academic experience but she is resisting pressure to write a book: 'I hate writing it all up. I just like the thrill of the chase'.¹⁰⁶ She does however write histories of places her ancestors lived.¹⁰⁷ Charles calls himself 'a factual man' and prefers collecting and publishing data to 'speculating' about it. He will, however, write articles for his local society's newsletter describing his work. While even the most supportive faculty historian might criticise local historians for failing to turn description into analysis,¹⁰⁸ for non-faculty historians, the prospect of publication is complicated because it changes the context of the original research from pastime to something that will be judged on different terms from those originally conceived. For example, Doris is compiling a memoir of a great-aunt, largely for other family members who she feels will want something interesting, accurate, entertaining and well-presented. While she has been collecting material to make the memoir suitably descriptive, she is also aware of another audience who might want something more from her account, which would require a 'shift' in her approach. Fear of being asked to engage with theoretical questions can also intimidate non-academic

¹⁰⁵ The fact that many of the family historian interviewees were retired and felt they had an uncertain but finite number of years in which to work was clearly a factor.

¹⁰⁶ Similarly, a 2005 study on genealogists by information scientist Crystal Fulton suggests that they gained 'a sense of accomplishment and discovery' from locating information about a particular ancestor. Crystal Fulton, 'Finding Pleasure in Information Seeking: Leisure and Amateur Genealogists Exploring Their Irish Ancestry', *Proceedings of the American Society for Information Science and Technology* 42, no. 1 (2005), doi:10.1002/meet.14504201228.

¹⁰⁷ Yakel found that 'narrators of the family history' was a role some consciously assumed, alongside 'archivist' and 'navigator'; these roles were also present in the interview data to various degrees. Yakel, 'Seeking Information, Seeking Connections, Seeking Meaning'.

¹⁰⁸ Michael Drake, 'From Old Bills to Sick Pigs: Our Ways to Capture Community', *Family & Community History* 1, no. 1 (1998): 7-26, <http://www.maneyonline.com/doi/abs/10.1179/fch.1998.1.1.002>.

historians who have not kept up with trends in historiography. As Doris said, 'if people were to come at me with saying, "well how would you deconstruct this?" I wouldn't be able to start, I couldn't deal with that kind of intervention'.

Sharing

Key findings from interview questions asking how family, local and faculty historians share data include the processes for negotiating sharing with unknown correspondents. Also presented are a range of methods for sharing resources, the positive and negative effects of sharing, and the importance of credit and attribution.

Family historians generally intend from the start to share the results of their work with their immediate families, and in some cases with distant relatives and other interested parties. Larry, who has been doing family history research on-and-off for fifty years, found that the ability to publish material online has greatly aided 'being able to search and find other people who were doing similar stuff'. Several participants also reported selectively sharing information online as 'cousin bait', a term for posting names online so that they can be found by others ('cousins') researching the same people.¹⁰⁹ For example, Ed posts 'skeletal' trees on *Ancestry* to enable contact with other researchers. Protecting the privacy of other family members was important for the interviewees who shared family trees online; living people are usually excluded, and a few interviewees are generally wary about putting information online at all. Conversely, Larry has posted a relative's war diary and other community and family history material online.

The experience of family historians shows that making content discoverable online leads

¹⁰⁹ The term also applies offline - Nell has found people through notices in family history magazines. 'Cousin Bait', *WikiTree*, accessed 23 April 2014, http://www.wikitree.com/wiki/Cousin_bait.

to connections. Being discoverable via name searches also occasionally leads to offers of information. Peter has been contacted by other researchers offering him information or further pointers for his *F&S* research, and Charles has been sent images of archival documents after meeting a family historian with ancestors from his village. Faculty historians like Kathryn and Anne had not previously considered that family or local historians might find their records valuable, but the experience of other historians shows that name-rich sources are particularly likely to be found and used after searches online.

The various cooperative projects discussed,¹¹⁰ mostly by community historians, engage in what information scientist Sanna Talja called 'strategic sharing', where information sharing is a 'conscious strategy [for] maximizing efficiency in a research group'.¹¹¹ For example, John credits the success of a current community history project to a shared *Access* database that is 'the core of the exercise'.

Faculty historians tend to be wary about sharing ideas and work in progress in circumstances that might jeopardise their ability to publish on their research or get credit for their work (although sharing unpublished work is necessary during peer review, and early versions of potential articles may be tested out at conferences). Some early career faculty historians blog about their research or post papers on *academia.edu*. Emily and Helen both post about research that is relevant to their expertise but not directly related to their academic scholarship. Daniel reports that other academics will share data, 'if asked, on condition that unpublished work is not cited'. He himself would share research findings rather than data, as he does not think sharing data 'would help me put my

¹¹⁰ For example, projects by FACHRS, *Founders & Survivors*, and local history societies.

¹¹¹ Sanna Talja, 'Information Sharing in Academic Communities: Types and Levels of Collaboration in Information Seeking and Use', *New Review of Information Behavior Research* 3 (2002).

thinking across'. Kathryn believes that 'the publication of data is important too in allowing others to verify your results, which generally speaking historians haven't done', but like other faculty historians, she would not herself share data until 'the prospect of formal publication was within sight'. Helen reports 'strictures about what you can publish yourself' when writing articles for peer-reviewed journals that further inhibit the sharing of data. The faculty historians interviewed sometimes pass on tips about documents to colleagues who do not share the same research interest or subject; Talja described this 'social sharing' as a part of building and maintaining social relationships.¹¹²

How is material shared?

When sharing his research, Ed adapts to his audience, printing and visiting with or mailing material for 'people that aren't computer literate', emailing others or showing documents on his laptop. Some interviewees were careful in how they showed their data to others. Anne has shown people her maps on her laptop and shown a screenshot at a conference, but 'I haven't given anyone the link, and I won't until it's finished'.

While some researchers put information online for anyone to find, others only share in response to specific requests for information. Family history websites have an established tradition of 'look-ups' (also mentioned in *Chapter 1*), which may inform current behaviours. Bob used to offer look-ups on *GENUKI* and will still respond to specific questions about particular individuals, usually 'privately, via email'. Similarly, Claire does not publish information online but will respond 'if I see somebody out there saying "does anybody know anything about this"'; Nell will share 'if they tell me who they are'. This willingness to share in response to specific requests may be because it is less an act of sharing material and more a matter of helping another researcher. It may also be because

¹¹² Talja, 'Information Sharing in Academic Communities'.

some interviewees prefer sharing with known individuals, a process I have labelled 'directed sharing'. Scholars working on specific topics or archives will also send requests for examples of specific things they are working on within small, closely bound networks.

Reported barriers to sharing data include the use of copyrighted material and issues related to the sustainability of digital resources. One senior faculty historian is concerned with digital preservation on two levels: a student has been unable to use the Cambridge Group for the History of Population and Social Structure's family reconstitution software as 'nobody can be found who can put it onto a platform'; the Arts and Humanities Data Service (AHDS) has not been funded since 2008 and the future of the UK Data Archive is considered uncertain. Anne and Charles' ability to share their work on maps is restricted by the copyright of the underlying map image.¹¹³ Kathryn points out that not all data can be shared because some archives place restrictions on the re-use of their data. Doris would like to let relevant libraries and museums know about her final written research, but isn't sure how to do it: 'do you share just from your own computer, or how do you make yourself known?'.¹¹⁴

One section of the interview asked 'are you part of any collaborative research projects?', and if so, how data, particularly work in progress, was shared in that collaboration. Some family historians pool resources and discuss their work with others researching the same

¹¹³ Anne's data is overlaid on *Google Maps*, and Charles' on an image of a map from an archive.

¹¹⁴ *GENUKI* suggest donating copies of family histories to 'relevant local and national genealogical libraries and societies, and to the LDS Family History Library, so that others may subsequently benefit from your work' but this recommendation may not be widely known. Brian Randell and Louis R. Mills, 'Getting Started in Genealogy and Family History', *GENUKI*, December 2013, <http://www.genuki.org.uk/gs/>.

families, but the main forms of collaboration (or to be more precise, cooperation)¹¹⁵ represented in the data were local and community history projects. Interviewees included participants in the crowdsourcing project *Founders and Survivors*, grassroots local history projects, and FACHRS projects. As the data shared through these projects is collected specifically for these projects, the interviewees' relationship to it and the implications of sharing it are different. It has neither the personal resonance of family history data, nor the career-building role of gathering data for publication by faculty. However, Charles' comment that his village society seeks income from selling their newsletter is a useful reminder that there are many reasons for not sharing material freely.

Positive results from sharing

Many family and local historians I interviewed have benefitted from pooling their resources with others researching the same individuals. For family historians like Bob, Claire, Nell and Ed, the connections (including meeting previously unknown living relatives) made through their research are rewarding in themselves. Larry has conducted joint research with family historians all over the world descended from the same two family members. Martin, Ed and Nell gained access to photos held by another branch of a family, and a researcher with whom Claire cooperated wrote up their research for the benefit of all contributors. Nell has swapped certificates with other researchers, but also feels that she benefitted when more experienced historians confirmed her own research.

As mentioned earlier in this chapter, some interviewees have benefited from 'serendipitous

¹¹⁵ As discussed in the *Introduction*, according to Denning and Yaholkovsky's heuristic in which collaboration requires the 'support and agreement of others before you can take action', very few projects are strictly collaborative. Peter J. Denning and Peter Yaholkovsky, 'Getting to "We": Solidarity, Not Software, Generates Collaboration', *Communications of the ACM* 51, no. 4 (2008): 19–24, <http://dl.acm.org/citation.cfm?id=1330316>.

altruism' when others share pointers to resources and articles known to be relevant to their research with them; they, in turn, have done the same for others.¹¹⁶ Otherwise, the benefits of sharing seem more hypothetical for the faculty historians interviewed, although Helen reports increased awareness of her work after publishing some information online. Anne hopes that after publishing her data she might hear from other researchers with more specific information, and Helen thinks that receiving changes or corrections to her data would be 'the most positive reason for sharing'.

Negative results from sharing

Interviewees reported both their own and second-hand accounts of negative experiences with sharing data. Bob and Larry have had bad experiences with their family history material being 'pinched' or reappearing 'without acknowledgment, and almost owned by others'. Helen has 'had some work reproduced in other contexts without due referencing to its source and that has put me off a bit'.

Faculty historians (and those who want faculty jobs) are wise to think carefully about how they share resources before publication. Information scientists Jenny Fry and Sanna Talja found that in fields with fewer collaborative projects and coordinated research efforts, such as history, there are 'lower levels of interpersonal recognition' which means that 'unique contributions to knowledge are less well protected outside the formal publication system'.¹¹⁷ This is probably particularly the case for less-established scholars. Faculty historians also have more at stake in terms of reputational risk. One said 'we are wary of others trawling through our research looking for errors or inconsistencies. [...] if we have

¹¹⁶ For example, Alison has a file called 'References for Other People' for material she will share.

¹¹⁷ Jenny Fry and Sanna Talja, 'The Intellectual and Social Organization of Academic Fields and the Shaping of Digital Resources', *Journal of Information Science* 33, no. 2 (February 2007): 115–133, doi:10.1177/0165551506068153.

made mistakes we don't want to have them used against us'. Faculty historians may also give an advantage to a scholarly rival if they share the resources they have gathered. While Benkler's work on peer production defines information as 'nonrival', in that 'its consumption by one person does not diminish its availability for use by any other person',¹¹⁸ I argue that information about archival discoveries may be considered 'rivalrous' for historians.

Some family historians reported that a reluctance to have their family collected by 'name collectors'¹¹⁹ deterred them from sharing. Ed and Peter explained that some name collectors use unverified information to get the biggest tree possible, to find a link to royalty, or to go back as many centuries as possible. Bob reported that name collectors had made 'incorrect corrections' on sites like *Ancestry* in order to link names to their families. Others had seen inexperienced family historians mistakenly link trees by misidentifying a historical person; these errors can propagate rapidly as family trees are copied. Claire and Peter additionally judged name collectors for having 'a whole tree of names with nothing else'. Complaints about name collectors fell into three categories - 'stealing' trees, making mistakes or publishing claims without evidence, and collecting 'names and dates' without researching the lives of those individuals.¹²⁰

¹¹⁸ Benkler, 'Coase's Penguin'.

¹¹⁹ Claire calls them 'name gatherers' and describes them as 'wishful thinkers' while Larry calls them 'trophy hunters'.

¹²⁰ I did not find signs of the conflict reported by Willever-Farr and Forte, perhaps because the interviewees were not engaged in 'memorialisation' or because my recruitment process excluded those with a more emotional approach. When interviewees mentioned contacting people about incorrect information in family trees, their corrections were either welcomed or ignored. Heather L. Willever-Farr and Andrea Forte, 'Family Matters: Control and Conflict in Online Family History Production', in *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work and Social Computing* (CSCW 2014, Baltimore, MD, USA.: ACM, 2014), <http://dl.acm.org/citation.cfm?id=2531737>.

Conditional sharing

Several factors affected whether interviewees shared, whom they shared with and how they shared resources. Yakel and Torres describe 'giving back' as an 'underlying ethos' that reinforces many activities in the genealogical community',¹²¹ and the interview data shows reciprocation and altruistic motivations in action. Ed shares family research information knowing that other people would want it 'as much as you'd want information that they've got'. Peter appreciates that he has benefited from 'information off a database or a microfiche that someone's compiled in the past' and regards his contributions to *Founders & Survivors* as a way of repaying that. Larry is motivated by altruism to the extent that he has donated diaries and photos passed through his family to appropriate archives and libraries, believing that holding onto them would be 'unreasonable' in the face of 'the greater benefit'.

The importance of trust emerged clearly from the interview data. Faculty and local historians seemed more likely to share with people they met through existing networks, whether scholarly events or community history groups. Family historians were the most likely group to post information about named historical individuals online, and consequently more likely to be contacted by a stranger online. However, a range of interviewees reported making decisions about whether or not to share resources in correspondence. Faculty historians default to refusing requests to share unpublished resources with unknown correspondents. Determining whom to trust is particularly important for family historians. They could potentially benefit hugely if they are able to

¹²¹ Elizabeth Yakel and Deborah A. Torres, 'Genealogists as a "Community of Records"' 70, no. 1 (2007): 93–113. This can be seen in forum posts and social media comments such as Guest Blogger, 'Indexing and Perseverance', *FamilySearch Blog*, 5 November 2013, <https://familysearch.org/blog/en/indexing-perseverance-gp/>. and Anderson, 'The Key to Unlocking Their Stories'.

combine their research with another's, but they also fear the consequences of sharing with a 'bad' historian. Gina and Martin are very careful about sharing their data as both have been told family secrets in the course of their research. The interviewees developed an iterative process for determining whether they could trust someone who had contacted them online. As one family historian said, 'When I determine that these people are for real we then do a little bit of a swap and see how that bit goes and then maybe send a little bit more'.¹²² Two different stages are described in this statement: the initial approach, and a test to see whether information provided is reciprocated. The approach made when requesting data can affect the outcome. For Larry, sharing data is part of a relationship: 'You just don't come to me and say, hey, I need this, goodbye'. Peter is more likely to help those 'who are genuinely interested when they make the personal approach'. Requests from those who fail to observe these niceties may be ignored. In the second stage, the other researcher is evaluated through the quality of the information they provide, their motivations for research, and the tone of their correspondence. Larry calls the process of establishing a correspondent's trustworthiness over several encounters or reciprocal exchanges of data an 'informational onion', a phrase that encapsulates the successive revealing of increasingly precious or personal information.

The point at which sharing might occur has an impact. Anne appreciated working in private for the ability to 'make mistakes and not worry that it's not ready'. Some interview data suggested that some historians might consider sharing 'depleted' collections of resources from which they had wrung every possible publication. However, without strong motivations to do, it is unlikely that time would be found to prepare the data for sharing, and therefore it is unlikely that data in personal research collections will be shared unless

¹²² Discovering that the other is interested in a different individual with the same name is another reason for correspondence ending.

appropriate rewards for the time required are devised.¹²³

Sharing also changes the scope and size of a project, not only in the work required to prepare the data before sharing, but also by potentially starting conversations with other researchers and making the researcher aware of the scrutiny of others. Doris said it was not that 'I want to keep the material to myself', but that 'each time you push out a bit, you can get back a lot of material and then you have to be able to cope with that and to know how to handle it', which further 'extends the project'. Logistical issues aside, she also wanted to have her interpretation and presentation of the material completed before others were able to interpret it: 'I have to know what my own understanding of it is before I can take up other people's ideas'. The possibility of engaging with more academic historians was also intimidating for someone who did not feel up-to-date with the latest theoretical trends. Sharing material puts it into new, unknown contexts, and managing any resulting interactions or comments takes time away from research.

Nell is happy to share 'with someone who's really doing it for their own families' but will not respond to anyone she thinks is trying to make money out of research.¹²⁴ Bob mildly objects to the prospect of someone using his data simply to avoid paying for their own. Larry will not share with people who are 'diamond mining', 'just out to get another bit of

¹²³ For example, a report on institutional repositories ('infrastructures through which universities and colleges seek to safeguard and share digital content created by faculty and staff') found that only 10-15% of American faculty had deposited material in one. David Seaman, 'Discovering the Information Needs of Humanists When Planning an Institutional Repository', *D-Lib Magazine* 17, no. 3/4 (March 2011), doi:10.1045/march2011-seaman.

¹²⁴ This echoes Fulton's finding that money-generating activities may decrease openness in communities. Crystal Fulton, 'Quid pro Quo: Information Sharing in Leisure Activities', *Library Trends* 57, no. 4 (2009): 753-768, <http://muse.jhu.edu/journals/lib/summary/v057/57.4.fulton01.html>.

information that they want to plug in and move on'. Ed regards his family history research as 'an interesting thing that I do in spare time' so he is not 'precious' about other people using his information, but his feelings are different if someone is writing articles or publishing information without attributing his work.

The importance of credit and authorship

When asked how much credit / attribution / authorship mattered to them, the responses show that everyone appreciates being credited for their work, and most would resent someone else taking credit for their work or using it without acknowledging their work. However, the importance of credit is a key difference between faculty and non-faculty historians. Family historian Bob says that it is rewarding when someone recognises the work he has put into research, but accepts that there is always a risk that 'people will pinch your work if it's of value to them'. Nell describes being credited as 'a good feeling', and Peter said it was not high on his priorities but he felt 'a little bit of a buzz' when he was credited in an academic paper for his work compiling information.¹²⁵ Larry said he finds it 'disappointing' when people copy his work without acknowledgement, and Martin might ask the person about it 'but I don't think I'd be too stroppy about it'. Conversely, early career scholars like Helen, Daniel and Anne describe credit and attribution as 'essential', and credit means 'a great deal' to mid-career academic Emily. One senior academic has forthright opinions on sharing data: academics are not free to publish on sites where they cannot be certain they will be credited because 'the attribution of work to authors is now a significant part of demonstrating that you are fulfilling the research part of your contract'. As a retired academic, John can perhaps afford to be more sanguine, saying 'I suppose if somebody lifted something directly and didn't mention it, it would piss me off a bit

¹²⁵ The approval of academically-trained historians is clearly valued by some, perhaps as a form of validation for their own work. However, one community historian was warned about talking about his work to a specific faculty historian who might 'snaffle' it.

[...but...] I don't suppose it would worry me for more than an hour or two'.

There was also a relationship between the importance of being credited and the amount of work required to gather or create a resource. Larry and Peter mentioned the amount of effort they have put in over the years and Doris described 'the slog' of transcribing. Ed is 'not too worried' about people using basic information from his family trees, but would be concerned if the biographical information he'd written was re-used without credit. Claire and Peter seem to mind more when someone 'snitches' information that took more effort, skill or ingenuity to find. Anne is concerned about the lack of 'academic mechanisms' for citing the work that has gone into her database were she to share it online, particularly given that it represents a huge accumulation of labour.¹²⁶ These attitudes might also reflect the difference between attitudes to sharing 'found' objects (e.g. physical or digitised primary sources) and 'created' objects (e.g. concepts, arguments, augmentations and annotations based on found objects).¹²⁷ These interviews suggest a link between attitudes to sharing information with the amount of personal effort, skill or knowledge required to create the item.

Discussion of historians' sharing

Most of the sharing behaviours discussed were targeted at known individuals or localised within small groups. I found that historians who will consider sharing resources will often do so progressively, tailoring how much they share subsequently to the response of the

¹²⁶ It takes 'anywhere from 5 minutes to half an hour to find a single address', a process repeated up to 600 times.

¹²⁷ The differentiation between 'found' and 'created' objects was inspired by Bernadou et al.'s description of 'physical' and 'conceptual' objects. Their physical objects are those discovered or gathered during research, while conceptual objects include 'created, represented and illustrated, and logical propositions formulated, supported, countered, proved, disproved or refuted'. Benardou et al., 'A Conceptual Model for Scholarly Research Activity'.

other party. Human-computer interaction researchers Preece and Shneiderman have described 'generalised reciprocity' as 'a process in which an individual gives back to the community, rather than directly to the person from whom the contribution was received'.¹²⁸ Building on their definition and on my data, I suggest 'generalised sharing' to describe situations in which data or information is shared with a broad community or unknown audiences. The converse - sharing with known individuals - as described by many interviewees, I have called 'directed sharing'.

One of my research questions was whether researchers are more likely to share information or data when their personal research collections contain more digital material. The interview data suggests several barriers to sharing the contents of personal record collections still exist. For example, faculty historians Kathryn and Anne have created bespoke databases of addresses for historical people tailored to their specific datasets and research questions; both mentioned the time that would be required to prepare this data for publication by documenting it appropriately (explaining column headers or database field names, expanding abbreviations, etc.).¹²⁹ As Kathryn said, 'one would need to explain what the categories one had used were and so on'; outlining the scope and exclusions of the dataset might also require explaining the research question and other context around the creation of the dataset. The time required to prepare data for sharing - ensuring it has been reviewed for completeness, consistency and record quality - is also a barrier to sharing for non-academic historians. While *FamilySearch* developed the GEDCOM format used by family tree software for exchanging genealogical data,¹³⁰ other

¹²⁸ Preece and Shneiderman, 'The Reader-to-Leader Framework'.

¹²⁹ Anne points out that work put forward for (peer-reviewed) publication has been 'checked and verified' for accuracy, and that any 'contribution to the world's knowledge' should be similarly verified before it is shared.

¹³⁰ GEDCOM is a file format, based on open specifications, for exchanging genealogical data.

researchers use bespoke recording systems, or rely on the hierarchy of files and folders on their computer to convey meaning. Any platform designed to host personal research collections would need to support a range of source formats and software.

Providing tools that support bespoke data structures may encourage those who have designed their own databases or spreadsheet formats around their specific sources and questions to share their data. Researchers need to be able to convey the context in which a dataset was created, whether based on a particular archive, period, place, topic or research question, and explain any gaps in the dataset etc. Family trees are shared within a relatively long-established genre of online resources that includes both personal websites and specialist genealogy sites.¹³¹ Situating their research within the genre of family trees triggers a range of contextual associations that helps others understand how to assess the information published, including a sense of the motivations and goals of researchers, the shared issues they face (such as name disambiguation and the variability in source records by time and place), the software and archives typically used, and the types of information likely to be present or absent in the dataset. However, there are no such genres for other forms of personal research collections,¹³² particularly those that could be shared by faculty historians. This may be one cultural barrier to sharing personal research collections. For example, Anne would prefer to share data on a specialist site where the context of the original archival sources and her research questions would be 'made really clear and really explicit' rather than something general like *Wikipedia*.¹³³ She also expressed concern about other people making 'the mistake of thinking it's more accurate than it is'. I suggest that

GEDCOM X, 'GEDCOM X', *GEDCOM X*, accessed 8 June 2015, <http://www.gedcomx.org/>.

¹³¹ One-place studies and national associations of family and local history societies provide a range of examples.

¹³² The nearest analogy is probably the deposition of historians' papers in archives.

¹³³ This is not generally an option as topics on *Wikipedia* must pass a 'notability' test.

the lack of conventions for citing edited and enhanced collections like Kathryn and Anne's databases is also a barrier that could be addressed by extending existing genres such as scholarly editions to include newer formats. Finally, it was important for interviewees that their view of the reliability of their research is conveyed with the resource, so that it is not taken as being more authoritative or verified than it actually was. Further research is needed to understand whether sharing personal research collections within a genre designed to accommodate these factors might address some of these issues. However, the time required to articulate the tacit knowledge behind the gathering of resources, and to check the accuracy of data, makes it unlikely that faculty historians would prepare resources for sharing unless it matched traditional forms of publication in terms of career progress and recognition.¹³⁴

Digitality in practice: historians using place, maps and geospatial tools

Here I focus on interviewees' responses to questions about historians' use of place, maps, mapping and geospatial technologies, as they provide a focused example of the challenges and opportunities in their use of digital technologies. This helps us understand the impact of digitality on faculty and community historians, and provides context for discussion of the impact of digital tools, resources and methods in the concluding chapter. Topics covered in my interviews included interviewees' goals in using place-based data or mapping, their use of historical maps or place names, the formats in which they store data,

¹³⁴ The inaugural issue of *Notes and Queries* made a case for the benefit of reviewing personal research collections, concluding with their 'humble conviction that we are doing a service to writers and readers, by calling forth materials which they themselves thought worth notice, but which, for want of elaboration, and the "little leisure" that has not yet come, are lying, and may lie for ever, unnoticed by others'. 'Notes and Queries', *Notes and Queries* 1, no. 1 (3 November 1849), <http://www.bodley.ox.ac.uk/cgi-bin/ilej/image1.pl?item=page&seq=1&size=1&id=nq.1849.11.3.1.1.x.1>. p. 3.

and the geospatial skills and 'things' (a deliberately general term) to which they would like access.¹³⁵

Eighteen interviewees were based in the UK, three were based elsewhere in Europe, and eight were based in Australia or America. Most 'remote' researchers had visited, and in a few cases had lived in the UK at some point. The local historians interviewed appear to favour research about their own location or region, while the number of places a family historian has researched depends on the movements of their ancestors. In many cases, interviewees undertook local or family history research in addition to their main interests. The interviewees' research questions or projects that involved place or geospatial analysis included: mapping historical addresses for members of particular groups of interest in order to look for patterns correlated with location; changing land uses in specific neighbourhoods or towns;¹³⁶ linking pre-enclosure maps and documents;¹³⁷ and mapping routes for different types of transport between early 20th century villages.

Interviewees mentioned a range of specific uses for place and geospatial sources and tools in both physical and digital formats. The historical sources most commonly mentioned are maps,¹³⁸ street and trade directories,¹³⁹ census records, and historical descriptions of places. Other sources include the Pevsner Architectural Guides 'Buildings of England' series and Victoria County History (accessed via British History Online). Digital tools ranged from 'office' software to extremely specialised tools. Structured data like addresses may be

¹³⁵ The interview questions are provided in *Appendix C*.

¹³⁶ For example, to help answer specific research questions, as context for understanding other sources and information, or to trace the development of a region over time

¹³⁷ That is, maps and other documentation created as common land was enclosed.

¹³⁸ Maps ranging from large-scale maps showing farm names to old A-Zs are used or collected because they contain specific or contextual information relevant to research undertaken.

¹³⁹ For example, the *Historical Directories of England & Wales* hosted by the University of Leicester.

recorded in *Word* or *Excel* tables, or *FileMaker Pro*.¹⁴⁰ One local historian is overlaying names from contemporaneous sources on a digitised version of the historical map using *Photoshop*, which he compares favourably to GIS software for his purposes. One interviewee uses a combination of three websites to plot (geolocate) historical addresses - comparing a large-scale digitised version of a historical map with other maps on *Hypercities* that have been geocoded with modern coordinates and *Google Maps*. Another creates maps for current and historical land-use and 'space syntax network maps'¹⁴¹ in *ArcMap*¹⁴² and *Depthmap*,¹⁴³ while another uses Google's *Maps Engine* and the family history software *PHPGedView*. One amateur prosopographer uses a tool called *Parish Locator* which 'allows you to put in either the full parish name or part of a parish name' and see its location, which helps him rule out similarly named parishes when researching individuals' lives. The *GENUKI* gazetteers are mentioned as providing 'a lot of information about where parishes were and also how boundaries have changed'.¹⁴⁴ Physical methods include drawing lines and points that represent information drawn from other sources onto copies of historical maps. These interviews described various forms of physical and digital map-making used to discover, gather, create and share research data and information.

Digitisation has made place-based research a lot more efficient and effective, but the

¹⁴⁰ However, discussions in the *Civil War Pathways* forum are an important reminder that not everyone has access to *Word* or image processing software, as some participants tried to work with PDF readers or Microsoft *Paint*.

¹⁴¹ Space syntax is 'a theory and set of methods for the socio-spatial analysis of settlements and buildings of all kinds and sizes'. 'Research', *The UCL Space Syntax Laboratory*, accessed 17 April 2015, <http://www.bartlett.ucl.ac.uk/space-syntax/research>.

¹⁴² <http://www.esri.com/>. Last accessed 9 June 2015.

¹⁴³ <http://www.spacesyntax.net/software/ucl-depthmap/>. Last accessed 9 June 2015.

¹⁴⁴ Useful, for example, to work out where records would be kept for a village that was once in Oxfordshire but is now in Berkshire.

interviewees felt that there are still some limits to digital methods and some advantages to traditional methods and resources. One prefers to view large-scale maps at the library or archive rather than online. Paper-based copies of maps are easily annotated with pencils, post-it notes or pinned notes, and these annotations are not limited by the ability of a systems programmer to anticipate the uses of those annotations.

When Duff and Johnson reported on genealogists' research in 2003, knowing the place names and administrative boundaries used in the periods being researched was vital, because historical place names were used to organise or index archival records.¹⁴⁵

However, by 2012 the availability of large online repositories had reduced the importance of internal archival record structures, and removed the requirement to search for (some) records in individual institutions. While location is still important, the interviewees were largely able to use name searches to find records independent of their location. Search engines capable of locating obscure references, secondary reading and conversations with more experienced researchers all seem important in helping novice researchers learn to deal with the complexities of place.

The interviewees also discussed the problems they had encountered when searching for or using place-based materials. Places change over time in many ways, from the redrawing of boundaries to the growth or death of towns and the renaming of houses, streets, regions and entire countries. Some novice researchers might learn this the hard way.¹⁴⁶ One

¹⁴⁵ Duff and Johnson, 'Where Is the List with All the Names? Information-Seeking Behavior of Genealogists'.

¹⁴⁶ For example, one less experienced researcher concluded that a listed location is 'not actually a parish' because the only modern reference he could find for it was a named crossroad in a rural location. However, sometimes traces of places may be discoverable by searching *Google Maps*, even though no modern feature remains. The source of these named features is not provided.

interviewee called orthography 'the biggest problem in consulting mapping data'.¹⁴⁷

Immigration records seem particularly prone to errors, as non-literate people with a range of accents provided information to local officials. Examples provided by interviewees included the location 'Stokenchurch' being written as 'Stockenchurch' (the additional 'c' causing it to seem a completely different string to most search engines); officials recording place names assuming that locals dropped their Hs and adding an H to the front of a place name; and an instance in which Hoxton was mis-transcribed as 'Hogsden'.

Digital tools exist to help solve these problems, but, as with many tools for digital history, people may be unaware of them, or the services they use might not have incorporated them. Historical gazetteers provide lists of previous names for places but do not appear to have been used in all instances where they might have been useful. This may be because existing gazetteers do not cover some places at the level of detail some researchers require, because general search engines like Google and those provided on genealogical sites generally suffice, or because any relevant specialist gazetteers¹⁴⁸ may not be easily discoverable in search engines or from academic libraries.¹⁴⁹ There is still a clear need for historical gazetteers, particularly computationally-friendly gazetteers that could help search engines provide better results. For example, gazetteers can include place names (such as administrative units) that are placed within a hierarchy but do not appear on

¹⁴⁷ Irregular 18th century spelling is a problem for another interviewee to the extent that the same street name will be spelt differently within the same document.

¹⁴⁸ Southall, Mostern, and Berman reported that 'many historical gazetteers are small scale projects developed by and for specialists in particular domains'. Humphrey Southall, Ruth Mostern, and Merrick Lex Berman, 'On Historical Gazetteers', *International Journal of Humanities and Arts Computing* 5, no. 2 (October 2011): 127–145, doi:10.3366/ijhac.2011.0028.

¹⁴⁹ Place name gazetteers such as *GeoNames* do not currently deal well with historical place names. However, services such as the *Digital Exposure of English Place-Names (DEEP)* gazetteer of English place names, *A vision of Britain* and some *EDINA* projects offer alternatives. See also Southall, Mostern, and Berman, 'On Historical Gazetteers'.

maps.¹⁵⁰ Gazetteers can also support coordinate-based spatial searches as well as keyword-based searches.¹⁵¹ Tools that allowed people to note place names of interest could help improve historical gazetteers by increasing the number of dated attestations¹⁵² about the use of obscure place names.¹⁵³ The suppliers of platforms could also make searches more powerful by applying phonetic search technologies such as *Soundex*¹⁵⁴ or *NameX*¹⁵⁵ that are currently available in some systems, or specialist services such as the *JRC Fuzzy Gazetteer*.

In other cases, the digital tools that exist might not be appropriate for historical data.¹⁵⁶

¹⁵⁰ Southall, Mostern, and Berman, 'On Historical Gazetteers'.

¹⁵¹ Despite the well-known issues with historical place names, most historical repositories offer keyword-based rather than spatial search. (In spatial searches, the parameters of the search query are determined by the area displayed in a map. *Google Maps* is the most common example but *OldMapsOnline* has spatial searches alongside keyword searches.) This may be because more historians think in terms of place names rather than spatial coordinates, or because spatialising digitised records requires more resources. In 2005 the National Library of Scotland were surprised by analysis of their web statistics that showed that text-based lists were 'significantly more popular than graphical methods' to access maps. Chris Fleet, "'Locating Trees in the Caledonian Forest': A Critical Assessment of Methods for Presenting Series Mapping over the Web', *E-Perimtron* 1, no. 2 (2006): 99–112. While this may reflect the usability of mapping platform used at the time, it also chimes with my interview data. Future research could usefully investigate historians' preferences for place name or spatial search in order to improve the discoverability of geolocated historical resources.

¹⁵² The need for these suggested by Southall, Mostern, and Berman, 'On Historical Gazetteers'.

¹⁵³ For example, some interviewees had records that used field or farm names, and in my own research I encountered records referring to the former sites of long-vanished pubs.

¹⁵⁴ The U.S. National Archives and Records Administration, 'The Soundex Indexing System'.

¹⁵⁵ 'What Is NameX?'

¹⁵⁶ See also Abby Smith Rumsey, 'Full Report: Scholarly Communication Institute 7 Report: Spatial Technologies and the Humanities' (Scholarly Communication Institute, 2009), <http://uvasci.digress.it/full-draft-report/>. Michael F Goodchild and Donald G Janelle, 'Toward Critical Spatial Thinking in the Social Sciences and Humanities.', *GeoJournal* 75, no. 1 (February 2010): 3–13, doi:10.1007/s10708-010-9340-3. Jessop, 'The Inhibition of Geographical Information in Digital Humanities Scholarship'.

The ability to locate records on consumer-facing services like *Google Maps* is valuable, but commercial, general-use mapping tools are not always suitable for historical data, which is often fuzzy, messy, and has highly variable coverage and precision. For example, placing text or points on maps can suggest a degree of certainty not supported by the data.

Locating historical addresses can be inherently uncertain in instances where street numbers were not yet in use, but most systems expect a location to be placed as a precise dot ('point of interest') on a map; drawing a line to mark a location would at least allow the length of a street to be marked as a possible address. Tools more suited to historical data would allow for orthographic variations (so that, for example, a single point of interest could be linked to all the versions of a personal or place name recorded in historical documentary sources).

In these geospatial examples, the complexities of place names can make discovering material more difficult. However, the ability to search the full text of documents without having to know which institution issued or holds which records has greatly aided the discovery process. The interviewees had found a range of solutions for gathering data but the process of turning data into knowledge by creating maps is less straightforward than it perhaps should be. While excellent examples of geospatial tools exist, the resources or knowledge required to use them can be prohibitive.¹⁵⁷ Factors such as concerns about copyright and the lack of tools that can suitably express the uncertain nature of much historical data further inhibit data sharing. The contrast between what historians would like to do with geospatial tools and what they are currently able to do is perhaps indicative of the current state of digital history.

¹⁵⁷ This may change as knowledge of tools like *Neatline* and *Hypercities* grows.

Conclusion

My analysis of these interviews shows the impact that digital tools, resources and methods have had on the processes of discovering, evaluating, gathering, creating, and sharing information for historical research. Discovering resources is more convenient, faster, and more comprehensive. Digitisation and the ability to conduct name searches through powerful general search engines and site-specific full-text searches has reduced the need to understand different types of sources before locating relevant resources. However, record types are still considered in the evaluation process once records have been found. One or two interviewees linked working with original documents with paying close attention to sources or spending time immersed in reading primary sources, but others felt that digital resources were read just as closely as they had previously read original documents. Working closely with source material - transcribing text, noting subjects and entities - is valuable for getting to know the material, but when those transcriptions and notes are in digital formats, they are newly searchable, more easily backed-up and shared, and they can be organised and re-organised as necessary during the writing process. While the practices of historians are clearly evolving as the tools and sources they use change alongside the cultural environments in which they work, most historians seem unreflexive about those changes. Perhaps, as historian Tim Hitchcock said, historians are currently using digital technologies 'to make our lives easier, while pretending that they do not exist',¹⁵⁸ in line with a broader pattern of not discussing the impact of archival structures on their work.¹⁵⁹ It may also reflect the extent to which the impact of digitality on research

¹⁵⁸ Tim Hitchcock, 'Confronting the Digital: Or How Academic History Writing Lost the Plot', *Cultural and Social History* 10, no. 1 (1 March 2013): 9–23, doi:10.2752/147800413X13515292098070. p. 87.

¹⁵⁹ Tim Hitchcock, 'Digital Searching and the Re-Formulation of Historical Knowledge', in *The Virtual Representation of the Past*, ed. Mark Greengrass and Lorna M. Hughes (Ashgate, 2008), 81–90.

practices has been normalised.

As described by the interviewees, the process of evaluating resources is relatively unchanged by the emergence of digitised and online resources. The interview data shows that scholars apply the same criteria when assessing traditional and digital resources. Online resources are not in themselves regarded as any less authoritative - as one said, 'information is information'. The initial evaluation process resembles Don Norman's 'visceral' response: some resources immediately fall into the 'authoritative' category by virtue of the scholarly credentials of the author or publisher. The visible presence of scholarly apparatus - footnotes, properly formed citations, the appropriate use of credible primary and secondary sources - provides a visual proxy for the creator's credibility and the extent of the research underlying the resource. Providing images of any documents mentioned on a site is vital if they are to be used by scholars, as transcriptions without images are not considered authoritative. Interviewees will look at amateur-created resources online, if only out of curiosity or the hope that they might find information that they can corroborate with credible primary sources. Resources not automatically categorised as authoritative may still be used. This echoes earlier research that found researchers are happy to turn to 'amateur websites when they are the only—or best—source of information about the given topic'.¹⁶⁰ Faculty historians are understandably conservative when evaluating interpretive rather than factual statements, and will tend to disregard material found outside authoritative resources. Resources that include information on who created it, when, and why are more likely to be regarded as credible, as are modern-looking sites with appropriate search and discovery functions. The credentials of individual contributors to a participatory site are less relevant than the evidence they provide for any statements. The broad agreement among interviewees that

¹⁶⁰ Terras, 'Digital Curiosities'.

cooperatively created sources can be evaluated positively, and the summary of the functional and design attributes that contribute to that positive evaluation presented here should be useful for the creators of digital resources.

Analysis of the interview data resulted in my categorisation of 'directed' and 'generalised' sharing, in which directed sharing is sharing with known individuals and generalised sharing is sharing with a broad community or unknown audiences. These behaviours build on pre-digital practices, but have reached a greater scale and wider range of people, as digitised material is easier to discover and share. Directed sharing is a form of controlled sharing, suggesting a fruitful vein for future research into the relationship between controlled sharing, motivations and rewards systems. I also drew the categories of 'found' information objects (pre-existing physical or digitised primary sources) and 'created' information objects (concepts, arguments, augmentations and annotations created from found objects) from the data. Further research could help understand the relationship between the personal effort and skill that results in created objects, and attitudes to the directed or generalised sharing of created and found objects.

As expected, faculty and non-faculty historians reported different attitudes to sharing both data and work-in-progress. Family and local historians may work in areas with a greater sense of the common good, but the interactions between local and family history researchers Charles reported, and the intersection of local and family history interests present in the interview sample, suggest that even greater cooperation and resource sharing between family and local historians would be fruitful. While some faculty historians may present a view of family historians happily sharing everything with all

comers,¹⁶¹ most family historians only share information selectively, using public family trees as 'cousin bait', and evaluating the bona fides of those who request information before deciding whether or not (and what) to share. For faculty historians, resources collected are an asset, the value of which is realised when they publish scholarship that situates it within their field. Despite a general feeling that sharing data would have theoretical benefits, there are currently very few individual rewards, and many perceived risks, for sharing data for faculty historians. As one historian said, 'everyone wants access to everyone else's data but no one wants to give up theirs'.

As others have reported, many family historians begin by looking for family members and move on to researching the places, events and broader historical context that affected their lives.¹⁶² The new discoverability of historical resources via name searches in internet search engines may have significantly lowered the barriers to beginning historical research. Future research into the role that family and local history play in providing an accessible entry point for historical research would deepen our understanding of the impact of digitality on public participation and scholarly practices in historical research.

Finally, the ethos of altruistic information sharing that is evident in some practices of local and family historians echoes the rhetoric of peer production movements in other domains. However, the underlying belief that information is a 'nonrival' good may not hold true for faculty historians. The process of sharing (or preparing to share) is also less time-

¹⁶¹ To quote one, 'Venture on to any genealogical website and you will find "lay" historians sharing countless examples of their hard-won archival victories and findings' with 'easy camaraderie'. W. Caleb McDaniel, 'Open Notebook History', *W. Caleb McDaniel*, 22 May 2013, <http://wcm1.web.rice.edu/open-notebook-history.html>.

¹⁶² Butterworth, 'Information Seeking and Retrieval as a Leisure Activity'.
Duff and Johnson, 'Where Is the List with All the Names?'

consuming for those who intended from the start to create resources that will need to be legible for others, but preparing to share personal research collections is a substantial task. Requirements to deposit data alongside scholarly texts might lead to the emergence of a more formal 'ending' (or tidying up) phase that does not currently exist for faculty historians.

These interviews hinted at the size of the gap between what historians are currently able to do and what they would like to do with digital tools. For example, family history software does not currently meet all the needs of the interviewees. Biographical software, particularly when used for family history, collaborative prosopographical or community history projects would benefit from the ability to record the degree of certainty for potential-but-not-yet-proven relationships or identifications, and to link uncertain information to specific individuals.¹⁶³ The interviews revealed contradictions between the activities or goals designed into software systems, and the goals expressed by historians. Future research could also usefully ask researchers which tools they have not tried for lack of time, information about their benefits, or appropriate instructions to learn how to use them. The skills required to use some tools can be a barrier to their wider use for digital history.

The *Conclusion* that follows draws together information from these interviews about the impact of digital tools, resources and methods on scholarly research practices, and insights from previous chapters on public participation in the process of creating digital resources, in order to consider the impact of digitality on historical research.

¹⁶³ Also suggested by Willever-Farr and Forte, 'Family Matters'.

Conclusion: The impact of digitality on public participation and scholarly practices in history

Digitality describes the state in which 'being digital' has been normalised and networked, and digital technologies have been embedded in our personal and professional lives. This thesis has highlighted different aspects of the impact of digitality on history, from the divergent approaches to participatory digital history discussed in the first chapter, to the role of crowdsourcing in exposing volunteer participants to a range of historical materials, and the ability of remote researchers to instantly view the holdings of a record office after discovering a resource online. Digital technologies have created myriad possibilities for public participation in history, and enhanced each stage of scholarly research, but, for a variety of reasons, their full potential has not yet been realised. This final chapter draws together the threads that link public participation and scholarly practices. It looks at the impact of digital tools, resources and methods on the practices that make different types of history, following the categories used in the previous chapter: discovering, gathering, creating and sharing historical materials and knowledge.

Analysing the impact of digitality means asking more than 'how has technology transformed this activity?'. We must also consider how experiences of digital technologies have influenced our understanding of what is possible in historical research, and consider the factors that limit the impact of those technologies. Innovative new technologies and methods may be quickly integrated into working practices, making it difficult to isolate

¹ Negroponte, *Being Digital*.

the impact of digitality - 'being digital' feels natural when it is hard to recall a pre-digital state. It now seems unremarkable that a historian working in one archive can use a device that fits into their pocket to wirelessly access images of documents held in another archive, but the ability to do so is relatively new and relies on a series of cumulative improvements in technologies and organisational practices. Digital technologies have enhanced many of the tasks involved in historical research, providing what historians Dan Cohen and Roy Rosenzweig call 'quantitative advantages' - the ability to 'do more, reach more people' and access more varied sources.² For some historians, the use of 'everyday' digital tools - the word processors and such like that 'enhance proficiency'³ - is no longer considered worthy of notice.⁴ Yet I would argue that the cumulative, quotidian effect of tools that 'enhance proficiency' should not be undervalued. The ease with which historians transform data from text notes to spreadsheets to maps to publications and presentations is almost taken for granted, but it shows the impact of digitality on enhancing everyday research practices.⁵ The changes introduced through digital technologies and mindsets are not limited to scholarly processes. Franco Moretti theorised that the object of study changes when literary historians start using computers, as new practices create previously unimaginable 'constructed' objects of study at scales simultaneously bigger and smaller than traditional objects like the book.⁶ Not everyone is interested in a shift from individual books to corpus of millions of words, but it may be that historians will see opportunities

² Daniel J. Cohen and Roy Rosenzweig, 'Introduction: Promises and Perils of Digital History', in *Digital History: A Guide to Gathering, Preserving, and Presenting the Past on the Web* (University of Pennsylvania Press, 2005), <http://chnm.gmu.edu/digitalhistory/introduction/>.

³ Sternfeld, 'Archival Theory and Digital Historiography'.

⁴ The motivations behind rhetoric that downplays the impact of everyday digital technologies on historical research processes is a question for another day.

⁵ For example, when I asked my interviewees 'what tools or resources have you stopped using?' only one of them thought to say 'the library card catalogue'.

⁶ Franco Moretti, 'Changes', *Public Books*, 1 March 2014, <http://www.publicbooks.org/briefs/changes>.

for new research questions in these new objects of study.

Furthermore, digitality changes the field of history by lowering the barriers to historical research. More members of the public can help make history at times and places convenient for them, and with more accessible resources and support than ever before. The examples discussed in the first three chapters show how crowdsourcing and citizen history projects create new paths for people to discover and nurture an interest in making history. The availability of long-standing communities of practice in online forums (many *Rootsweb* mailing list archives date back to the early 2000s)⁷ and active local and family history groups mean that novice researchers can find answers and support in formats that suit them. Genealogy software and freely-available websites help guide others through the process of assessing and compiling family history records. While some community historians I interviewed took up family or local history upon retirement, other interviewees have been able to fit their research around their working lives. One interviewee, Gina, described how she used overnight business trips as opportunities to work on her family history research, an opportunity created by the widespread availability of wifi in hotels and family history software that she could install on her laptop.

However, this study has shown that technological changes are only part of the story when it comes to enabling new practices. My interview data has shown that the most experienced faculty and community historians have similar approaches to discovering, evaluating and gathering digital resources. But it has also shown that their different motivations and goals are reflected in their attitudes to sharing and creating resources. While all historians benefit from the resources created through crowdsourcing projects, it

⁷ Much activity has now moved to social media sites like *Facebook*, which is a cause for concern when digital preservation is yet to find ways to reliably capture material from these sites.

seems that faculty historians are less able than community historians to explore new forms of digital creation, hindered (as they are to an extent) by the difficulties of collaborating on interdisciplinary digital projects and by their need for credit and attribution when publishing data or research. Conversely, community historians and participants in crowdsourcing are usually motivated by altruism and the intrinsic rewards of learning, discovery and enjoyment, and thereby encounter fewer barriers to data sharing and collaboration. The success of grassroots community history projects like *FreeBMD* and *GENUKI*, unpolished as they may be, demonstrates the transformative effect that digital technologies have had on the ability to collaboratively create and share resources.

The impact of digitality on history

In order to unpack the impact of digitality on historical research, I consider the effect of digital tools, methods and resources in enhancing or transforming each stage of the research process⁸ - 'discovering', 'gathering', 'creating' and 'sharing', as in the previous chapter. However, it is worth noting that, in some cases, digitality has already blurred the lines between the processes of gathering data, turning it into knowledge, and disseminating the results. Knowledge is simultaneously created, discovered and shared in the communities of practice (whether comprised of crowdsourcing participants or community historians) that form around some historical resources. However, faculty historians' need to protect their ideas until they have been polished for, and protected by, publication, so the boundaries between these stages remain largely intact for them.

Discovering

Search technologies have transformed the task of discovering historical sources. Powerful search engines mean that any phrase or line of text can become a search result, and

⁸ University of Minnesota Libraries, 'A Multi-Dimensional Framework for Academic Support'.

phonetic and fuzzy searches have also improved the ability to discover sources.⁹ The ability to search for records that mention a specific named individual, place or event, and receive results from millions of digitised books, historical documents, personal web pages and institutional repositories has opened up new avenues for historians. However, it is possibly more transformative for historians searching for records about people unlikely to have been regarded as notable during their own lifetimes than it is for historians researching well-documented individuals. More broadly, by making very specific or speculative queries possible, search engines have enabled a promiscuous use of documentary sources that has transformed research questions one at a time.

However, the historians I interviewed, and others such as Tim Hitchcock and Andrew Prescott, have rightly raised concerns about the potential for the study of history to be skewed by concentrations of digital resources in certain periods, regions and formats. Digital resources may have the appearance of completeness, but copyright, the commercial imperatives of family and academic history, and historiographic trends lead to gaps in the digital record.

Gathering

The availability of free (or comparatively inexpensive) historical records through GLAM¹⁰ and community sites means the public can access historical material without having to work around archive opening hours, negotiate entry to archives (some of which require users to be 'bona fide scholars'), or navigate unknown etiquettes. Text transcription allows readers who lack the skills to read manuscript or hand-written documents to make use of

⁹ Semantic search and machine learning technologies will further transform search. For example, computational systems that can automatically identify and describe things depicted in images are being developed. Linn, 'Picture This'.

¹⁰ Galleries, libraries, archives and museums.

these resources. Powerful search engines can find information about specific topics through sites like *Wikipedia*¹¹ or even *YouTube*. Computational techniques can link contextual information directly to historical records to aid understanding.¹² However, access alone does not guarantee good history: historians such as Leslie Madsen-Brooks sound a note of caution about potentially 'uninformed' and 'decontextualised readings' of digitised primary sources.¹³ Similarly, Brett Hirsch points out that access is not enough to make digitised materials such as historical plays comprehensible to modern readers; scholarly annotations and commentary are also necessary.¹⁴

Everyday technologies like bibliographic management software (*EndNote*, *Zotero*, and the like) have greatly enhanced historical research. Digital images and transcribed text are easier to collect than physical documents. Unlike the piles and boxes of paper that some interviewees reported still sit on their shelves, digital documents are instantly searchable. The ability to retain every version of a document, and to keep documents for decades (assuming no catastrophic data loss) has also made research more efficient.

¹¹ While the coverage of some historical topics in *Wikipedia* is still uneven, on-going work aims to address the gaps.

¹² *Old Bailey Online* provides a good example of contextual information, mapping the locations of crimes and defendant's homes over modern London, and providing an 'Associated Records' database which includes 'other manuscript and printed materials relating to specific trials' and other associated records. Hitchcock and Shoemaker, 'Digitising History From Below'.

¹³ Leslie Madsen-Brooks, "'I Nevertheless Am a Historian': Digital Historical Practice and Malpractice around Black Confederate Soldiers (Spring 2012 Version)", in *Writing History in the Digital Age*, ed. Jack Dougherty and Kristen Nawrotzki, 2012, <http://writinghistory.trincoll.edu/crowdsourcing/madsen-brooks-2012-spring/>.

¹⁴ Brett D. Hirsch, 'Digital Renaissance Editions', *Journal for Early Modern Cultural Studies* 13, no. 4 (2013), http://muse.jhu.edu/journals/journal_for_early_modern_cultural_studies/vo13/13.4.hirsch.html.

However, copyright and the commercialisation of archival documents makes some documents less accessible. The historical record online, as viewed from outside 'pay walls', is missing both some records valuable to family historians, 'orphan works' of unknown status, and those primary and secondary sources under copyright restrictions. As one community historian (who, like others I interviewed, had previously had access to academic databases while studying) said, there is no point to digitised material if he cannot access it. Finally, as some interviewees pointed out, it is not even the case that every faculty historian has access to commercial repositories through their institutional libraries, let alone researchers outside academia.

Creating

The University of Minnesota Libraries' definition of 'creating' encompasses various processes for turning data into knowledge, including comparing, annotating, relating, referring, writing, illustrating, contextualising, analysing, enriching and interpreting materials.¹⁵ Digital tools have enhanced many of these tasks, but the underlying process may remain the same. For example, while annotations might be recorded as comments in *Word* documents, attached to a digital image in *iPhoto* or through an online tool, or left as a post-it note in a printed book, the intellectual activity involved has not changed.

However, digital tools can change the context in which those activities take place, which in turn changes the activity. For example, digital tools can represent multiple commentaries on a single image or document through linked annotations. The *Digital Mellini* project makes use of a collaborative workspace to incorporate the views of different researchers on specific aspects of a historical document, rather than attempt to find a single 'correct answer'.¹⁶ This allows several interpretations to co-exist and be discoverable from the same

¹⁵ University of Minnesota Libraries, 'A Multi-Dimensional Framework for Academic Support'.

¹⁶ Murtha Baca, 'Digital Mellini: Project Update and Observations on Translating Historical Texts',

document. However, they report that it has also required additional work to make those annotations comprehensible to other scholars.¹⁷ Digital tools that put these creative tasks in public or semi-public spaces also introduce new challenges, and may therefore change the associated intellectual activity. For example, while some recent books have been reviewed 'in public',¹⁸ it is not clear how many scholars are comfortable publicly critiquing work under their own names.¹⁹ Public comments are also more liable to be read by scholars from different disciplines, potentially leading to misinterpretations or judgements based on assumptions drawn from different genres of writing or research practices.

Digital text and media allows for new forms of analysis, using methods such as data visualisation, topic modelling²⁰ or data mining.²¹ These methods can yield new insights and provoke new research questions, but most are not yet accessible to the ordinary historian. In part, this is because computational tools often require intensive data preparation, and

Getty Research Journal, no. 4 (1 January 2012): 153–60, <http://www.jstor.org/stable/41413138>.

¹⁷ As they explained, 'How many people would you expect to understand a post-it note that you wrote to remind yourself about something? Probably not that many'. Francesca Albrezzi, 'Creating "Getty Scholars" Workspace': Lessons from the Digital Humanities Trenches', *The Getty Iris*, 6 March 2013, <http://blogs.getty.edu/iris/creating-getty-scholars-workspace-lessons-from-the-digital-humanities-trenches/>.

¹⁸ For some discussion of the process of writing in public see Shawn Graham, Ian Milligan, and Scott Weingart, 'Writing The Historian's Macroscopic in Public', *Perspectives on History*, October 2014, <http://www.historians.org/publications-and-directories/perspectives-on-history/october-2014/writing-the-historian%E2%80%99s-macroscopic-in-public>. An earlier book reviewed in public was Jack Dougherty and Kristen Nawrotzki, eds., *Writing History in the Digital Age*, 2012, <http://writinghistory.trincoll.edu/>.

¹⁹ For example those used to anonymous peer review may wish to adapt their tone, and others may spend more time fact-checking before commenting in case their comment leaves their ignorance exposed.

²⁰ A statistical method for collecting 'topics' within a text.

²¹ For example, the work of historians Benjamin Schmidt, Michelle Moravec and Tim Sherratt; a glance at the Digital Humanities conference programme will yield other names and projects.

the rewards for doing so are not yet clear to every historian. The development of tools that do not require command-line access is also probably necessary for these new methods to have a wider impact on the discipline of history. Some projects are addressing these gaps. For example, the *Old Bailey Online* project have provided instructions for using *Voyeur Tools*, an online interface for analysing text, with their data.²² Adding structure to data can enable innovative research activities. For example, Tim Hitchcock and Robert Shoemaker describe the results of creating structured data (through the use of XML tags) in the *Old Bailey Online* corpus as a resource 'which can concurrently be both textually searched and statistically analysed', supporting both quantitative and qualitative research.²³ The markup is flexible enough to support the exploration of questions that were not envisaged when the data was created. The creation of *Old Bailey Online*'s data was a large-scale project, but adding structure to other datasets may become less resource-intensive as new computational techniques become available. As discussed previously, computational techniques such as a machine learning could be combined with crowdsourcing tasks in mutually beneficial ways.²⁴ This may reduce the gap between well-funded projects and others, thereby reducing the likelihood of digital history projects reinscribing the canon.

The emergence of crowdsourcing platforms, such as *Pybossa*, *Scripto*, *FromThePage*, the Open University's *nQuire-it* (which allows anyone to set up 'scientific missions' to collect and share data),²⁵ the *Zooniverse's Panoptes*,²⁶ and the New York Times' *Hive*,²⁷ further

²² Stéfán Sinclair, 'Voyeur Tools & Old Bailey', *Criminal Intent*, 29 May 2011, <http://criminalintent.org/voyeur-tools/>.

²³ Hitchcock and Shoemaker, 'Digitising History From Below'.

²⁴ For example, the Public Catalogue Foundation has an interface in which participants can quickly verify image categories suggested by software. Collings, 'The Art of Computer Image Recognition'.

²⁵ nQuire It, 'About NQuire It', *NQuire It*, accessed 23 February 2015, <http://www.nquire-it.org/#/about>.

²⁶ Lintott, 'Open Zooniverse: Beta Testers Required'.

contributes to public participation by reducing the amount of work required to set up a crowdsourcing project. Positive publicity around crowdsourcing has helped raise public awareness of the mutual benefits of participation in crowdsourcing, making it easier for subsequent projects to attract attention. This combination of publicity and entry-level technical platforms may also lower the barriers for institutions or individuals who wish to experiment with crowdsourcing.

This thesis has discussed a number of crowdsourced transcription and indexing projects that make major contributions to history by digitising and classifying information.

Transcription and indexing projects are also popular with grassroots or community history groups such as FreeUKGenealogy, *Online Parish Clerks*, and local history societies. The public can volunteer to help enhance heritage material through dozens of crowdsourcing sites or community history projects. As discussed previously, undertaking microtasks that help transcribe or describe historical materials also provides opportunities for learning technical skills or developing new interests, potentially leading to further historical research.

Attitudes to sharing the data created through transcription reveals key differences between these groups. Participants in crowdsourcing projects are often motivated by the desire to make an altruistic contribution to other people's research, and to the general good; for them, the act of creating and sharing are intrinsically linked. Family and local historians, like faculty historians, transcribe records for their own purposes. However, while most family and local historians either intend to share their data from the start, or negotiate sharing with others as their research develops, faculty historians tend not to

²⁷ Jacqui Maher, 'Hive: Open-Source Crowdsourcing Framework', *Nytlabs Blog*, 9 December 2014, <http://blog.nytlabs.com/2014/12/09/hive-open-source-crowdsourcing-framework/>.

share their data, focusing on disseminating the results of their research in publications and presentations.

Through other crowdsourcing projects, members of the public can contribute information to help identify historical photographs, or upload their own historical photographs to sites like *Flickr*, *HistoryPin* and various genealogy sites. Niche sites based around specific topics (for example, projects commemorating the centenary of the First World War or collecting memories of the fall of the Berlin Wall) or locations (*My Brighton and Hove's Letter in the Attic*) provide services where the public can upload personal documents. Again, this contributes to digital history resources for the benefit of all. In turn, the serendipitous connections that these projects can provide helps increase the public's interest in making history. However, most projects aim to enhance existing collections. There are as yet no widely used services to help preserve and provide access to 'shoebox archives' of historical material held by families or private collectors.²⁸

Sharing

Sharing includes dissemination activities such as publishing, teaching, conferences, discussion, and sharing data and ideas.²⁹ As discussed in earlier chapters, digital technologies have lowered the barriers to publishing material online. Local history societies and family historians have found ways to share selected forms of their resources and research online, through family trees or community archives. Faculty historians may

²⁸ Another outcome from my research was a proposal for the 'digital history commons', in part to address this. See Mia Ridge, 'Creating a Digital History Commons through Crowdsourcing and Participant Digitisation' (Herrenhausen Conference: '(Digital) Humanities Revisited', Herrenhausen Palace, Hanover, Germany, 2013), <http://www.miaridge.com/herrenhausen/>.

²⁹ University of Minnesota Libraries, 'A Multi-Dimensional Framework for Academic Support'.

also be encouraged by administrative or cultural means³⁰ to publish their research in open access journals or to make articles available under 'open access' licences. Early career researchers are particularly vulnerable to pressures both to share data (to be 'more digital') and to keep their data and research interests private (to protect the intellectual assets that may determine their future career path). Two interviewees mentioned that their funders or university courses expected them to blog about their research. One described how she negotiated the requirement to do so without prejudicing her chances of publishing the ideas behind her project in a 'more academic' outlet. The impact of the open access movement on academic publishing, and on public access to historical research, is also still to be seen. It may yet lead to the creation and cultural enforcement of models for crediting digital data deposits, drawing on the citation practices already established for other forms of scholarly publishing.

Faculty historians may be encouraged to deposit articles, teaching materials or datasets in institutional repositories. However, it appears that institutional repositories are under-populated,³¹ and, as mentioned in the first chapter, community repositories for primary sources like *Zotero Commons* did not appear to be much used.³² While 'commons'

³⁰ For example, the UK Data Service aims to work with 'data owners to identify and remove all unnecessary barriers to access'. UK Data Service, 'Open Access Data'. Funders may also require recipients to publish their research under an open access licence. Research Councils UK, 'RCUK Policy on Open Access', *Research Councils UK*, accessed 31 May 2015, <http://www.rcuk.ac.uk/research/openaccess/policy/>.

³¹ For example, in 2010 a US study found that only 15% of faculty members had deposited data. Reported in Seaman, 'Discovering the Information Needs of Humanists When Planning an Institutional Repository'.

³² One archivist's comment about historians' 'traditional assumption that documents lose value in proportion to how many people have seen or used them' may be pertinent. P. Botticelli, 'Comment on "Archivists and Historians—Am I Giving Archivists Too Much Credit?"', *ArchivesNext*, 4 January 2012, <http://www.archivesnext.com/?p=2434&cpage=1#comment-284643>.

platforms technically enable sharing, the cultural practices around sharing are yet to change, particularly for faculty historians.³³ Control over the extent to which thoughts-in-progress are shared is important for researchers, and I have drawn categories of controlled 'directed' and 'generalised' sharing from my interview data.³⁴ This suggests that online tools that do not take sufficient care to reassure users that their data will remain private are less likely to be used. Attitudes to sharing may particularly be an issue for scholarly workbenches or niche sites with a small, closed community (i.e. where requests for an account may have to be approved) in which not every member of the community is known to each other. The liminal space between directed and generalised sharing this creates may be particularly problematic for faculty historians. The risk of losing scholarly face by posting an incorrect or misinterpreted statement or a mundane question, combined with the risk of having one's best ideas stolen, may be why much scholarly discussion takes place in-person or using more private media such as email. Conversely, forum posts in communities of practice are public but are often directed at specific individuals. It may be that posters forget that their posts are public, that they do not think their posts will attract any particular attention, or that being found to be in error or having ideas 'stolen' does not have the same consequences for those posters. The requirement for generalised sharing and the uncertain outcomes associated with doing so may also be one cause of resistance to projects attempting to implement a scholarly 'commons'.

The difference between behaviours enabled by sharing technologies, and the behaviours that actually occur, are an example of what computer-supported cooperative work (CSCW)

³³ The fact that there are different models for 'scholarly commons' can only add to the confusion. While some 'commons' are designed around sharing common cultural objects, others seem to be social spaces in which like-minded individuals can meet.

³⁴ As defined in *Chapter 4*, directed sharing is sharing with known individuals, and generalised sharing is sharing with a broad community or unknown audiences.

researchers call the 'social-technical gap'³⁵ between the social needs of users and the technical affordances of digital tools. My interviews showed that data sharing is a negotiated, conditional process, and that different conditions apply in different circumstances. For example, one family historian may not want to share information about living relatives, while another may be happy to share information available in public archives but would not share information that was more difficult to track down with anyone except trusted peers. Faculty historians may carefully choose cross-disciplinary colleagues or those working on the same material from a different angle in order not to feel anxiety about discussing their research in progress. Conversely, most tools have a binary model - information is either private or public - for sharing information, which may not meet the needs of researchers. CSCW research shows that implementing more granular sharing models is still a challenge, and that technical solutions for controlling access to information cannot exist outside social systems. Faculty historians may find sharing particularly difficult, as even the act of negotiating sharing can be fraught.

The apparent failure of historical wikis³⁶ may be due to a clash between the potential to write entries related to historical sources collaboratively, and a research culture in which open collaboration is relatively rare. As an example, the *London Lives* wiki was set up for 'writing biographies of eighteenth-century Londoners',³⁷ but the lack of existing models for

³⁵ Ackerman, 'The Intellectual Challenge of CSCW: The Gap Between Social Requirements and Technical Feasibility'.

³⁶ The UK's National Archive, the Science Museum (London), the British Postal Museum & Archive and the Public Record Office of Victoria have experimented with and subsequently closed wikis. Alexandra Eveleigh's doctoral thesis at University College London should provide further information on the National Archives project.

³⁷ London Lives, 'Biographies', *London Lives Wiki*, 2010, <http://www.hrionline.ac.uk/londonliveswiki/tiki-index.php?page=Biographies>.

collaboratively writing life histories might be one reason it was little used.³⁸ The uncertainty of the task is intensified when the potential readers and collaborators are unknown, there is no shared specific goal in writing (as when writing an article with another scholar, editing a book, or even writing for *Wikipedia*), and the topics to be written may relate closely to someone else's scholarly territory or family history. Sites that seek public contributions from scholars may be able to encourage more contributions by relating their request to activities that are already part of the research process. Another factor might be designing tasks so that the contributor does not have to take on the overhead of collaborating with others while making their contribution.³⁹

Faculty historians seem more constrained than community historians in devising new dissemination formats that take full advantage of digital methods. Very few peer-reviewed journals are able to host formats other than articles, inhibiting historians' ability to explore emerging digital formats for presenting research. However, formats such as *Scalar* offer 'media-rich scholarly publishing that's as easy as blogging', and can be given the same stable 'digital object identifiers' as online journal articles. Faculty historians might dream

³⁸ Other reasons suggested by project staff include the need for a separate login, and the lack of facilitation (or scaffolding) for the task, as people do contribute to more structured tasks. Sharon Howard, London Lives staff member interview, interview by Mia Ridge, 15 April 2014. One local historian I interviewed had learnt similar lessons from a collaborative community history project: 'if you want someone to help you it's useful to give them a very clear set of instructions [...] you've got to have a template, make it easy for them'. See also Simon, 'Principles of Participation'.

³⁹ For example, it may be more natural to leave a comment with factual information than to edit another scholar's sentence unasked. It would then fall to the project to integrate the information submitted. Research on *Wikipedia* has shown that a large proportion of the time writing articles is spent coordinating and communicating between users. Aniket Kittur and Robert E. Kraut, 'Harnessing the Wisdom of Crowds in Wikipedia: Quality Through Coordination', in *Computer Supported Cooperative Work (CSCW)* (Computer Supported Cooperative Work (CSCW) 2008, San Diego, CA, USA, 2008), 37-46.

of creating digital projects tailored for the specific requirements of their historical dataset, research question and audience, but their peers may not be confident in their ability to evaluate the results.⁴⁰ Ludmilla Jordanova lists three criteria for judging historical writings ('the use and interpretation of sources; the aptness and effectiveness of conceptual frameworks; and the quality of the writing itself')⁴¹ that could be adapted for assessing digital scholarship, if criteria for judging additional outputs (for example, the aptness and effectiveness of digital platforms) can be agreed. As Kathleen Fitzpatrick points out, 'digital work demands its own medium-specific modes of assessment'.⁴² Bethany Nowviskie notes that digital scholarship is rarely the work of only one individual, and is rarely completely 'done', complicating notions of authorship and appropriate moments for assessment.⁴³ Understanding the particular requirements of digital scholarship may require additional training for faculty historians, but this may also encourage more historians to explore new forms of sharing their research. The difficulties in creating and assessing digital scholarship suggest that the discipline of history has not reached a state of 'post-digitality', in which the choice of specific digital tools or methodologies is considered alongside, or as part of, the scholarly argument.⁴⁴

⁴⁰ My own experience bears this out. Originally I had proposed building a software tool for aggregating spatially-indexed manuscripts, letters and diaries. However, the pace of change and the requirement to produce a traditional thesis combined to make the project unfeasible.

⁴¹ Jordanova, *History in Practice*, 2006. p. 94.

⁴² Kathleen Fitzpatrick, 'Evolving Standards and Practices in Tenure and Promotion Reviews', *Planned Obsolescence*, 11 February 2014, <http://www.plannedobsolescence.net/blog/evolving-standards-and-practices-in-tenure-and-promotion-reviews/>.

⁴³ Bethany Nowviskie, 'Evaluating Collaborative Digital Scholarship (or, Where Credit Is Due)', *Journal of Digital Humanities* 1, no. 1 (2012), <http://journalofdigitalhumanities.org/1-4/evaluating-collaborative-digital-scholarship-by-bethany-nowviskie/>.

⁴⁴ For a view of post-digitality from other fields, see Ross Parry, 'The End of the Beginning: Normativity in the Postdigital Museum', *Advances in Research - Museum* 1, no. 1 (1 July 2013): 24–39, doi:10.3167/armw.2013.010103.

The impact of digitality on collaboration in historical research

In the twentieth century, pre-networked technologies enhanced the work of community history projects that sought to compile indexes to materials such as wills, marriage, birth and baptism records held in local record offices. However, the ability to send electronic versions of documents on floppy disks for central compilation, and to copy and paste those electronic records into a central record did not reduce the need for manual coordination. A fundamental change wrought by networked technology is that scholarly collaboration no longer needs central coordination. Software can track which sections of text have been handed to individuals for tasks, then compare and merge the results of those tasks (or re-assign them for manual validation). The validation systems used by some crowdsourcing platforms can also reduce the reliance on gatekeepers to weed out poor quality or malicious contributions. This means that tasks can be more confidently opened to contributions from the public.⁴⁵ Finally, the relative simplicity of publishing content online has made it easier for individuals or community groups to share the results of their collaborative indexing projects. In the past, local history societies may have sent CD-ROMs of their resources to local libraries, but now they can post them, or at least advertise their existence, online.

Local history and special interest groups have established successful models for collaborative projects that collect, digitise and organise resources (particularly primary sources). Their collaborations may benefit from a shared, altruistic motivation to make historical sources about their area of interest available to all. Community historians also

⁴⁵ To draw on an entirely pre-digital example, Anne Secord's work on 19thC natural history correspondence networks noted the time and effort that had to be 'invested' when corresponding with those of unknown social status and connections. By enabling computational validation of the contributions instead, digital tools have reduced the need to invest time in vetting contributors. Secord, 'Corresponding Interests'.

benefit from relationships of trust, aided by their shared histories or goals, and sometimes from the ability to meet in person.⁴⁶ They are also free from some of the impediments that scholars within academia must overcome, particularly the complicated link between collaboration on digital publications, and the credit and attribution necessary for career progression.⁴⁷ Analysis of successful 'grassroots' community history projects and other participatory history projects yields examples that might provide models for institutional projects. These community history projects tend to have a collaborative initiation phase,⁴⁸ in which stakeholders negotiate mutually agreeable goals related to the historical materials and/or questions, and select or design the necessary recording systems. This is usually followed by a cooperative phase, in which distributed recording systems (whether Access databases or web-based crowdsourcing interfaces) manage the coordination of data inputs. In future, the collaborative phase may require less technical development or specification as more participatory platforms become available. However, as each dataset, research topic or niche interest is slightly different, some initial work will still be required. Writing project updates and explaining how the results are aiding the wider project may require further collaboration or could be delegated to a 'community manager' role. Discussion amongst participants (possibly as a community of practice) may in turn lead to new projects that initiate another form of collaborative action. Understanding this model

⁴⁶ Further research into the advantages that groups able to meet in person, or formed around a strong, shared goal or approach (such as the FACHRS members, many of whom had studied a family and community history course at the Open University), have over purely online, distant groups would contribute to our understanding of these grassroots projects.

⁴⁷ For example, Bethany Nowviskie posited the 'tacit notion of scholarly credit as a zero-sum game, which functions as an underlying inhibitor to generous sharing'. Nowviskie, 'Evaluating Collaborative Digital Scholarship (or, Where Credit Is Due)'.

⁴⁸ Following Denning and Yaholkovsky's formulation in which collaboration 'requires [the] support and agreement of others before you can take action' and cooperation just means following the rules of a project. Denning and Yaholkovsky, 'Getting to "We": Solidarity, Not Software, Generates Collaboration'.

may help academics and GLAMs when planning projects. Institutionally-led projects may have a different initiation phase, and, as potential participants may not have been involved from the start, they may need a dedicated marketing/outreach phase in addition to on-going communications.

This division between collaborative and cooperative phases in grassroots projects can be connected to Caroline Haythornthwaite's work on light and heavyweight models of peer production.⁴⁹ The collaborative phase, like her 'heavyweight peer production', requires 'attention to the actions and contributions of others' in a community who 'themselves determine and enact the goals, purposes and processes' for their project. However, based on my analysis of grassroots community history projects, these 'heavyweight' collaborations may produce lightweight, microtask-based systems to which any member of the 'crowd' can contribute. While institutional projects tend to be better funded, and therefore more polished, the institutional or community origin or organisational structure of a project might not be apparent in the final interface online.

'Nonrivalous' information and other concepts from peer production

The discourse around peer production and open source models for digital history sometimes tends to the utopian (or at least the hopeful, as various proposals for 'open notebook history'⁵⁰ or content commons show). Benkler's influential work on the peer production of information relies on the 'commonplace' idea that information is a 'purely' or 'perfectly nonrival good', in that 'its consumption by one person does not diminish its availability for use by any other person'.⁵¹ However, the reality may be different for faculty, family and local historians. The interaction model created on *Find A Grave*, in which a

⁴⁹ Haythornthwaite, 'Crowds and Communities'.

⁵⁰ McDaniel, 'Open Notebook History'.

⁵¹ Benkler, 'Coase's Penguin'.

record is 'owned' by a single contributor who controls the information added to it, shows that it is possible to create models of competitive, rivalrous peer production from theoretically non-rivalrous information such as the existence of a gravestone. While others are free to use the information posted, or to post their own information elsewhere, they are excluded from directly contributing to that particular record. The discussion on 'sharing' in *Chapter 4* shows that the pressure on faculty historians to publish research may also lead to their regarding information as rivalrous. The close relationship between personal research collections and publications based on those collections (often built up after years of work with historical materials) means that data sharing currently offers more risks than rewards for faculty historians.

Data sharing also has a logistical overhead. While the interviews contained hints that academic historians might be willing to share parts of their personal research collections once they had wrought every last publication from them, this is unlikely to happen before retirement (if then). The discussion of personal research collections indicated the level of effort required to clean up and document a dataset or ad hoc collection of records before depositing it in a public space. The preparation of this information is neither cost- nor risk-free, while the benefits are still nebulous for academic historians. However, the sharing practices of family and local historians, in addition to the comments of those experimenting with collaborations in digital history,⁵² show that data is more readily shareable when it is known from the start that it will be viewed by others.

Collaborative, born-digital productions such as open source software are sometimes held up as models. However, open source software models do not automatically translate to historical research. One motivation for software engineers contributing to open source

⁵² Albrezzi, 'Creating "Getty Scholars" Workspace'".

projects is the potential career benefits that may result from their contributions.⁵³

However, if contributing to 'open source history', whether through editing *Wikipedia* articles or sharing data, might be considered equivalent activities within history, there are currently no (or very few) equivalent career benefits for academic historians.

The gap between the potential and reality of digital history

While digital technologies have enhanced many research practices, there are still barriers that prevent historians from taking better advantage of new tools and methods. As discussed earlier, I asked community and faculty historians what additional geospatial skills, tools, software or data they would like access to. The results showed that, despite large-scale commercial mapping work and some high profile historical mapping projects, there is an unmet need for everyday geospatial tools suitable for historians. Many wished to create maps to help analyse or understand their data. For example, those with datasets containing historical locations would appreciate the ability to map addresses from specific periods on historical maps that are georeferenced, georectified and displayable on a historical map or a modern, copyright-free map.⁵⁴ Some interviewees turned to readily available, 'vernacular'⁵⁵ tools such as *Google Maps* or *Google Earth*. Others would just like access to 'old maps superimposed on current maps' and through them to modern coordinate systems. One who wanted to learn GIS⁵⁶ also expressed a desire for a better user interface than those in GIS tools she has seen. However, and quite apart from any

⁵³ Oreg and Nov, 'Exploring Motivations for Contributing to Open Source Initiatives'.

⁵⁴ Three academic historians also mentioned a desire for others with relevant knowledge to be able to add comments, or suggest corrections or new data points; all three recognised that receiving corrections might also be personally challenging.

⁵⁵ i.e. tools that are 'not designed for expert use exclusively'. Smith Rumsey, 'Full Report: Spatial Technologies and the Humanities'.

⁵⁶ Geographic Information Systems, software for manipulating spatial data.

usability issues, learning to use GIS software also requires a shift in mental models.⁵⁷ For example, GIS organises information in layers placed over a base map, with 'each layer dedicated to a particular type of data',⁵⁸ which may require some adjustment for less structured data (and to mental models for querying data). It may also require access to specialist software such as *ArcGIS*, which may require the purchase of compatible computer equipment.⁵⁹ The complexity of some software packages (or the combination of packages assembled to meet various needs) is a barrier for those short on time, unable to access dedicated support or training, or who do not feel capable of learning the specialist jargon and skills required to assess and procure software to meet their needs.⁶⁰ The need for equipment and software licences can be a financial barrier.⁶¹ Unclear licensing requirements and costs for purchasing high-resolution historical maps are another. One interviewee pointed out that data or maps that are free for academic use might not be re-

⁵⁷ The learning curve for GIS has also been described as 'steep'. Jessop, 'The Inhibition of Geographical Information in Digital Humanities Scholarship'.

⁵⁸ J. B. Owens, 'Toward a Geographically-Integrated, Connected World History: Employing Geographic Information Systems (GIS)', *History Compass* 5, no. 6 (November 2007): 2014–40, doi:10.1111/j.1478-0542.2007.00476.x.

⁵⁹ *ArcGIS* seems to be the software most commonly taught but is only compatible with Windows, requiring Mac users to have machines capable of running 'parallel' software and to pay for additional operating system licences <http://www.esri.com/software/arcgis/arcgis-for-desktop/system-requirements>, last accessed 10 June 2015.

⁶⁰ Similarly, Suri's 2011 study of the use of Geographical Information Systems (GIS) by people researching historical topics found five 'stumbling blocks': the 'lack of technology training, spatial literacy, lack of support at their home institution, the pragmatics and impracticalities of getting involved in a project with strangers, and finding a potential collaborative partner in their own institution'. Venkata Ratnadeep Suri, 'The Assimilation and Use of GIS by Historians: A Sociotechnical Interaction Networks (STIN) Analysis', *International Journal of Humanities and Arts Computing* 5, no. 2 (October 2011): 159–188, doi:10.3366/ijhac.2011.0030.

⁶¹ For example, one interviewee requires two large external hard drives to store his larger GIS maps, and another had to reduce the resolution of a map he was making so it could be manipulated by his home computer.

usable in public-facing sites, forcing her to choose between affordable software and public access for her project.

The interviews also demonstrated how difficult it is for people without technical experience to imagine and describe the tools that might help their research. High profile or widely used digital history projects might help by acting as 'boundary objects'.

Boundary objects are abstract or physical objects that can help translate concepts between domains by providing a common object which people from disciplines, or specialists and generalists, can view and discuss.⁶² These boundary objects might make the process of specifying requirements easier for historians by allowing them to discuss their requirements in relation to existing projects.

While there is a great interest in geospatial tools that would help scholars explore and publish spatially-oriented historical research data on historical maps, there seems to be a lack of suitable tools or platforms, or at least a lack of awareness of suitable options.⁶³ This is partly the result of the specific requirements of faculty historians, which include the need to keep their research data private until they are ready to publish it, and to include appropriate citation information when it is published. The ability to mark a probable region of interest by drawing a polygon or line would alleviate some of the false precision so common when mapping historical data. In 2009, historian Onno Boonstra wrote of the need for a 'GIS data infrastructure', with more 'digitised, vectorised and georeferenced'

⁶² Susan Leigh Star and James R. Griesemer, 'Institutional Ecology, "Translations" and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39', *Social Studies of Science* 19, no. 3 (1 August 1989): 387-420, <http://www.jstor.org/stable/285080>.

⁶³ For example, *Hypercities* or *Google Earth* may meet some requirements, and GIS software should support the display and query of more structured data.

maps and geolocated content.⁶⁴ While much progress has been made in the provision of digital historical maps,⁶⁵ it seems the next step is providing tools for ordinary historians to gather, create and share their research data on historical maps. The lack of easy-to-use tools that can integrate historical maps with different types of data to study change over time is a barrier to historians asking more interesting questions of their sources.

This is symptomatic of the wider state of digital history. This research has discussed the impact of good design on participation rates in crowdsourcing projects, and I would argue that academic projects and tools could well pay more attention to user experience design. The uptake of scholarly platforms may be improved if they planned outreach phases, designed onboarding processes, and created tutorials and instructions focused on helping historians understand their benefits in relation to their own data. Designing tasks in order to reduce cognitive load would free more working memory for other tasks. Finally, showing scholars how their data can be exported from a platform might help alleviate fears about sustainability and investing time and effort into learning a new tool.

Reconsidering 'the historian'

Participatory digital history projects have encouraged, or made more visible, a shift in the characterisation of 'historians' that has been underway for many years. Projects discussed throughout this thesis have contributed to a broadening of interest and participation in historical research. This is also linked to the availability of primary sources online, and the narratives of intellectual discovery and emotional journeys presented in television shows like the BBC's *Who Do You Think You Are?*.⁶⁶ While movements in popular culture have

⁶⁴ Onno Boonstra, 'Barriers between Historical GIS and Historical Scholarship', *International Journal of Humanities & Arts Computing* 3, no. 1–2 (2009): 3–7, doi:10.3366/E1753854810000480.

⁶⁵ Some of it due to private individuals (for example, the *David Rumsey Historical Map Collection*).

⁶⁶ Philip Riden has argued that the growth in local history was linked to the rise in adult education,

helped make the practices of history feel more relevant and perhaps more attainable, the barriers to 'making history' have been significantly lowered by digital technologies, particularly the web. Accordingly, discussion of the intersection of public participation in crowdsourcing, community history and academic research in the creation and use of digital resources may benefit from more precise vocabulary to describe key groups. In *Chapter 3*, I defined citizen history projects as those that require or teach some historical skills beyond the technical skill of transcribing text, but I did not address the question of what we call the people participating in and learning through those projects. At what point does 'a participant' become 'a historian'?

To explore this, I return to *Operation War Diary*,⁶⁷ which mentions 'citizen historians' on

and adult education courses have certainly influenced community history groups like the Family and Community Historical Research Society. Riden, *Local History: A Handbook for Beginners*. p. 10. It is also important to remember that debates about the roles of amateur and academic historians are centuries old. See for example Beckett and Watkins, 'Natural History and Local History in Late Victorian and Edwardian England'. On the more recent impact of television, film and other popular histories see also Hilda Kean, 'People, Historians, and Public History: Demystifying the Process of History Making', *The Public Historian* 32, no. 3 (August 2010): 25–38, doi:10.1525/tph.2010.32.3.25. Jerome de Groot, 'Invitation to Historians', *Rethinking History* 18, no. 4 (2014): 599–612, doi:10.1080/13642529.2014.893662. David Thelan, 'Afterthoughts. David Thelen: A Participatory Historical Culture', in *The Presence of the Past: Popular Uses of History in American Life*, by Roy Rosenzweig and David Thelan, 1998, <http://chnm.gmu.edu/survey/afterdave.html>. Roy Rosenzweig, 'Afterthoughts. Roy Rosenzweig: Everyone a Historian', in *The Presence of the Past: Popular Uses of History in American Life*, by Roy Rosenzweig and David Thelan, 1998, <http://chnm.gmu.edu/survey/afterroy.html>. Twells, 'Community History'. Julie Hill, 'Online Family History Trends Report', *Archives*, 2 February 2011, <http://www.archives.com/blog/miscellaneous/online-family-history-trends-1.html>. Georgia-Lee Hoe, 'Australian History: Online and Insatiable', in *Peer Reviewed Proceedings of the 4th Annual Conference, Popular Culture Association of Australia and New Zealand (PopCAANZ)* (4th Annual Conference, Popular Culture Association of Australia and New Zealand, Brisbane, Australia, 2013), http://popcaanz.com/conferenceproceedings_2013/History_Hoe_AussieHistoryOnline.pdf.

⁶⁷ The *Zooniverse*, Imperial War Museum and National Archives project aimed at indexing unit

its front page and in various publicity materials. As on other sites, participants on the site are clearly learning technical skills, and forum discussions also demonstrate aspects of historical thinking, including posting questions, using evidence, and looking for corroboration between sources. However, as discussed in *Chapter 3*, while it is an excellent crowdsourcing project, its claims to be a 'citizen history' project are somewhat overstated, in part because it has not been able to support participants in learning more advanced historical skills. The project had invited academic historians to take part in its advisory board and participate in discussion, but the way this has been implemented may have unintentionally highlighted a further tension within the project. When viewing the *OWD* forum,⁶⁸ one may notice that some messages have the label 'historian' next to the poster's name. This raises questions about the status of the purported 'citizen historians' who are not so labelled - does it mean they are 'potential historians', or even 'wannabe historians'? Or, for the moment, are they more accurately called 'transcribers'?⁶⁹ The existence of these 'real' historians also make it unclear at whom forum posts with appeals for further information from 'historians' are aimed, which might have a quelling or dampening effect as non-'historian' participants dutifully defer their speculation until the 'expert historian' responds (which might never happen). The *Operation War Diary* forum contains further hints of the division between participants and 'real' historians. On one thread in which participants could suggest new hashtags,⁷⁰ a well-meaning response by a volunteer moderator suggests asking 'how might this information be useful to researchers in future?'. However, this inadvertently discounts the possibility that the participants might

diaries from the First World War.

⁶⁸Or 'Talk' pages, in *Zooniverse's* parlance

⁶⁹ The Papers of the War Department project has used the term 'Transcription Associates'. Papers of the War Department, 'Become a Transcription Associate', *Papers of the War Department*, accessed 29 May 2015, <http://wardepartmentpapers.org/transcribe.php>.

⁷⁰ A special keyword which can link potentially interesting documents together.

themselves be researchers in the future. These unintentional but revealing juxtapositions of 'real' and 'citizen' historians raise the spectre of the citizen historian as 'faux' historian.

My exploration of these issues was, in part, a response to online discourse around 'citizen' historians and other 'citizen' scholars. For example, one excitable tweet said Easter is 'a wonderful time to become a citizen archeologist!'.⁷¹ While transcribing text might help archaeology, it is, however, nothing like being or becoming an archaeologist.⁷² Similarly, calling people who have transcribed a few lines of text 'citizen historians' undermines the time and effort that others have put into developing their skills as amateur historians. It also flattens expertise: if a participant is labelled a specialist after transcribing some text, then what do we call someone with a degree and years of experience? Calling 'someone who has transcribed some text' a 'citizen historian' undervalues the skills and experience of the actual citizen historian, who (to my mind) is equivalent with 'amateur historian'; the term 'citizen' then functions as a reminder that they are helping create a shared, public resource. Some amateur historians (including many of the community historians I interviewed) have long years of experience and are highly skilled, while others are less so; but they are all judged in comparison to skilled historians. 'Contributors', or more specific terms like 'transcribers', might look less impressive on a funding application, but they are at least accurate phrases. If projects have not included an explicit plan for helping

⁷¹ @Crowdcrafting, 'Easter... a Wonderful Time ...', *Twitter*, 17 April 2014, <https://twitter.com/crowdcrafting/status/456887826563751936>. This tweet about a transcription project was written by an enthusiast rather than by the project itself.

⁷² The use of 'citizen x' labels is a widely debated issue, but to date no other term has been widely taken up. For example, a 2014 workshop found the political connotations of 'citizen' and the religious associations of 'lay' problematic, and suggested alternatives including 'volunteer researchers, explorers, quarriers and even prospectors'. Gowan Dawson, 'Crowds and Clouds Workshop, Chicago', *Constructing Scientific Communities*, 26 June 2014, <http://conscicom.org/2014/06/26/crowds-and-clouds-workshop-chicago/>. See also Theimer, 'Why We Need to Find a Term to Replace "Citizen Archivist"'.

participants learn new skills, it would be more honest to talk about 'citizen transcribers' or simply 'volunteers'.⁷³

This brings us to the question of defining 'historians'. Public participation in historical research itself has a long history,⁷⁴ but digital technologies have made the work of local and family historians more visible to academic historians. In 1994, historian Raphael Samuel said that if history 'was thought of as an activity rather than a profession, then the number of its practitioners would be legion'.⁷⁵ Are the behaviours, goals and motivations of citizen and community historians sufficiently different from the 'explanation and analysis required in the practice of academic history'⁷⁶ for their practices to form a different type of history? Definitions of 'historians' that assume publication is the final goal may not adequately reflect history as practised by those outside memory institutions and formal education. Do we need to revisit traditional definitions of 'historians' to encompass those acting outside academe? As a thought experiment in teasing out the different models for being 'a historian' made more visible through digital technologies, I have grouped them into categories that represent different types of historians.

My interviews and explorations of local and family history forums showed that some family and local historians are content to collect evidence for specific personal or local

⁷³ As discussed in *Chapter 3*, promising that transcribers can become 'citizen historians' without supporting the process and ultimately over-selling the crowdsourcing experience could reduce the ability of citizen history projects to attract participants and undermine the emerging field of citizen scholarship. It may also be unethical to promise opportunities for learning but not support them.

⁷⁴ Beckett and Watkins, 'Natural History and Local History in Late Victorian and Edwardian England'.

⁷⁵ Samuel, *Theatres of Memory*. p. 17.

⁷⁶ Andrew Jackson, 'Local and Regional History as Heritage: The Heritage Process and Conceptualising the Purpose and Practice of Local Historians', *International Journal of Heritage Studies* 14, no. 4 (July 2008): 362–79, doi:10.1080/13527250802155877.

histories. These 'field historians', like field archaeologists, collect sources and information but do not focus on interpreting them in relationship to published historiography. They may be family historians, focused on building evidence to construct a family tree, or local historians who are simply interested in seeking out and collecting material from the past. The meaning contained within the material they collect is created through its relationship with the people or places in which the researcher is interested.⁷⁷

Other local and family historians may write and share narrative histories.⁷⁸ While these narrative histories meet some definitions of historical scholarship,⁷⁹ they do not necessarily contain an explicit historical argument or theoretical framing.⁸⁰ For most non-faculty historians I interviewed, their research is an intrinsically motivated activity, undertaken for the pleasures it brings. They may have the experience or training to publish historiographic works but most were not motivated to undertake the work required to turn their research into narratives or arguments.⁸¹ This may be because they have other priorities for their free time, or because, unlike professional historians, there is

⁷⁷ In some ways this seems similar to the interest in 'then and now' sites such as *HistoryPin* that juxtapose historical images with modern 'street views'.

⁷⁸ For example, they may be shared within families, posted to the web, published in society newsletters or in peer-reviewed journals.

⁷⁹ Various definitions of historical scholarship can be summarised as: making original research about the past available for evaluation and use by others. Under these definitions, family trees and other formats for sharing original historical research would also count as scholarship. American Historical Association, 'Statement on Standards of Professional Conduct'. Lee S. Shulman, 'Taking Learning Seriously', *Change*, 1999.

⁸⁰ Here I am assuming that an implicit argument is contained in their selection of sources or the weight given to them, much as there is in the selection and description of sources in digital history projects. Based on my interviews with family and local historians, those with formal academic training in history would feel most confident but not necessarily any more keen to engage with broader historiographic arguments.

⁸¹ It is worth noting that this is also the case for at least one faculty historian I interviewed!

no extrinsic or intrinsic reward for their doing so, and the curiosity that drove their original research has been satisfied. Is this a form of what Dorn calls 'non-argument scholarship',⁸² and if so, does this form of scholarly history need to be distinguished from what could be called 'argument-based scholarship'? Since the most scholarly community historians demonstrate nearly all the skills and practices required of historians, there may be nothing more at stake than greater precision when discussing expectations for different types of historians.

Conclusion

This research has demonstrated the potential for using methods and theoretical frameworks from the field of user experience research and human-computer interaction in digital historiography.⁸³ The application of user-focused methods has allowed a focus on the reception of digital projects and tools by the public and by historians. Similar benefits may derive from further explorations of the links between concepts from history of science and technology, such as Pickering's mangle of practice,⁸⁴ with theoretical frameworks from human-computer interaction such as computer-supported cooperative work, cultural-historical activity theory (CHAT),⁸⁵ and other concepts such as social informatics.⁸⁶ These

⁸² Sherman Dorn, 'Writing History in the Digital Age» More Than an Argument about the Past?', in *Writing History in the Digital Age*, ed. Jack Dougherty and Kristen Nawrotzki, accessed 8 March 2014, <http://writinghistory.trincoll.edu/revisioning/dorn-2012-spring/>.

⁸³ The 'interdisciplinary study of the interaction of digital technology with historical practice'. Sternfeld, 'Archival Theory and Digital Historiography'.

⁸⁴ Andrew Pickering and Keith Guzik, eds., *The Mangle in Practice: Science, Society, and Becoming* (Durham: Duke University Press, 2008).

⁸⁵ Explaining 'social and cultural work practices by relating them to the cultural and historic context in which the activity is being performed'. Christine Rivers, Janko Calic, and Amy Tan, 'Combining Activity Theory and Grounded Theory for the Design of Collaborative Interfaces', *Interface*, 2009, 312–321.

⁸⁶ An 'analytical perspective which theorises and questions the nature and roles that ICT plays in social, institutional and cultural contexts'. McLoughlin and Alam, 'A Case Study of Instructor

different approaches all seek to articulate the extent to which technologies are both shaped by and embedded in social/cultural environments and practices, and could profit from an exchange of methods and knowledge.

Following my overview of participatory, digital history projects, and analysis of the effects of digital tools, methods and resources on the processes of discovering, gathering, creating and sharing, it is clear that the practice of history has already been shaped by digitality. The tangible and intangible benefits of crowdsourcing for organisations has encouraged them to create more opportunities for the public to connect meaningfully with historical materials, and the projects they create are able to provide smaller, more accessible activities because digital networks can manage the overhead of coordinating, distributing and validating tasks. The development of historical skills through participatory digital projects is possible at immense scale with the support of online communities of practice and resources. Primary sources are increasingly accessible regardless of distance and opening hours. Research can now be fitted in around other work - as useful for the faculty historian with a pile of marking as it is for the community historian with a busy professional or family life. Family and local history may once have been considered a retirees' hobby, but digital resources and tools have made it accessible to many more people. However, both faculty and community historians have reasons to limit their sharing practices, despite the technologies available for sharing resources and knowledge - the 'social-technical gap' remains an issue.

Another factor limiting the transformative effective of digital history is the gap between existing software that supports historical work (such as *Word* or *Excel*), and future software that does some of the work of historical research. Digital tools make digital

resources searchable, shareable, re-organisable, but they cannot (yet) do the work of turning text notes into structured data - understanding a string of text to be a historical address, adding geospatial coordinates and plotting the results on a map. It is possible to build processes that manage some of these tasks, but it takes persistence and some computational abilities. Most historical methods still involve processes that move at human, not computational speeds. When digital tools can produce the equivalent of *Old Bailey Online*'s marked-up text, we will see a further transformation in the scale and speed of historical research.

It can be difficult to predict the future impact of digital technologies and resources. In 1979, historian Lawrence Stone fretted about the 'gigantic amount of effort' then being spent on assembling digital data from parish registers, 'only some of which is likely to produce worthwhile results'.⁸⁷ He could not have predicted the role that digital images and transcriptions of historical documents would have in transforming the discoverability of historical sources, once they were combined with powerful internet search engines. It is hard now to predict what the intersection of digitised material, machine learning and unforeseen computational techniques will lead to. The impact of digitality is also dependent on the types of history being practiced, with the infrastructure of academia both hindering and helping historians in the adoption of digital tools and methods. However these different strands affect each other in the future, it is clear that digital tools and resources have already greatly expanded the number of people who can help make history, whether through their participation in crowdsourcing, collaborative community history projects, or by transforming each stage of the research process.

⁸⁷ Stone, 'The Revival of Narrative'.

Bibliography of works cited

Books

- Barthes, Roland. *Camera Lucida: Reflections on Photography*. London: Vintage, 2000.
- Carr, Edward Hallett. *What Is History?* New York: Knopf, 1962.
- Csikszentmihalyi, Mihaly. *Creativity: Flow and the Psychology of Discovery and Invention*. New York, NY, USA: Harper Perennial, 1997.
- . *Flow: The Psychology of Optimal Experience*. New York: Harper and Row, 1990.
- Dougherty, Jack, and Kristen Nawrotzki, eds. *Writing History in the Digital Age*, 2012.
<http://writinghistory.trincoll.edu/>.
- Dumas, Joseph S., and Janice C. Redish. *A Practical Guide to Usability Testing*. Revised edition. Chicago: University of Chicago Press, 1999.
<http://www.press.uchicago.edu/ucp/books/book/distributed/P/b05545036.html>.
- Farge, Arlette. *The Allure of the Archives*. Translated by Thomas Scott-Railton. New Haven, Connecticut: Yale University Press, 2013.
- Jordanova, Ludmilla. *History in Practice*. London: Arnold, 2000.
- . *History in Practice*. Second edition. London, U.K.: Bloomsbury Academic, 2006.
- Koster, Raph. *A Theory of Fun for Game Design*. Scottsdale, AZ: Paraglyph Press, 2005.
- Negroponte, Nicholas. *Being Digital*. London: Hodder & Stoughton, 1995.
- Oates, Briony J. *Researching Information Systems and Computing*. London: Sage, 2006.
- Pickering, Andrew, and Keith Guzik, eds. *The Mangle in Practice: Science, Society, and Becoming*. Durham: Duke University Press, 2008.
- Riden, Philip. *Local History: A Handbook for Beginners*. London, U.K.: Batsford Academic and Educational Ltd, 1983.
- Ridge, Mia, ed. *Crowdsourcing Our Cultural Heritage*. Digital Research in the Arts and Humanities. Farnham, Surrey, UK: Ashgate, 2014.
<http://www.ashgate.com/isbn/9781472410221>.
- Samuel, Raphael. *Theatres of Memory. Volume 1, Past and Present in Contemporary Culture*. London: Verso, 1994.
- Sharp, Helen, Yvonne Rogers, and Jenny Preece. *Interaction Design: Beyond Human-Computer Interaction*. Vol. Second. Chichester: Wiley, 2007.
- Shirky, Clay. *Cognitive Surplus: Creativity and Generosity in a Connected Age*. London, U.K.: Penguin, 2011.
- Wineburg, Sam. *Historical Thinking And Other Unnatural Acts: Charting The Future Of Teaching The Past*. Philadelphia: Temple University Press, 2001.

Book chapters

- Causser, Tim, and Melissa Terras. “Many Hands Make Light Work. Many Hands Together Make Merry Work”: Transcribe Bentham and Crowdsourcing Manuscript Collections’. In *Crowdsourcing Our Cultural Heritage*, edited by Mia Ridge. Digital Research in the Arts and Humanities. Farnham, Surrey, UK: Ashgate, 2014.
<http://www.ashgate.com/isbn/9781472410221>.
- Cohen, Daniel J., and Roy Rosenzweig. ‘Introduction: Promises and Perils of Digital History’. In *Digital History: A Guide to Gathering, Preserving, and Presenting the Past on the Web*. University of Pennsylvania Press, 2005.
<http://chnm.gmu.edu/digitalhistory/introduction/>.
- Csikszentmihalyi, Mihaly, and Kim Hermanson. ‘Intrinsic Motivation in Museums: Why Does One Want to Learn?’ In *Public Institutions for Personal Learning: Establishing*

- a *Research Agenda*, edited by John Falk and Lynn D. Dierking, 66–77. Washington D.C.: American Association of Museums, 1995.
- Dafis, Lyn Lewis, Lorna M. Hughes, and Rhian James. 'What's Welsh for "Crowdsourcing"? Citizen Science and Community Engagement at the National Library of Wales'. In *Crowdsourcing Our Cultural Heritage*, edited by Mia Ridge. Farnham, Surrey, UK: Ashgate, 2014. <http://www.ashgate.com/isbn/9781472410221>.
- Dorn, Sherman. 'Writing History in the Digital Age » More Than an Argument about the Past?' In *Writing History in the Digital Age*, edited by Jack Dougherty and Kristen Nawrotzki. Accessed 8 March 2014. <http://writinghistory.trincoll.edu/revisioning/dorn-2012-spring/>.
- Eccles, Kathryn, and Andrew Greg. 'Your Paintings Tagger: Crowdsourcing Descriptive Metadata for a National Virtual Collection'. In *Crowdsourcing Our Cultural Heritage*, edited by Mia Ridge. Farnham, Surrey, UK: Ashgate, 2014. <http://www.ashgate.com/isbn/9781472410221>.
- Gilliland, Anne J. 'Setting the Stage'. In *Introduction to Metadata*, edited by Murtha Baca, 2nd ed. The Getty Research Institute Publications Program. Los Angeles, CA: Getty Publications, 2008. http://www.getty.edu/research/publications/electronic_publications/intrometadata/index.html.
- Haklay, Muki. 'Citizen Science and Volunteered Geographic Information – Overview and Typology of Participation'. In *Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice*, edited by Daniel Z. Sui, Sarah Elwood, and Michael F. Goodchild. Dordrecht: Springer Netherlands, 2013. <http://www.springerlink.com/index/10.1007/978-94-007-4587-2>.
- Hitchcock, Tim. 'Digital Searching and the Re-Formulation of Historical Knowledge'. In *The Virtual Representation of the Past*, edited by Mark Greengrass and Lorna M. Hughes, 81–90. Ashgate, 2008.
- Kelly, T. Mills. 'Thinking: How Students Learn About the Past'. In *Teaching History in the Digital Age*. University of Michigan Press, 2013. <http://hdl.handle.net/2027/spo.12146032.0001.001>.
- Lakhani, Karim, and Robert Wolf. 'Why Hackers Do What They Do: Understanding Motivation and Effort in Free/Open Source Software Projects'. In *Perspectives on Free and Open Source Software*, edited by Joseph Feller, Brian Fitzgerald, Scott A. Hissam, and Karim R. Lakhani. MIT Press, 2005. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=443040.
- Lascarides, Michael, and Ben Vershbow. 'What's on the Menu?: Crowdsourcing at the New York Public Library'. In *Crowdsourcing Our Cultural Heritage*, edited by Mia Ridge. Farnham, Surrey, UK: Ashgate, 2014. <http://www.ashgate.com/isbn/9781472410221>.
- Leon, Sharon M. 'Build, Analyse and Generalise: Community Transcription of the Papers of the War Department and the Development of Scripto'. In *Crowdsourcing Our Cultural Heritage*, edited by Mia Ridge. Farnham, Surrey, UK: Ashgate, 2014. <http://www.ashgate.com/isbn/9781472410221>.
- Madsen-Brooks, Leslie. "I Nevertheless Am a Historian": Digital Historical Practice and Malpractice around Black Confederate Soldiers (Spring 2012 Version). In *Writing History in the Digital Age*, edited by Jack Dougherty and Kristen Nawrotzki, 2012. <http://writinghistory.trincoll.edu/crowdsourcing/madsen-brooks-2012-spring/>.
- Marick, Brian. 'A Manglish Way of Working: Agile Software Development'. In *The Mangle in Practice: Science, Society, and Becoming*, edited by Andrew Pickering and Keith Guzik. Durham: Duke University Press, 2008.
- Meyer, Eric T. 'Socio-Technical Interaction Networks: A Discussion of the Strengths, Weaknesses and Future of Kling's STIN Model'. In *Social Informatics: An Information Society for All? In Remembrance of Rob Kling*, edited by J. Berleur, M.I.

- Numinen, and J. Impagliazzo, 223:37–48. IFIP International Federation for Information Processing. Boston: Springer, 2006.
- Oomen, Johan, Riste Gligorov, and Michiel Hildebrand. 'Waisda?: Making Videos Findable through Crowdsourced Annotations'. In *Crowdsourcing Our Cultural Heritage*, edited by Mia Ridge. Farnham, Surrey, UK: Ashgate, 2014.
<http://www.ashgate.com/isbn/9781472410221>.
- Pickering, Andrew. 'New Ontologies'. In *The Mangle in Practice: Science, Society, and Becoming*, edited by Andrew Pickering and Keith Guzik. Durham: Duke University Press, 2008.
- Ridge, Mia. 'Crowdsourcing Our Cultural Heritage: Introduction'. In *Crowdsourcing Our Cultural Heritage*, edited by Mia Ridge. Farnham, Surrey, UK: Ashgate, 2014.
<http://www.ashgate.com/isbn/9781472410221>.
- Rosenzweig, Roy. 'Afterthoughts. Roy Rosenzweig: Everyone a Historian'. In *The Presence of the Past: Popular Uses of History in American Life*, by Roy Rosenzweig and David Thelan, 1998. <http://chnm.gmu.edu/survey/afterroy.html>.
- Simon, Nina K. 'Chapter 1: Principles of Participation'. In *The Participatory Museum*, 2010.
<http://www.participatorymuseum.org/chapter1/>.
- . 'Participatory Design and the Future of Museums'. In *Letting Go? Sharing Historical Authority in a User-Generated World*, edited by Bill Adair, Benjamin Filene, and Laura Koloski. Left Coast Press, 2012.
- Smart, Paul R., Elena Simperl, and Nigel Shadbolt. 'A Taxonomic Framework for Social Machines'. In *Social Collective Intelligence: Combining the Powers of Humans and Machines to Build a Smarter Society*, edited by Daniele Miorandi, Vincenzo Maltese, Michael Rovatsos, Anton Nijholt, and James Stewart. Berlin, Germany: Springer, 2014. <http://eprints.soton.ac.uk/362359/>.
- Thelan, David. 'Afterthoughts. David Thelen: A Participatory Historical Culture'. In *The Presence of the Past: Popular Uses of History in American Life*, by Roy Rosenzweig and David Thelan, 1998. <http://chnm.gmu.edu/survey/afterdave.html>.
- Wenger, Etienne. 'Communities of Practice and Social Learning Systems: The Career of a Concept'. In *Social Learning Systems and Communities of Practice*. Springer Verlag and the Open University, 2010.

Journal articles

- Ackerman, Mark S. 'The Intellectual Challenge of CSCW: The Gap Between Social Requirements and Technical Feasibility'. *Human-Computer Interaction* 15 (2000): 179–203.
- Alberti, Samuel J. M. M. 'Amateurs and Professionals in One County: Biology and Natural History in Late Victorian Yorkshire'. *Journal of the History of Biology* 34 (2001): 115–147.
- Allen, David, Stan Karanasios, and Mira Slavova. 'Working With Activity Theory: Context, Technology, and Information Behavior'. *Journal of the American Society for Information Science* 62, no. 4 (2011): 776–788. doi:10.1002/asi.
- Allen, James B., Jessie L. Embry, and Kahlile B. Mehr. 'Appendix I: Historical Milestones'. *Brigham Young University Studies* 34, no. 2 (1994): 343–346.
<https://journals.lib.byu.edu/spc/index.php/BYUStudies/article/viewFile/6198/5848>.
- . 'Chapter 8: Automating the Records'. *Brigham Young University Studies* 34, no. 2 (1994): 303–337. <http://www.jstor.org/stable/43042189>.
- Anderson, Sheila, Tobias Blanke, and Stuart Dunn. 'Methodological Commons: Arts and Humanities e-Science Fundamentals'. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 368, no. 1925 (18 July

- 2010): 3779–96. doi:10.1098/rsta.2010.0156.
- Andrews, Thomas, and Flannery Burke. 'What Does It Mean to Think Historically?' *Perspectives on History* 45, no. 1 (January 2007).
<https://www.historians.org/publications-and-directories/perspectives-on-history/january-2007/what-does-it-mean-to-think-historically>.
- Baca, Murtha. 'Digital Mellini: Project Update and Observations on Translating Historical Texts'. *Getty Research Journal*, no. 4 (1 January 2012): 153–60.
<http://www.jstor.org/stable/41413138>.
- Bates, Marcia J, Deborah N Wilde, and Susan Siegfried. 'Research Practices of Humanities Scholars in an Online Environment: The Getty Online Searching Project Report No. 3'. *Art History* 540, no. 3 (1995).
- Becher, Tony. 'Disciplinary Discourse'. *Studies in Higher Education* 12, no. 3 (1987): 261–74. doi:10.1080/03075078712331378052.
- Beckett, John, and Charles Watkins. 'Natural History and Local History in Late Victorian and Edwardian England: The Contribution of the Victoria County History'. *Rural History* 22, no. 01 (March 2011): 59–87. doi:10.1017/S0956793310000142.
- Benkler, Yochai. 'Coase's Penguin, or, Linux and "The Nature of the Firm"'. *The Yale Law Journal* 112, no. 3 (December 2002): 369–446. <http://www.jstor.org/stable/1562247>.
- Betts, Kristen. 'Lost in Translation: Importance of Effective Communication in Online Education'. *Online Journal of Distance Learning Administration* XII, no. II (2009).
<http://www.westga.edu/~distance/ojdla/summer122/betts122.html>.
- Biggs, John. 'What the Student Does: Teaching for Enhanced Learning'. *Higher Education Research & Development* 18, no. 1 (April 1999): 57–75. doi:10.1080/0729436990180105.
- Boonstra, Onno. 'Barriers between Historical GIS and Historical Scholarship'. *International Journal of Humanities & Arts Computing* 3, no. 1–2 (2009): 3–7. doi:10.3366/E1753854810000480.
- Bowman, Joel P., and Debbie A. Renshaw. 'Desktop Publishing: Things Gutenberg Never Taught You'. *Journal of Business Communication* 26, no. 1 (1989): 57–77.
<http://jjob.sagepub.com/content/26/1/57.short>.
- Brabham, Daren C. 'Moving the Crowd at IStockphoto: The Composition of the Crowd and Motivations for Participation in a Crowdsourcing Application'. *First Monday*, 2 June 2008.
<http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2159/1969>.
- Bradley, John, and Paul Vetch. 'Supporting Annotation as a Scholarly Tool—Experiences From the Online Chopin Variorum Edition'. *Literary and Linguistic Computing* 22, no. 2 (June 2007): 225–41. doi:10.1093/lc/fqm001.
- Braun, Virginia, and Victoria Clarke. 'Using Thematic Analysis in Psychology'. *Qualitative Research in Psychology* 3, no. 2 (2006): 77–101. doi:10.1191/1478088706qp0630a.
- Brenna, Brita. 'Clergymen Abiding in the Fields: The Making of the Naturalist Observer in Eighteenth-Century Norwegian Natural History'. *Science in Context* 24, no. 02 (April 2011): 143–166. doi:10.1017/S0269889711000044.
- Brewer, John. 'Microhistory and the Histories of Everyday Life'. *Cultural and Social History* 7, no. 1 (2010): 87–109. doi:10.2752/147800410X477359.
- Brown, John Seely, Allan Collins, and Paul Duguid. 'Situated Cognition and the Culture of Learning'. *Educational Researcher* 18, no. 1 (1989): 32–42.
<http://edr.sagepub.com/content/18/1/32.short>.
- Butterworth, Richard. 'Information Seeking and Retrieval as a Leisure Activity'. *DL-CUBA'06*, 2006, 29–32.
http://web.archive.org/web/*/http://www.ucl.ac.uk/events/dl-cubaz006/papers/Butterworth.pdf.
- Cacioppo, John T., Richard E. Petty, Jeffrey A. Feinstein, and W. Blair G. Jarvis.

- 'Dispositional Differences in Cognitive Motivation: The Life and Times of Individuals Varying in Need for Cognition'. *Psychological Bulletin* 119, no. 2 (1996): 197–253.
- Cardamone, Carolin, Kevin Schawinski, Marc Sarzi, Steven P. Bamford, Nicola Bennert, C. M. Urry, Chris Lintott, et al. 'Galaxy Zoo Green Peas: Discovery of a Class of Compact Extremely Star-Forming Galaxies'. *Monthly Notices of the Royal Astronomical Society* 399, no. 3 (November 2009): 1191–1205. doi:10.1111/j.1365-2966.2009.15383.x.
- Case, Donald Owen. 'The Collection and Use of Information by Some American Historians: A Study of Motives and Methods'. *The Library Quarterly* 61, no. 1 (1991): 61–82.
- Causer, Tim, and Melissa Terras. 'Crowdsourcing Bentham: Beyond the Traditional Boundaries of Academic History'. *International Journal of Humanities and Arts Computing* 8, no. 1 (2014): 46–64.
<http://www.eupublishing.com/doi/abs/10.3366/ijhac.2014.0119>.
- Causer, Tim, and Valerie Wallace. 'Building A Volunteer Community: Results and Findings from Transcribe Bentham'. *Digital Humanities Quarterly* 6, no. 2 (2012).
<http://www.digitalhumanities.org/dhq/vol/6/2/000125/000125.html>.
- Clary, E. Gil, Mark Snyder, Robert D Ridge, John Copeland, Arthur A Stukas, Julie Haugen, and Peter Miene. 'Understanding and Assessing the Motivations of Volunteers: A Functional Approach'. *Journal of Personality and Social Psychology* 74, no. 6 (June 1998): 1516–30. <http://www.ncbi.nlm.nih.gov/pubmed/9654757>.
- Cooper, Caren B., Janis Dickinson, Tina Phillips, and Rick Bonney. 'Ecology and Society: Citizen Science as a Tool for Conservation in Residential Ecosystems'. *Ecology and Society* 12, no. 2 (2007). <http://www.ecologyandsociety.org/vol12/iss2/art11/>.
- de Groot, Jerome. 'Invitation to Historians'. *Rethinking History* 18, no. 4 (2014): 599–612. doi:10.1080/13642529.2014.893662.
- Denning, Peter J., and Peter Yeholkovsky. 'Getting to "We": Solidarity, Not Software, Generates Collaboration'. *Communications of the ACM* 51, no. 4 (2008): 19–24.
<http://dl.acm.org/citation.cfm?id=1330316>.
- DeRidder, Jody L., and Kathryn G. Matheny. 'What Do Researchers Need? Feedback On Use of Online Primary Source Materials'. *D-Lib Magazine* 20, no. 7/8 (2014). doi:doi:10.1045/july2014-deridder.
- Drake, Michael. 'From Old Bills to Sick Pigs: Our Ways to Capture Community'. *Family & Community History* 1, no. 1 (1998): 7–26.
<http://www.maneyonline.com/doi/abs/10.1179/fch.1998.1.1.002>.
- Duff, Wendy M., and Catherine A. Johnson. 'Where Is the List with All the Names? Information-Seeking Behavior of Genealogists'. *American Archivist* 66, no. Spring/Summer (2003): 79–95.
- Edwards, Deborah, and Margaret Graham. 'Museum Volunteers and Heritage Sectors'. *Australian Journal on Volunteering* 11, no. 1 (2006): 19–27.
- Estelles-Arolas, Enrique, and Fernando Gonzalez-Ladron-de-Guevara. 'Towards an Integrated Crowdsourcing Definition'. *Journal of Information Science* 38, no. 2 (March 2012): 189–200. doi:10.1177/0165551512437638.
- Evans, Max J. 'Archives of the People, by the People, for the People'. *American Archivist* 70, no. 2 (2007): 387–400.
<http://archivists.metapress.com/index/d157t6667g54536g.pdf>.
- Evans, Tanya, and Patricia Curthoys. 'Family History, Identity, and Public History: Writing a History of The Benevolent Society in Its 200th Year'. *Journal of Australian Studies* 37, no. 3 (September 2013): 285–301. doi:10.1080/14443058.2013.814582.
- Filene, Benjamin. 'Passionate Histories: "Outsider" History-Makers and What They Teach Us'. *The Public Historian* 34, no. 1 (February 2012): 11–33. doi:10.1525/tph.2012.34.1.11.

- Flanagan, Mary, and Peter Carini. 'How Games Can Help Us Access and Understand Archival Images'. *American Archivist* 75, no. 2 (2012): 514–537.
<http://archivists.metapress.com/content/b424537w2797ogu4/>.
- Fleet, Chris. "Locating Trees in the Caledonian Forest": A Critical Assessment of Methods for Presenting Series Mapping over the Web'. *E-Perimtron* 1, no. 2 (2006): 99–112.
- Fry, Jenny, and Sanna Talja. 'The Intellectual and Social Organization of Academic Fields and the Shaping of Digital Resources'. *Journal of Information Science* 33, no. 2 (February 2007): 115–133. doi:10.1177/0165551506068153.
- Fulton, Crystal. 'Finding Pleasure in Information Seeking: Leisure and Amateur Genealogists Exploring Their Irish Ancestry'. *Proceedings of the American Society for Information Science and Technology* 42, no. 1 (2005).
 doi:10.1002/meet.14504201228.
- . 'Quid pro Quo: Information Sharing in Leisure Activities'. *Library Trends* 57, no. 4 (2009): 753–768.
<http://muse.jhu.edu/journals/lib/summary/v057/57.4.fulton01.html>.
- Goldstein, Daniel. "Yours for Science": The Smithsonian Institution's Correspondents and the Shape of Scientific Community in Nineteenth-Century America'. *Isis* 85, no. 4 (December 1994): 573–599. <http://www.jstor.org/stable/235279>.
- Goodchild, Michael F, and Donald G Janelle. 'Toward Critical Spatial Thinking in the Social Sciences and Humanities'. *GeoJournal* 75, no. 1 (February 2010): 3–13.
 doi:10.1007/s10708-010-9340-3.
- Graham, Shawn, Ian Milligan, and Scott Weingart. 'Writing The Historian's Macroscopic in Public'. *Perspectives on History*, October 2014.
<http://www.historians.org/publications-and-directories/perspectives-on-history/october-2014/writing-the-historian%E2%80%99s-macroscopic-in-public>.
- Guest, Greg, Arwen Bunce, and Laura Johnson. 'How Many Interviews Are Enough?: An Experiment with Data Saturation and Variability'. *Field Methods* 18, no. 1 (1 February 2006): 59–82. doi:10.1177/1525822X05279903.
- Gura, Trisha. 'Citizen Science: Amateur Experts'. *Nature* 496, no. 7444 (10 April 2013): 259–61. doi:10.1038/nj7444-259a.
- Hirsch, Brett D. 'Digital Renaissance Editions'. *Journal for Early Modern Cultural Studies* 13, no. 4 (2013).
http://muse.jhu.edu/journals/journal_for_early_modern_cultural_studies/v013/13.4.hirsch.html.
- Hitchcock, Tim. 'Confronting the Digital: Or How Academic History Writing Lost the Plot'. *Cultural and Social History* 10, no. 1 (1 March 2013): 9–23.
 doi:10.2752/147800413X13515292098070.
- Hitchcock, Tim, and Robert Shoemaker. 'Digitising History From Below: The Old Bailey Proceedings Online, 1674–1834: The Old Bailey Proceedings Online, 1674–1834'. *History Compass* 4, no. 2 (March 2006): 193–202. doi:10.1111/j.1478-0542.2006.00309.x.
- Holland, Michael. 'Swing Revisited: The Swing Project'. *Family & Community History* 7, no. 2 (2004): 87–100. <http://www.maneyonline.com/doi/abs/10.1179/fch.2004.7.2.002>.
- Holley, Rose. 'Crowdsourcing: How and Why Should Libraries Do It?' *D-Lib Magazine* 16, no. 3/4 (April 2010). doi:10.1045/march2010-holley.
- . 'How Good Can It Get? Analysing and Improving OCR Accuracy in Large Scale Historic Newspaper Digitisation Programs'. *D-Lib Magazine* 15, no. 3/4 (2009).
 doi:10.1045/march2009-holley.
- . 'Resource Sharing in Australia: Find and Get in Trove – Making "Getting" Better'. *D-Lib Magazine* 17, no. 3/4 (2011). doi:10.1045/march2011-holley.
- Holmes, Kirsten. 'Volunteers in the Heritage Sector: A Neglected Audience?' *International Journal of Heritage Studies* 9, no. 4 (December 2003): 341–355.

- doi:10.1080/1352725022000155072.
- Huvila, Isto. 'Participatory Archive: Towards Decentralised Curation, Radical User Orientation, and Broader Contextualisation of Records Management'. *Archival Science* 8, no. 1 (September 2008): 15–36. doi:10.1007/s10502-008-9071-0.
- Jackson, Andrew. 'Local and Regional History as Heritage: The Heritage Process and Conceptualising the Purpose and Practice of Local Historians'. *International Journal of Heritage Studies* 14, no. 4 (July 2008): 362–79. doi:10.1080/13527250802155877.
- Jeater, Diana. 'Terence Ranger: Life as Historiography'. *History Workshop*, 16 July 2011. <http://www.historyworkshop.org.uk/terence-ranger-life-as-historiography/>.
- Jessop, Martyn. 'The Inhibition of Geographical Information in Digital Humanities Scholarship'. *Literary and Linguistic Computing* 23, no. 1 (December 2007): 39–50. doi:10.1093/llc/fqmo41.
- Jordan Raddick, M., Georgia Bracey, Pamela L. Gay, Chris J. Lintott, Carie Cardamone, Phil Murray, Kevin Schawinski, Alexander S. Szalay, and Jan Vandenberg. 'Galaxy Zoo: Motivations of Citizen Scientists'. *Astronomy Education Review* 12, no. 1 (December 2013). doi:10.3847/AER2011021.
- Journal, Jennifer International. 'Motivation, Ability and Opportunity to Participate: A Reconceptualization of the RAND Model of Audience Development'. *International Journal of Arts Management* 7, no. 1 (2004).
- Kean, Hilda. 'People, Historians, and Public History: Demystifying the Process of History Making'. *The Public Historian* 32, no. 3 (August 2010): 25–38. doi:10.1525/tph.2010.32.3.25.
- Kling, Rob. 'Computerization and Social Transformations'. *Science, Technology, & Human Values* 16, no. 3 (1991): 342–67. <http://www.jstor.org/stable/689919>.
- Kohler, Robert E. 'Finders, Keepers: Collecting Sciences and Collecting Practice'. *History of Science* 45, no. 4 (2007): 428–455. doi:10.1177/007327530704500403.
- Kraut, Robert, Mary Lou Maher, Judith Olson, Thomas W. Malone, Peter Pirolli, and John C. Thomas. 'Scientific Foundations: A Case for Technology-Mediated Social-Participation Theory'. *Computer* 43, no. 11 (November 2010): 22–28. doi:10.1109/MC.2010.324.
- Lindquist, Thea, and Holley Long. 'How Can Educational Technology Facilitate Student Engagement with Online Primary Sources?: A User Needs Assessment'. *Library Hi Tech* 29, no. 2 (2011): 224–41. doi:10.1108/07378831111138152.
- Litman, Jordan. 'Curiosity and the Pleasures of Learning: Wanting and Liking New Information'. *Cognition & Emotion* 19, no. 6 (September 2005): 793–814. doi:10.1080/02699930541000101.
- Mahoney, Anne. 'Tachypaedia Byzantina: The Suda On Line as Collaborative Encyclopedia'. *Digital Humanities Quarterly* 3, no. 1 (2009). <http://www.digitalhumanities.org/dhq/vol/003/1/000025.html>.
- Mathisen, Ralph W. 'Where Are All the PDBs?: The Creation of Prosopographical Databases for the Ancient and Medieval Worlds'. *Chronicle of Higher Education*, 1989, A7. <http://prosopography.modhist.ox.ac.uk/images/04%20Mathisen%20pdf.pdf>.
- Mayer, Richard E., and Roxana Moreno. 'Nine Ways to Reduce Cognitive Load in Multimedia Learning'. *Educational Psychologist* 38, no. 1 (2003): 43–52. http://www.tandfonline.com/doi/full/10.1207/S15326985EP3801_6.
- McInerney, Claire. 'Knowledge Management and the Dynamic Nature of Knowledge'. *Journal of the American Society for Information Science and Technology* 53, no. 12 (2002): 1009–1018. <http://onlinelibrary.wiley.com/doi/10.1002/asi.10109/full>.
- McLoughlin, Catherine. 'Learner Support in Distance and Networked Learning Environments: Ten Dimensions for Successful Design'. *Distance Education* 23, no. 2

- (2002): 149–162.
<http://www.tandfonline.com/doi/abs/10.1080/0158791022000009178>.
- McLoughlin, Catherine E., and Sultana Lubna Alam. 'A Case Study of Instructor Scaffolding Using Web 2.0 Tools to Teach Social Informatics'. *Journal of Information Systems Education* 25, no. 2 (2014): 125.
<https://www.questia.com/library/journal/1G1-406164672/a-case-study-of-instructor-scaffolding-using-web-2-0>.
- Meho, Lokman I., and Helen R. Tibbo. 'Modeling the Information-Seeking Behavior of Social Scientists: Ellis's Study Revisited'. *Journal of the American Society for Information Science and Technology* 54, no. 6 (2003): 570–587.
<http://onlinelibrary.wiley.com/doi/10.1002/asi.10244/full>.
- Moyle, Martin, Justin Tonra, and Valerie Wallace. 'Manuscript Transcription by Crowdsourcing: Transcribe Bentham'. *Liber Quarterly* 20, no. 3–4 (2011): 347–356.
<http://liber.library.uu.nl/index.php/lq/article/view/7999>.
- Mugar, Gabriel, Carsten Østerlund, Katie DeVries Hassman, Kevin Crowston, and Corey Brian Jackson. 'Planet Hunters and Seafloor Explorers: Legitimate Peripheral Participation Through Practice Proxies in Online Citizen Science'. Accessed 3 June 2014.
http://crowston.syr.edu/sites/crowston.syr.edu/files/paper_revised%20copy%20to%20post.pdf.
- Norman, Donald A. 'Affordance, Conventions, and Design'. *Interactions* 6, no. 3 (1999): 38–43. <http://dl.acm.org/citation.cfm?id=301168>.
- 'Notes and Queries'. *Notes and Queries* 1, no. 1 (3 November 1849).
<http://www.bodley.ox.ac.uk/cgi-bin/ilej/image1.pl?item=page&seq=1&size=1&id=nq.1849.11.3.1.1.x.1>.
- Nowviskie, Bethany. 'Evaluating Collaborative Digital Scholarship (or, Where Credit Is Due)'. *Journal of Digital Humanities* 1, no. 1 (2012).
<http://journalofdigitalhumanities.org/1-4/evaluating-collaborative-digital-scholarship-by-bethany-nowviskie/>.
- Oreg, Shaul, and Oded Nov. 'Exploring Motivations for Contributing to Open Source Initiatives: The Roles of Contribution Context and Personal Values'. *Computers in Human Behavior* 24, no. 5 (September 2008): 2055–2073.
doi:10.1016/j.chb.2007.09.007.
- Owens, J. B. 'Toward a Geographically-Integrated, Connected World History: Employing Geographic Information Systems (GIS)'. *History Compass* 5, no. 6 (November 2007): 2014–40. doi:10.1111/j.1478-0542.2007.00476.x.
- Paas, Fred, Alexander Renkl, and John Sweller. 'Cognitive Load Theory and Instructional Design: Recent Developments'. *Educational Psychologist* 38, no. 1 (March 2003): 1–4. doi:10.1207/S15326985EP3801_1.
- Paas, Fred, Juhani E. Tuovinen, Huib Tabbers, and Pascal W.M. Van Gerven. 'Cognitive Load Measurement as a Means to Advance Cognitive Load Theory'. *Educational Psychologist* 38, no. 1 (2003): 63–71.
http://www.tandfonline.com/doi/full/10.1207/S15326985EP3801_8.
- Parry, Ross. 'The End of the Beginning: Normativity in the Postdigital Museum'. *Advances in Research - Museum* 1, no. 1 (1 July 2013): 24–39. doi:10.3167/armw.2013.010103.
- Peddibhotla, Naren B., and Mani R. Subramani. 'Contributing to Public Document Repositories: A Critical Mass Theory Perspective'. *Organization Studies* 28, no. 3 (1 March 2007): 327–46. doi:10.1177/0170840607076002.
- Pennock, Maureen, and Chris Clark. 'Saving the Sounds of the UK in the UK SoundMap'. *Ariadne*, no. 66 (30 January 2011). <http://www.ariadne.ac.uk/issue66/pennock-clark>.
- Preece, Jennifer, and Ben Shneiderman. 'The Reader-to-Leader Framework: Motivating

- Technology-Mediated Social Participation'. *AIS Transactions on Human-Computer Interaction* 1, no. 1 (2009): 13–32.
<http://www.cs.umd.edu/~ben/papers/Jennifer2009Reader.pdf>.
- Raddick, M. Jordan, Georgia Bracey, Pamela L. Gay, Chris J. Lintott, Phil Murray, Kevin Schawinski, Alexander S. Szalay, and Jan Vandenberg. 'Galaxy Zoo: Exploring the Motivations of Citizen Science Volunteers'. *Astronomy Education Review* 9, no. 1 (2010): 18. <http://dx.doi.org/10.3847/AER2009036>.
- Ransom, Joann. 'Kete Horowhenua: The Story of the District as Told by Its People'. Accessed 15 June 2014.
http://www.valaconf.org.au/vala2008/papers2008/31_Ransom_Final.pdf.
- Reed, Jason, M. Jordan Raddick, Andrea Lardner, and Karen Carney. 'An Exploratory Factor Analysis of Motivations for Participating in Zooniverse, a Collection of Virtual Citizen Science Projects'. 2013 46th Hawaii International Conference on System Sciences, January 2013, 610–619. doi:10.1109/HICSS.2013.85.
- Ridge, Mia. 'From Tagging to Theorizing: Deepening Engagement with Cultural Heritage through Crowdsourcing'. *Curator: The Museum Journal* 56, no. 4 (October 2013).
- Rivers, Christine, Janko Calic, and Amy Tan. 'Combining Activity Theory and Grounded Theory for the Design of Collaborative Interfaces'. *Interface*, 2009, 312–321.
- Robyns, Marcus C. 'The Archivist as Educator: Integrating Critical Thinking Skills into Historical Research Methods Instruction'. *American Archivist* 64, no. 2 (2001): 363–384. <http://archivists.metapress.com/index/Q4742X2324J10457.pdf>.
- Rogers, Yvonne. 'New Theoretical Approaches for Interaction'. *Human-Computer Interaction*, 2002, 87–143.
- Rosenzweig, Roy. 'Can History Be Open Source? Wikipedia and the Future of the Past'. *The Journal of American History* 93, no. 1 (June 2006): 117–46.
<https://chnm.gmu.edu/essays-on-history-new-media/essays/?essayid=42>.
- Rowley, J. 'The Wisdom Hierarchy: Representations of the DIKW Hierarchy'. *Journal of Information Science* 33, no. 2 (15 February 2007): 163–80.
doi:10.1177/0165551506070706.
- Seaman, David. 'Discovering the Information Needs of Humanists When Planning an Institutional Repository'. *D-Lib Magazine* 17, no. 3/4 (March 2011).
doi:10.1045/march2011-seaman.
- Secord, Anne. 'Corresponding Interests: Artisans and Gentlemen in Nineteenth-Century Natural History'. *The British Journal for the History of Science* 27, no. 04 (1994): 383–408. doi:10.1017/S0007087400032416.
- Sen, Shilad, Shyong K. Lam, Al Mamunur Rashid, Dan Cosley, Dan Frankowski, Jeremy Osterhouse, F. Maxwell Harper, and John Riedl. 'Tagging, Communities, Vocabulary, Evolution'. *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work - CSCW '06*, 2006, 181.
doi:10.1145/1180875.1180904.
- Sharma, Priya, and Michael J. Hannafin. 'Scaffolding in Technology-Enhanced Learning Environments'. *Interactive Learning Environments* 15, no. 1 (April 2007): 27–46.
doi:10.1080/10494820600996972.
- Silvertown, Jonathan. 'A New Dawn for Citizen Science.' *Trends in Ecology & Evolution* 24, no. 9 (September 2009): 467–71. doi:10.1016/j.tree.2009.03.017.
- Smith, Sue. 'Work in Progress Victorian Stationmasters: Origins and Outcomes'. *Family & Community History* 15, no. 2 (October 2012): 113–17.
doi:10.1179/1463118013Z.0000000007.
- Southall, Humphrey, Ruth Mostern, and Merrick Lex Berman. 'On Historical Gazetteers'. *International Journal of Humanities and Arts Computing* 5, no. 2 (October 2011): 127–145. doi:10.3366/ijhac.2011.0028.
- Star, Susan Leigh, and James R. Griesemer. 'Institutional Ecology, "Translations" and

- Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39'. *Social Studies of Science* 19, no. 3 (1 August 1989): 387-420. <http://www.jstor.org/stable/285080>.
- Stebbins, Robert A. 'Casual Leisure: A Conceptual Statement'. *Leisure Studies* 16, no. 1 (January 1997): 17-25. doi:10.1080/026143697375485.
- Sternfeld, Joshua. 'Archival Theory and Digital Historiography: Selection, Search, and Metadata as Archival Processes for Assessing Historical Contextualization'. *American Archivist* 74, no. 2 (2011): 544-575. <http://archivists.metapress.com/index/644851P6GMG432Ho.pdf>.
- Stone, Lawrence. 'The Revival of Narrative: Reflections on a New Old History'. *Past & Present*, no. 85 (1 November 1979): 3-24. <http://www.jstor.org/stable/650677>.
- Suri, Venkata Ratnadeep. 'The Assimilation and Use of GIS by Historians: A Sociotechnical Interaction Networks (STIN) Analysis'. *International Journal of Humanities and Arts Computing* 5, no. 2 (October 2011): 159-188. doi:10.3366/ijhac.2011.0030.
- Talja, Sanna. 'Information Sharing in Academic Communities: Types and Levels of Collaboration in Information Seeking and Use'. *New Review of Information Behavior Research* 3 (2002).
- Tally, Bill, and Lauren B. Goldenberg. 'Fostering Historical Thinking With Digitized Primary Sources'. *Journal of Research on Technology in Education* 38, no. 1 (2005): 1-21.
- Teddlie, C., and F. Yu. 'Mixed Methods Sampling: A Typology With Examples'. *Journal of Mixed Methods Research* 1, no. 1 (1 January 2007): 77-100. doi:10.1177/2345678906292430.
- Terras, Melissa. 'Digital Curiosities: Resource Creation via Amateur Digitization'. *Literary and Linguistic Computing* 25, no. 4 (2010): 425-438. doi:10.1093/lilc/fqq019.
- Theimer, Kate. 'A Distinction Worth Exploring: "Archives" and "Digital Historical Representations"'. *Journal of Digital Humanities* 3, no. 2 (Summer 2014). <http://journalofdigitalhumanities.org/3-2/a-distinction-worth-exploring-archives-and-digital-historical-representations/>.
- Tibbo, Helen R. 'Primarily History in America: How US Historians Search for Primary Materials at the Dawn of the Digital Age'. *American Archivist* 6 (2003): 9-50. <http://archivists.metapress.com/index/b1203701g718n74.pdf>.
- Tosh, John. 'Public History, Civic Engagement and the Historical Profession in Britain'. *History* 99, no. 335 (2014): 191-212. <http://onlinelibrary.wiley.com/doi/10.1111/1468-229X.12053/full>.
- Twidale, Michael B., David M. Nichols, and Chris D. Paice. 'Browsing Is a Collaborative Process'. *Information Processing & Management* 33, no. 6 (November 1997): 761-83. doi:10.1016/S0306-4573(97)00040-X.
- Van Merriënboer, Jeroen J. G., Paul A. Kirschner, and Liesbeth Kester. 'Taking the Load off a Learner's Mind: Instructional Design for Complex Learning'. *Educational Psychologist* 38, no. 1 (2003): 5-13. http://www.tandfonline.com/doi/full/10.1207/S15326985EP3801_2.
- Vetter, Jeremy. 'Introduction: Lay Participation in the History of Scientific Observation'. *Science in Context* 24, no. 02 (April 2011): 127-141. doi:10.1017/S0269889711000032.
- . 'Lay Observers, Telegraph Lines, and Kansas Weather: The Field Network as a Mode of Knowledge Production'. *Science in Context* 24, no. 02 (April 2011): 259-280. doi:10.1017/S0269889711000093.
- White, David S., and Alison Le Cornu. 'Visitors and Residents: A New Typology for Online Engagement'. *First Monday* 16, no. 9 (2011). doi:10.5210/fm.v16i9.3171.
- Willett, Kyle W., Chris J. Lintott, Steven P. Bamford, Karen L. Masters, Brooke D. Simmons, Kevin R. V. Casteels, Edward M. Edmondson, et al. 'Galaxy Zoo 2: Detailed Morphological Classifications for 304 122 Galaxies from the Sloan Digital

- Sky Survey'. *Monthly Notices of the Royal Astronomical Society*, 2013, stt1458.
<http://mnras.oxfordjournals.org/content/early/2013/09/15/mnras.stt1458.short>.
- Wineburg, Sam. 'Historical Problem Solving: A Study of the Cognitive Processes Used in the Evaluation of Documentary and Pictorial Evidence.' *Journal of Educational Psychology* 83, no. 1 (1991): 73. <http://psycnet.apa.org/journals/edu/83/1/73/>.
- Wood, David, Jerome S. Bruner, and Gail Ross. 'The Role of Tutoring in Problem Solving'. *Journal of Child Psychology and Psychiatry, and Allied Disciplines* 17, no. 2 (1976): 89–100. <http://www.ncbi.nlm.nih.gov/pubmed/932126>.
- Wrigley, E. A. 'Small-Scale but Not Parochial: The Work of the Cambridge Group for the History of Population and Social Structure'. *Family & Community History* 1, no. 1 (1998): 27–36. <http://www.maneyonline.com/doi/abs/10.1179/fch.1998.1.1.003>.
- Yakel, Elizabeth. 'Seeking Information, Seeking Connections, Seeking Meaning: Genealogists and Family Historians.' *Information Research* 10, no. 1 (October 2004). <http://informationr.net/ir/10-1/paper205.html>.
- Yakel, Elizabeth, and Deborah A. Torres. 'AI: Archival Intelligence and User Expertise'. *The American Archivist* 66, no. Spring/Summer (2003): 51–78.
- . 'Genealogists as a "Community of Records"' 70, no. 1 (2007): 93–113.
- Zhang, Ping, and Gisela M. von Dran. 'Satisfiers and Dissatisfiers: A Two-Factor Model for Website Design and Evaluation'. *Journal of the American Society for Information Science* 51, no. 14 (2000): 1253–1268.
[http://onlinelibrary.wiley.com/doi/10.1002/1097-4571\(2000\)9999:9999%3C::AID-AS11039%3E3.o.CO;2-O/full](http://onlinelibrary.wiley.com/doi/10.1002/1097-4571(2000)9999:9999%3C::AID-AS11039%3E3.o.CO;2-O/full).

Newspaper articles

- Crown, Sarah. 'Good News! Dickens Detectives Flooding In'. *The Guardian*, 5 August 2011.
<http://www.theguardian.com/books/booksblog/2011/aug/05/dickens-detectives-found>.
- Drew, John. 'Thanks to You All for a Dickens of a Job Done'. *The Guardian*, 21 June 2012, sec. Letters. <http://www.theguardian.com/books/2012/jun/21/thanks-you-dickens-job-done>.
- 'Nasa Seeks Coders to Hunt Asteroids'. *BBC News*, 11 March 2014.
<http://www.bbc.co.uk/news/technology-26528516>.
- Olson, Elizabeth. 'Smithsonian Expands Its Reach Through Social Media and the Public'. *NYTimes.Com*, 16 March 2011.
<http://www.nytimes.com/2011/03/17/arts/design/smithsonian-expands-its-reach-through-social-media-and-the-public.html>.
- Rincon, Paul. 'Teacher Finds New Cosmic Object'. *BBC NEWS*, sec. Science/Nature. Accessed 25 June 2014. <http://news.bbc.co.uk/1/hi/sci/tech/7543776.stm>.
- Wright, Alex. 'Managing Scientific Inquiry in a Laboratory the Size of the Web'. *The New York Times*, 27 December 2010, sec. Science.
<http://www.nytimes.com/2010/12/28/science/28citizen.html?pagewanted=all>.
- . 'New Search Technologies Mine the Web More Deeply'. *The New York Times*, 22 February 2009, sec. Technology / Internet.
<http://www.nytimes.com/2009/02/23/technology/internet/23search.html>.

Dissertations

- Ridge, Mia. 'Playing with Difficult Objects: Game Designs for Crowdsourcing Museum Metadata'. MSc Dissertation, City University London, 2011.
<http://www.miaridge.com/my-msc-dissertation-crowdsourcing-games-for->

museums/.

Other (blog posts, conference papers, forum posts, interviews, magazine articles, presentations, reports, webpages)

- 18thConnect. 'About'. *18thConnect*. Accessed 19 February 2015.
<http://www.18thconnect.org/about/>.
- . 'Peer Review'. *18thConnect*. Accessed 16 May 2015.
<http://www.18thconnect.org/about/scholarship/peer-review/>.
- 1940 US Census Community Project. 'We Did It! The 1940 US Census Community Project', 2012. <http://us2.campaign-archive2.com/?u=bode542dc933cfc848d187ea&id=c6e095aa92>.
- 'About'. *Free UK Genealogy*. Accessed 11 February 2015.
<http://www.freeukgenealogy.org.uk/about/>.
- 'About'. *Know Your Bristol*. Accessed 1 March 2015. <http://knowyourbristol.org/about-us/>.
- 'About Ancestry.Co.Uk'. *Ancestry*. Accessed 26 February 2015.
<http://www.ancestry.co.uk/cs/legal/Overview>.
- 'About FamilySearch'. *FamilySearch.Org*. Accessed 12 February 2015.
<https://familysearch.org/about>.
- 'About LibriVox'. *LibriVox*. Accessed 27 February 2015. <https://librivox.org/pages/about-librivox/>.
- 'About the Ancestry.Com World Archives Project'. *Ancestry*. Accessed 26 December 2013.
<http://landing.ancestry.com/wap/learnmore.aspx>.
- 'About the Gravestone Photographic Resource Project Grave Directory'. *Gravestone Photographic Resource*, 2015.
<http://www.gravestonephotos.com/information/about.php?r=1&width=1920&browser=1110>.
- 'About the USGenWeb'. *USGenWeb Project*. Accessed 12 February 2015.
<http://www.usgenweb.org/about/index.shtml>.
- 'About the USGenWeb Archives'. *USGenWeb Archives*. Accessed 12 February 2015.
<http://www.usgwarchives.net/projecthistory.htm>.
- 'About Us'. Accessed 12 January 2012. <http://www.ucl.ac.uk/transcribe-bentham/about/>.
- 'About Us'. *Historic Graves*. Accessed 2 March 2015.
<http://historicgraves.com/content/about-us>.
- Abraham, Rebekkah. 'Reflections on Putting Art on the Map'. *Historypin Blog*, 20 April 2014. <http://blog.historypin.com/2014/04/20/reflections-on-putting-art-on-the-map/>.
- Adler, Rudy. 'Ancestry.Com Acquires 1000memories'. *1000memories*, 3 October 2012.
<http://blog.1000memories.com/170-ancestry-acquires-1000memories>.
- AHDS History. 'AHDS History - Depositing Data with the AHDS'. *AHDS History*, University of Essex, March 2007.
<http://www.ahds.ac.uk/history/depositing/index.html>.
- Alam, Sultana, and John Campbell. 'Crowdsourcing Motivations in a Not-for-Profit GLAM Context: The Australian Newspapers Digitisation Program'. In *ACIS 2012: Location, Location, Location: Proceedings of the 23rd Australasian Conference on Information Systems 2012*, 1–11, 2012. <http://dro.deakin.edu.au/eserv/DU:30049107/alam-crowdsourcingmotivations-2012.pdf>.
- Albrezzi, Francesca. 'Creating "Getty Scholars" Workspace": Lessons from the Digital Humanities Trenches'. *The Getty Iris*, 6 March 2013.
<http://blogs.getty.edu/iris/creating-getty-scholars-workspace-lessons-from-the-digital-humanities-trenches/>.

- Allori, Lorenzo, and Lisa Kaborycha. 'Opening Aladdin's Cave or Pandora's Box? The Challenges of Crowdsourcing the Medici Archives'. Lincoln, Nebraska, 2013. <http://dh2013.unl.edu/abstracts/ab-312.html>.
- American Historical Association. 'AHA History Tuning Project: History Discipline Core'. *American Historical Association*, September 2013. <http://www.historians.org/teaching-and-learning/current-projects/tuning/history-discipline-core>.
- . 'Statement on Standards of Professional Conduct'. *American Historical Association*, January 2011. <http://www.historians.org/about-aha-and-membership/governance/policies-and-documents/statement-on-standards-of-professional-conduct>.
- Ancestry World Archives Project. 'With about 90 Minutes Left There Are...' *Facebook*, 31 January 2015. <https://www.facebook.com/AncestryWorldArchivesProject/posts/10153104225303487>.
- 'Ancestry World Archives Project Profile Page'. *Facebook*. Accessed 11 February 2015. <https://www.facebook.com/AncestryWorldArchivesProject>.
- Anderson, Jennifer. 'A Message from the Indexing Workforce Team'. *FamilySearch Blog*, 6 May 2013. <https://familysearch.org/blog/en/message-indexing-workforce-team/>.
- . 'Arbitration Results Ruining Your Day?' *FamilySearch Blog*, 23 April 2012. <https://familysearch.org/blog/en/arbitration-results-ruining-day/>.
- . 'The Key to Unlocking Their Stories'. *FamilySearch Blog*, 28 August 2013. <https://familysearch.org/blog/en/key-unlocking-stories/>.
- Arends, Max, Josef Froschauer, Doron Goldfarb, Dieter Merkl, and Martin Weingartner. 'Analysing User Motivation in an Art Folksonomy'. In *Proceedings of the 12th International Conference on Knowledge Management and Knowledge Technologies (i-KNOW '12)*. New York: ACM, 2012. doi:10.1145/2362456.2362473.
- ARL Working Group on Special Collections. 'Special Collections in ARL Libraries: A Discussion Report from the ARL Working Group on Special Collections'. Washington, DC: Association of Research Libraries, March 2009. <http://www.arl.org>.
- Arnone, Marilyn P. 'Using Instructional Design Strategies To Foster Curiosity'. Syracuse, NY: ERIC Clearinghouse on Information and Technology, August 2003. <http://www.ericdigests.org/2004-3/foster.html>.
- Averkamp, Shawn, and Matthew Butler. 'The Care and Feeding of a Crowd'. In *Code4Lib Conference*, 2013. http://ir.uiowa.edu/lib_pubs/129.
- Badzmierowska, Karolina. 'Letters of 1916 Celebrates Its First Birthday!', 29 September 2014. <http://dh.tcd.ie/letters1916/letters-1916-celebrates-first-birthday/>.
- Barry, Nathan. 'A Lesson in Gradual Engagement'. *UX Booth*, 18 December 2012. <http://www.uxbooth.com/articles/a-lesson-in-gradual-engagement/>.
- Bedford, Aurora. 'Instructional Overlays and Coach Marks for Mobile Apps'. *Nielsen Norman Group*, 16 February 2014. <http://www.nngroup.com/articles/mobile-instructional-overlay/>.
- 'Beginner's Tutorial'. *Recogito*. Accessed 15 May 2015. <http://pelagios.org/recogito/static/documentation/index.html>.
- Benardou, Agiatis, Panos Constantopoulos, Costis Dallas, and Dimitris Gavriliis. 'A Conceptual Model for Scholarly Research Activity'. In *ICConference 2010 Proceedings*, edited by John Unsworth, Howard Rosenbaum, and Karen E. Fisher, 26–32. Urbana-Champaign, IL: University of Illinois, 2010. <http://hdl.handle.net/2142/14945>.
- Bitgood, Stephen. 'An Attention-Value Model of Museum Visitors'. Center for the Advancement of Informal Science Education, 2010.

- http://caise.insci.org/uploads/docs/VSA_Bitgood.pdf.
- Blaauboer, Rob. 'Nichesourcing helpt het Rijksmuseum collecties in kaart brengen'. *Frankwatching*, 20 March 2014.
<http://www.frankwatching.com/archive/2014/03/20/nichesourcing-helpt-het-rijksmuseum-collecties-kaart-brengen/>.
- Bonney, Rick, Heidi Ballard, Rebecca Jordan, Ellen McCallie, Tina Phillips, Jenifer Shirk, and Candie C. Wilderman. 'Public Participation in Scientific Research: Defining the Field and Assessing Its Potential for Informal Science Education. A CAISE Inquiry Group Report'. Washington D.C.: Center for Advancement of Informal Science Education (CAISE), July 2009.
<http://caise.insci.org/uploads/docs/PPSR%20report%20FINAL.pdf>.
- Botanical Society of Britain & Ireland. 'Herbaria'. *Botanical Society of Britain & Ireland*. Accessed 22 June 2014. <http://www.bsbi.org.uk/herbaria.html>.
- Botticelli, P. 'Comment on "Archivists and Historians—Am I Giving Archivists Too Much Credit?"' *ArchivesNext*, 4 January 2012.
<http://www.archivesnext.com/?p=2434&cpag=1#comment-284643>.
- Brohan, Philip. 'HMS Africa in Action against Orthomyxoviridae'. *Old Weather Blog*, 12 January 2011. <http://blog.oldweather.org/2011/01/12/hms-africa-in-action-against-orthomyxoviridae/>.
- . 'In Search of Lost Weather'. *Old Weather Blog*, 18 August 2014.
<http://blog.oldweather.org/2014/08/18/in-search-of-lost-weather/>.
- . 'New Crew Members at Naval-History.Net'. *Old Weather Blog*, 23 May 2014.
<http://blog.oldweather.org/2014/05/23/new-crew-members-at-naval-history-net/>.
- . 'There's a Green One and a Pink One and a Blue One and a Yellow One'. *Old Weather Blog*, 5 September 2012. <http://blog.oldweather.org/2012/09/05/theres-a-green-one-and-a-pink-one-and-a-blue-one-and-a-yellow-one/>.
- Brumfield, Ben W. 'Crowdsourcing at IMLS WebWise 2012'. *Collaborative Manuscript Transcription*, 17 March 2012.
<http://manuscripttranscription.blogspot.com.au/2012/03/crowdsourcing-at-imls-webwise-2012.html>.
- . 'Itinera Nova in the World(s) of Crowdsourcing and TEI'. *Collaborative Manuscript Transcription*, 29 April 2013.
<http://manuscripttranscription.blogspot.co.uk/2013/04/itinera-nova-in-worlds-of-crowdsourcing.html>.
- . 'Progress Report: GitHub, Archive.Org Integration, and General Availability'. *Collaborative Manuscript Transcription*, 4 January 2011.
<http://manuscripttranscription.blogspot.co.uk/2011/01/progress-report-github-archiveorg.html>.
- . 'Quality Control for Crowdsourced Transcription'. *Collaborative Manuscript Transcription*, 5 March 2012.
<http://manuscripttranscription.blogspot.co.uk/2012/03/quality-control-for-crowdsourced.html>.
- . 'Wikilinks in FromThePage'. *Collaborative Manuscript Transcription*, 14 March 2014. <http://manuscripttranscription.blogspot.co.uk/2014/03/wiki-links-in-fromthepage.html>.
- Bryant, Susan L., Andrea Forte, and Amy Bruckman. 'Becoming Wikipedian: Transformation of Participation in a Collaborative Online Encyclopedia'. In *Proceedings of the 2005 International ACM SIGGROUP Conference on Supporting Group Work*, 1–10. ACM, 2005. <http://dl.acm.org/citation.cfm?id=1099205>.
- Budiu, Raluca. 'Login Walls Stop Users in Their Tracks'. *Nielsen Norman Group*, 2 March 2014. <http://www.nngroup.com/articles/login-walls/>.
- . 'Memory Recognition and Recall in User Interfaces'. *Nielsen Norman Group*, 6 July

2014. <http://www.nngroup.com/articles/recognition-and-recall/>.
- Byrne, Ciara. 'Armchair Astronomers Help Discover Why Galaxies Stop Producing Stars'. *Fast Company*, 3 January 2014. <http://www.fastcolabs.com/3024378/open-company/armchair-astronomers-help-discover-why-galaxies-stop-producing-stars>.
- Cates, Jonathan. 'Discovery - Finding More Archives'. *The National Archives Blog*, 18 September 2014. <http://blog.nationalarchives.gov.uk/blog/discovery-finding-archives/>.
- Causer, Tim. 'Progress Update, 9 to 15 May 2015'. *UCL Transcribe Bentham*, 15 May 2015. <https://blogs.ucl.ac.uk/transcribe-bentham/2015/05/15/progress-update-9-to-15-may-2015/>.
- Cheng, Justin, Jaime Teevan, Shamsi T. Iqbal, and Michael S. Bernstein. 'Break It Down: A Comparison of Macro- and Microtasks', 4061–64. Seoul, Republic of Korea: ACM Press, 2015. doi:10.1145/2702123.2702146.
- Chionh, Claudine. Founders and Survivors staff member interview. Interview by Mia Ridge, 13 April 2012.
- Christian, Jon. 'Deleting the Family Tree'. *Slate*, 23 April 2015. http://www.slate.com/articles/technology/future_tense/2015/04/myfamily_shuttered_ancestry_com_deleted_10_years_of_my_family_history.html.
- Civil War Pathways. 'About'. *Civil War Pathways*. Accessed 27 March 2014. <http://web.archive.org/web/20130515144248/http://pathways.omeka.net/about>.
- . Survey: Crowdsourcing and public participation in digital history. Interview by Mia Ridge, June 2014.
- Clow, Doug, and Elpida Makriyannis. 'ISpot Analysed: Participatory Learning and Reputation'. In *Proceedings of the 1st International Conference on Learning Analytics and Knowledge*, 34–43. Banff, Alberta, Canada, 2011.
- Collings, Rachel. 'The Art of Computer Image Recognition'. *The Public Catalogue Foundation*, 5 May 2014. http://www.thepcf.org.uk/what_we_do/48/reference/862.
- Cornwall Online Parish Clerks. 'About Us'. *Cornwall Online Parish Clerks*. Accessed 14 May 2015. http://www.cornwall-opc.org/Structure/about_us.php.
- 'Cousin Bait'. *WikiTree*. Accessed 23 April 2014. http://www.wikitree.com/wiki/Cousin_bait.
- CrossyNZ. 'CrossyNZ Comments on The Panel of Historians Thread VII. Apply for Flair Here!' *Reddit AskHistorians*, March 2014. http://www.reddit.com/r/AskHistorians/comments/1yhjnh/the_panel_of_historians_thread_vii_apply_for/cfkwurh.
- @Crowdcrafting. 'Easter... a Wonderful Time ...' *Twitter*, 17 April 2014. <https://twitter.com/crowdcrafting/status/456887826563751936>.
- Crowston, Kevin, and Isabelle Fagnot. 'The Motivational Arc of Massive Virtual Collaboration', 1–35, 2013.
- Davis, Ellen. 'Collaborative Transcription Project', 9 February 2012. <http://www.southwestern.edu/live/news/6475-collaborative-transcription-project>.
- Davis, Jessie. 'Stepping Stones of Genealogy'. *FamilySearch Blog*, 20 November 2012. <https://familysearch.org/blog/en/stepping-stones-genealogy/>.
- Dawson, Gowan. 'Crowds and Clouds Workshop, Chicago'. *Constructing Scientific Communities*, 26 June 2014. <http://conscicom.org/2014/06/26/crowds-and-clouds-workshop-chicago/>.
- De Benetti, Tommaso. 'The Secrets of Digitalkoot: Lessons Learned Crowdsourcing Data Entry to 50,000 People (for Free)'. *Microtask*, 16 June 2011. <http://blog.microtask.com/2011/06/the-secrets-of-digitalkoot-lessons-learned-crowdsourcing-data-entry-to-50000-people-for-free/>.
- de Boer, Victor, Michiel Hildebrand, Lora Aroyo, Pieter De Leenheer, Chris Dijkshoorn, Binyam Tesfa, and Guus Schreiber. 'Nichesourcing: Harnessing the Power of

- Crowds of Experts'. In *Proceedings of the 18th International Conference on Knowledge Engineering and Knowledge Management, EKAW 2012*, 16–20. Springer, 2012. http://dx.doi.org/10.1007/978-3-642-33876-2_3.
- 'Digitalkoot: Crowdsourcing Finnish Cultural Heritage'. Accessed 9 January 2012. <http://blog.microtask.com/2011/02/digitalkoot-crowdsourcing-finnish-cultural-heritage/>.
- DigitalNZ. 'Selecting for Digitisation'. *Make It Digital*, 3 June 2009. <http://www.digitalnz.org/make-it-digital/selecting-for-digitisation>.
- Dijkshoorn, Chris, Mieke H. R. Leyssen, Archana Nottamkandath, Jasper Oosterman, Myriam C. Traub, Lora Aroyo, Alessandro Bozzon, et al. 'Personalized Nichesourcing: Acquisition of Qualitative Annotations from Niche Communities'. In *Proc. 6th Workshop on Personalized Access to Cultural Heritage - PATCH'13*. Rome, 2013. <http://www.few.vu.nl/~sna210/pubs/paper4.pdf>.
- Dillon-Scott, Piers. 'How Europeana, Crowdsourcing & Wiki Principles Are Preserving European History'. *The Sociable*, 31 March 2011. <http://sociable.co/business/how-europeana-crowdsourcing-wiki-principles-are-preserving-european-history/>.
- DiMeo, Michelle. 'First Monday Library Chat: University of Iowa's DIY History'. *The Recipes Project*, 3 February 2014. <http://recipes.hypotheses.org/3216>.
- Distributed Proofreaders. 'DP: Welcome'. *Distributed Proofreaders*. Accessed 27 July 2014. <http://www.pgdp.net/c/>.
- . 'Smooth Reading Pool Preview'. *Distributed Proofreaders*, 13 May 2014. http://www.pgdp.net/c/tools/post_proofers/smooth_reading.php.
- 'DIY History Code'. *The University of Iowa Libraries*, 2012. <http://web.archive.org/web/20130218122809/http://diyhistory.lib.uiowa.edu/code.html>.
- Dolph Briscoe Center for American History. 'Frequently Asked Questions'. *Dolph Briscoe Center for American History*. Accessed 28 February 2015. <http://www.cah.utexas.edu/projects/transcribe/faq>.
- Drew, John. 'About Us'. *Dickens Journals Online*. Accessed 14 February 2015. <http://www.djo.org.uk/about-us.html>.
- . 'The Online Text Correction (OTC) Project'. *Dickens Journals Online*. Accessed 14 February 2015. <http://web.archive.org/web/20130426050435/http://www.djo.org.uk/about-us/community-projects/online-text-correction.html>.
- Dunn, Stuart, and Mark Hedges. 'Crowd-Sourcing Scoping Study: Engaging the Crowd with Humanities Research'. London, U.K.: King's College, 2012. <http://www.humanitiescrowds.org>.
- Dymond, Robert. 'Community Heritage Digital Archive: Functional Description', n.d. 'Early Modern Manuscripts Online (EMMO) - Folgerpedia', July 2014. [http://folgerpedia.folger.edu/Early_Modern_Manuscripts_Online_\(EMMO\)](http://folgerpedia.folger.edu/Early_Modern_Manuscripts_Online_(EMMO)).
- Eng, Karen. 'You Found a Planet!: Robert Simpson Crowdsources Scientific Research'. *TED Blog*, 18 April 2014. <http://blog.ted.com/2014/04/18/you-found-a-planet-robert-simpson-on-zooniverse/>.
- Eveleigh, Alexandra, Charlene Jennett, Ann Blandford, Philip Brohan, and Anna L. Cox. 'Designing for Dabblers and Deterring Drop-Outs in Citizen Science', 2985–94. ACM Press, 2014. doi:10.1145/2556288.2557262.
- Eveleigh, Alexandra, Charlene Jennett, Stuart Lynn, and Anna L. Cox. 'I Want to Be a Captain! I Want to Be a Captain!: Gamification in the Old Weather Citizen Science Project'. In *Proceedings of the First International Conference on Gameful Design, Research, and Applications*, 79–82. ACM, 2013. <http://dl.acm.org/citation.cfm?id=2583019>.
- FACHRS. 'Projects'. *Family and Community Historical Research Society*. Accessed 10 June

2014. <http://www.fachrs.com/projects.html>.
- . 'The Swing Riots Project'. *Family and Community Historical Research Society*. Accessed 22 January 2015. <http://www.fachrs.com/Information/Swing.html>.
- 'Family History-Ancestry or Genealogy'. *Mormon.Org*. Accessed 12 February 2015. <http://www.mormon.org/values/family-history>.
- FamilySearch Wiki. 'Wiltshire Online Parish Clerks'. *FamilySearch.Org*, 2 January 2011. http://familysearch.org/learn/wiki/en/Wiltshire_Online_Parish_Clerks.
- 'FamilySearch Wiki:Introduction'. *FamilySearch.Org*, 28 March 2014. http://familysearch.org/learn/wiki/en/The_Vision:_Why_We_Built_FamilySearch_Wiki.
- 'FAQ'. *British Newspaper Archive*. Accessed 1 March 2015. <http://www.britishnewspaperarchive.co.uk/help-faq/what-should-i-do-if-i-notice-a-mistake-in-the-ocr-text#1>.
- 'FAQ: Why No Personal Messaging or Closed Groups?' *ISpot*. Accessed 13 June 2014. <http://www.ispotnature.org/faq-personal>.
- Fechter, Anna. 'Let's Start 2015 Off With a Bang'. *Ancestry World Archives Project News*, 2 January 2015. <http://blogs.ancestry.com/worldarchivesproject/?p=2322>.
- 'Feedback Request: Cardamine Flexuosa (21386)'. Accessed 22 June 2014. <http://herbariaunited.org/forum/viewtopic.php?t=2613>.
- Ferne, Tristan. 'Algorithms and Crowd-Sourcing for Digital Archives'. *Create Hub*, 29 May 2014. <http://www.create-hub.com/comment/algorithms-and-crowd-sourcing-for-digital-archives/>.
- 'File:IrelandFamineRelief List Key 1.JPG'. *Ancestry.Com Wiki*, 3 December 2014. http://www.ancestry.com/wiki/index.php?title=File:IrelandFamineRelief_List_Key_1.JPG.
- Fisher, Joy, Linda Lewis, Barry McGhee, Billie R. McNamara, Kim Harris Myers, John Rigdon, and George Waller. 'Overview of the USGenWeb Project', 10 April 1996. <http://usgenweb.org/about/whitepaper.shtml>.
- Fitzpatrick, Kathleen. 'Evolving Standards and Practices in Tenure and Promotion Reviews'. *Planned Obsolescence*, 11 February 2014. <http://www.plannedobsolescence.net/blog/evolving-standards-and-practices-in-tenure-and-promotion-reviews/>.
- FLGenWeb. 'Adopt a County'. *FLGenWeb*. Accessed 26 February 2015. <http://www.flgenweb.org/volunteers/adopt>.
- Flinders, Scott. 'The Future of Indexing'. *FamilySearch Blog*, 1 November 2013. <https://familysearch.org/blog/en/future-indexing/>.
- Fogg, B.J. 'A Behavior Model for Persuasive Design'. In *Proceedings of the 4th International Conference on Persuasive Technology - Persuasive '09*. New York, New York, USA: ACM Press, 2009. doi:10.1145/1541948.1541999.
- Forum poster. 'Civil War Pathways - 3-Way Brain Drain'. *Civil War Pathways*, 24 April 2013. <http://civil-war-pathways.12348.n7.nabble.com/3-Way-Brain-Drain-td60.html>.
- . 'Image AWD0002keb'. *Operation War Diary Talk*, May 2014. <http://talk.operationwardiary.org/#/subjects/AWD0002keb>.
- . 'Re: Peas Project'. *Galaxy Zoo Forum*, 5 November 2008. <http://www.galaxyzooforum.org/index.php?topic=270633.msg218401#msg218401>.
- . 'Re: Wiltshire Online Parish Clerks'. *Rootschat*, 6 April 2012. <http://www.rootschat.com/forum/index.php?topic=303500.msg4416364#msg4416364>.
- Forum posters. '2nd Batch of Unit War Diaries Added to Operation War Diary Today!' *Talk Operation War Diary History Board / Latest News*, March 2014. <http://talk.operationwardiary.org/#/boards/BWD0000003/discussions/DWD00006gw>.

- . '60,000 Pages Tagged in 2 Days! You Are Making History – THANK YOU.' *Talk Operation War Diary History Board*, January 2014.
<http://talk.operationwardiary.org/#/boards/BWD0000003/discussions/DWD0000opi>.
- . 'Algernon_Asimov Comments on [Meta]look Users, Eternal September, Rules, Moderators, and a Million Other Things.' *Reddit AskHistorians*, 2013.
http://www.reddit.com/r/AskHistorians/comments/18sbnb/meta1ook_users_eternal_september_rules_moderators/c8hxx6r.
- . 'Ask for Expert Advice'. *Old Weather Forum*, 2013.
<http://forum.oldweather.org/index.php?action=printpage;topic=3944.o>.
- . 'Biographies'. Accessed 22 June 2014.
<http://herbariaunited.org/core/forum/viewtopic.php?t=917>.
- . 'Casulties Tag - Room For Improvment'. *Talk Operation War Diary Help Board / Suggestings for Improvements*, 2014.
<http://talk.operationwardiary.org/#/boards/BWD0000008/discussions/DWD00006eo>.
- . 'Civil War Pathways - Re: A Plug for the Metadata Thesaurus!' Accessed 27 March 2014. <http://civil-war-pathways.12348.n7.nabble.com/Re-A-plug-for-the-Metadata-Thesaurus-td84.html>.
- . 'Close the Site down. It Seeks to Commercialize Remembrance and Is Poorly Designed. – Member Feedback for Lives of the First World War'. Accessed 25 May 2014. <http://support.livesofthefirstworldwar.org/forums/246136-feedback/suggestions/5926545-close-the-site-down-it-seeks-to-commercialize-re>.
- . 'Doncaster & Huntingdon Collector'. *Herbaria@home*, 2010.
<http://herbariaunited.org/core/forum/viewtopic.php?t=5627>.
- . 'Ease of Use and Linkage to Commercial Organisation – Member Feedback for Lives of the First World War'. Accessed 25 May 2014.
<http://support.livesofthefirstworldwar.org/forums/246136-feedback/suggestions/5917863-ease-of-use-and-linkage-to-commercial-organisation>.
- . 'Epidemic'. *Talk Operation War Diary*, 2014.
http://talk.operationwardiary.org/#/boards/BWD000000b/discussions/DWD00002ef?page=1&comment_id=52e80baef1e0413040001dc.
- . '[FEATURE REQUEST] Ability to Go Back and Change Tags'. *Talk Operation War Diary Help Board*, 2014.
<http://talk.operationwardiary.org/#/boards/BWD0000008/discussions/DWD0000ok3>.
- . 'Give Peas a Chance!' *Galaxy Zoo Forum*, 2007.
<http://www.galaxyzooforum.org/index.php?topic=3638.o>.
- . 'How Do I Use Hashtags?' *Talk Operation War Diary Help Board / FAQs - Tags and Tagging*, 2014.
<http://talk.operationwardiary.org/#/boards/BWD000000f/discussions/DWD000034v>.
- . 'In Person Meetings - Update'. *Civil War Pathways*, 4 October 2013. <http://civil-war-pathways.12348.n7.nabble.com/update-td503.html>.
- . 'Jargon Guide?' *Talk Operation War Diary History Board / First World War Jargon in the Diaries*, January 2014.
<http://talk.operationwardiary.org/#/boards/BWD000000e/discussions/DWD0000ogs>.
- . 'Not Impressed. This Is a Commercial Site Not a National Archive. Very Frustrated. Not How It Has Been Advertised. – Member Feedback for Lives of the First World War'. Accessed 25 May 2014.

- <http://support.livesofthefirstworldwar.org/forums/246136-feedback/suggestions/5918130-not-impressed-this-is-a-commercial-site-not-a-nat>.
- . 'Quench Project: A Proposal Aimed at Reviving and Completing It'. *Talk Galaxy Zoo Quench*, 2014.
<http://quenchtalk.galaxyzoo.org/#/boards/BGSooooooe/discussions/DGSoooo22f>.
- . 'Report Transcription'. *Talk Operation War Diary The Mess Hall Board / The Diary Pages*, January 2014.
<http://talk.operationwardiary.org/#/boards/BWDoooooo6/discussions/DWDoooo2by>.
- . 'Reviving and Completing the Quench Project: A Proposal'. *Talk Galaxy Zoo Quench*, January 2014.
http://quenchtalk.galaxyzoo.org/#/boards/BGSooooooe/discussions/DGSoooo22g?page=1&comment_id=52e6e579f1e0404a9000876.
- . 'Searchable Database'. *Old Weather Forum*, June 2012.
<http://forum.oldweather.org/index.php?topic=2872.o>.
- . 'Signs of OW Addiction ...'. *Old Weather Forum » Shore Leave » Dockside Cafe*, present 2010. <http://forum.oldweather.org/index.php?topic=1432.o>.
- . 'The British Public Has Already Paid to Collect the Information Concerning Our Armed Forces Who Fought in the First World War.' *Member Feedback for Lives of the First World War*. Accessed 25 May 2014.
<http://support.livesofthefirstworldwar.org/forums/246136-feedback/suggestions/5920281-the-british-public-has-already-paid-to-collect-the>.
- . 'The Unbearable Arbitrariness of Serendipitous Galaxy Zoo Discoveries?' *Galaxy Zoo Forum*, 2014. <http://www.galaxyzooforum.org/index.php?topic=281590.o>.
- . 'What Could We Improve?' *Talk Operation War Diary Help Board / Suggestions for Improvements*, January 2014.
<http://talk.operationwardiary.org/#/boards/BWDoooooo8/discussions/DWDoooo004>.
- . 'Where Have All The ... ?' *Talk Galaxy Zoo Quench*, 2013.
<http://quenchtalk.galaxyzoo.org/#/boards/BGSoooooo2/discussions/DGSoooo1yp>.
- . 'Why Can't I Tag Places / Activities / Other Units / Etc in Orders Pages? [Because We Are Interested in What Actually Happened]'. *Talk Operation War Diary History Board / Help Board / FAQ - Tags and Tagging*. Accessed 6 May 2014.
<http://talk.operationwardiary.org/#/boards/BWDooooooof/discussions/DWDoooo0qd>.
- . 'Why Have to Pay Again? – Member Feedback for Lives of the First World War'. Accessed 25 May 2014. <http://support.livesofthefirstworldwar.org/forums/246136-feedback/suggestions/5956766-why-have-to-pay-again>.
- . 'Why Is There No Back Button'. *Talk Operation War Diary Help Board / I Need Some Help!*, January 2014.
<http://talk.operationwardiary.org/#/boards/BWDooooood/discussions/DWDoooo03l>.
- . 'Zooniverse News'. *Home Port: Welcome to Old Weather => Old Weather News*, to present 2013.
<http://forum.oldweather.org/index.php?action=printpage;topic=4013.o>.
- Founders and Survivors. 'About the Project'. *Founders and Survivors*. Accessed 27 May 2014. <http://www.foundersandsurvivors.org/project>.
- . 'Ships Project News'. *Founders and Survivors*, 22 October 2012.
<http://foundersandsurvivors.org/content/ships-project-news>.
- Frankle, Elissa. 'Making History with the Masses: Citizen History and Radical Trust in Museums', 2013. <http://mith.umd.edu/podcasts/making-history-with-the-masses-citizen-history-and-radical-trust-in-museums/>.

- . 'More Crowdsourced Scholarship: Citizen History'. *Center for the Future of Museums*, 28 July 2011. <http://futureofmuseums.blogspot.co.uk/2011/07/more-crowdsourced-scholarship-citizen.html>.
- Frankle, Elissa, Sarah Hyams, David Klevan, and Michael Haley Goldman. Children of the Lodz Ghetto project staff members interview. Interview by Mia Ridge, 18 September 2014.
- 'Free Family Tree, Genealogy and Family History'. *MyHeritage*. Accessed 7 April 2015. <http://www.myheritage.com/>.
- 'FreeBMD Help: Welcome to the FreeBMD Signup Process'. *FreeBMD*. Accessed 23 March 2014. <http://freebmd.org.uk/Signup.html>.
- FreeBMD Newsletter. 'RootsWeb: FREEBMD-SYNDICATES-L FreeBMD Newsletter - February 2014'. *FREEBMD-SYNDICATES-L Archives*, 8 March 2014. <http://archiver.rootsweb.ancestry.com/th/read/FREEBMD-SYNDICATES/2014-03/1394314857>.
- Fugelstad, Paul, Patrick Dwyer, Jennifer Filson Moses, John S. Kim, Cleila Anna Mannino, Loren Terveen, and Mark Snyder. 'What Makes Users Rate (Share, Tag, Edit...)? Predicting Patterns of Participation in Online Communities'. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, 969–978. Seattle, 2012. <http://dl.acm.org/citation.cfm?id=2145349>.
- GEDCOM X. 'GEDCOM X'. *GEDCOM X*. Accessed 8 June 2015. <http://www.gedcomx.org/>.
- Geiger, R. Stuart, and David Ribes. 'Trace Ethnography: Following Coordination through Documentary Practices'. In *System Sciences (HICSS), 2011 44th Hawaii International Conference On*, 1–10. IEEE, 2011. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5718606.
- 'Getting Started'. *Transcribe Bentham: Transcription Desk*, 16 July 2013. http://www.transcribe-bentham.da.ulcc.ac.uk/td/Getting_Started.
- Gilliver, Peter. "Your Dictionary Needs You": A Brief History of the OED's Appeals to the Public'. *Oxford English Dictionary*, 4 October 2012. <http://public.oed.com/the-oed-appeals/history-of-the-appeals/>.
- Ginther, James. 'T-PEN Version 2.0'. *The Center for Digital Theology at Saint Louis University*, 5 May 2012. <http://digital-editor.blogspot.co.uk/2012/05/t-pen-version-20.html>.
- Greenstreet, Colin. 'Communicating MarineLives'. *The Shipping News*, 24 November 2013. <http://marinelives-theshippingnews.org/blog/2013/11/24/communicating-marinelives/>.
- . 'MarineLives - Call for Project Volunteers'. *Academia.Edu*, 2012. http://marinelives.academia.edu/ColinGreenstreet/Posts/351069/MarineLives_-_Call_for_Project_Volunteers.
- . 'MarineLives - Call for Project Volunteers'. *Institute of Historical Research*, 12 July 2012. <http://www.history.ac.uk/news/2012-08-31/marinelives-call-project-volunteers>.
- . 'On the Crest of a Wave'. *History Today*. Accessed 24 May 2014. <http://www.historytoday.com/colin-greenstreet/crest-wave>.
- . 'Patterns in Time'. *The Shipping News*, 14 May 2014. <http://marinelives-theshippingnews.org/blog/2014/05/15/patterns-in-time/>.
- . 'Progress Review Two'. *The Shipping News*, 21 January 2013. <http://marinelives-theshippingnews.org/blog/2013/01/21/progress-review-two/>.
- . 'Summer Programme 2014'. *The Shipping News*, 7 May 2014. <http://marinelives-theshippingnews.org/blog/2014/05/07/summer-programme-2014/>.
- Greenstreet, Colin, and Jill Wilcox. 'C17th London as Seen in the Archives of the High Court of Admiralty'. presented at the Docklands History Group, Museum of London Docklands, 3 September 2014.

- https://www.academia.edu/8188588/Text_and_Thumbnail_slides_Dockland_History_Group_Talk_London_Dockland_Museum_September_3rd_2014.
- Grey, Francois. 'Upscience: Inverting the Pyramid of Scientific Enquiry'. *Billion Brain Blog*, 11 May 2014. <http://www.billionbrainblog.com/?p=237>.
- 'GuardianWitness'. Accessed 2 March 2015. <https://witness.theguardian.com/moreabout>.
- Guest Blogger. 'Indexing and Perseverance'. *FamilySearch Blog*, 5 November 2013. <https://familysearch.org/blog/en/indexing-perseverance-gp/>.
- Guralnick, Rob. 'Making Progress Clear on Notes from Nature'. *Notes from Nature*, 24 February 2014. <http://blog.notesfromnature.org/2014/02/24/making-progress-clear-on-notes-from-nature/>.
- Haklay, Muki. 'European Citizen Science Association Suggestion for 10 Principles of Citizen Science'. *Po Ve Sham - Muki Haklay's Personal Blog*, 14 May 2014. <http://povesham.wordpress.com/2014/05/14/european-citizen-science-association-suggestion-for-10-principles-of-citizen-science/>.
- 'Hall of Fame'. *UCL Transcribe Bentham*, 22 July 2014. <http://blogs.ucl.ac.uk/transcribe-bentham/hall-of-fame/>.
- Hansen, Derek L., Patrick J. Schone, Doug Corey, Matthew Reid, and Jake Gehring. 'Quality Control Mechanisms for Crowdsourcing: Peer Review, Arbitration, & Expertise at FamilySearch Indexing'. In *Computer Supported Cooperative Work and Social Computing (CSCW)*. San Antonio: ACM, 2013.
- Hart, Michael. 'The History and Philosophy of Project Gutenberg'. *Project Gutenberg*, August 1992. http://www.gutenberg.org/wiki/Gutenberg:The_History_and_Philosophy_of_Project_Gutenberg_by_Michael_Hart.
- 'Has Galaxy Zoo Research Become Manufacturing? (Why Astronomers Are so Rare Here?)'. Accessed 24 June 2014. <http://www.galaxyzooforum.org/index.php?topic=281847.0>.
- Haythornthwaite, Caroline. 'Crowds and Communities: Light and Heavyweight Models of Peer Production'. In *Proceedings of the 42nd Hawaii International Conference on System Sciences - 2009*, 1-10, 2009. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4755627.
- Heglund, Chester R. 'Comment on 2014 Indexing Year in Review'. *FamilySearch Blog*, 24 January 2015.
- Hess, Whitney. 'Onboarding: Designing Welcoming First Experiences'. *UX Magazine*, 16 February 2010. <http://uxmag.com/articles/onboarding-designing-welcoming-first-experiences>.
- Hill, Julie. 'Online Family History Trends Report'. *Archives*, 2 February 2011. <http://www.archives.com/blog/miscellaneous/online-family-history-trends-1.html>.
- 'Hints'. *Ancestry*. Accessed 7 April 2015. <http://www.ancestry.com/hints>.
- 'History Working Papers Project'. *History Working Papers Project*. Accessed 19 February 2015. <http://www.historyworkingpapers.org/>.
- 'Historypin'. *Shift*. Accessed 1 March 2015. <http://www.shiftdesign.org.uk/products/historypin/>.
- Hoe, Georgia-Lee. 'Australian History: Online and Insatiable'. In *Peer Reviewed Proceedings of the 4th Annual Conference, Popular Culture Association of Australia and New Zealand (PopCAANZ)*. Brisbane, Australia, 2013. http://popcaanz.com/conferenceproceedings_2013/History_Hoe_AussieHistoryOnline.pdf.
- Holley, Rose. 'From Little Things Big Things Grow – Gold Star to Newspaper Text Correctors for E-Books!'. *Rose Holley's Blog*, 13 December 2011. <http://rose-holley.blogspot.co.uk/2011/12/from-little-things-big-things-grow-gold.html>.
- . 'National Archives of Australia Embraces Crowdsourcing and Releases "The Hive"'. *Rose Holley's Blog*, 10 November 2012. <http://rose->

- holley.blogspot.co.uk/2012/11/national-archives-of-australia-embraces.html.
- Howard, Sharon. 'Bloody Code: Reflecting on Ten Years of the Old Bailey Online and the Digital Futures of Our Criminal Past'. *Early Modern Notes*, 18 May 2013.
<http://earlymodernnotes.wordpress.com/2013/05/18/bloody-code-reflecting-on-ten-years-of-the-old-bailey-online-and-the-digital-futures-of-our-criminal-past/>.
- . London Lives staff member interview. Interview by Mia Ridge, 15 April 2014.
- . 'London Lives Wiki'. *London Lives News*, 5 October 2010.
http://www.hrionline.ac.uk/londonliveswiki/tiki-view_blog_post.php?postId=2.
- Howe, Jeff. 'Crowdsourcing: A Definition', 2 June 2006.
http://crowdsourcing.typepad.com/cs/2006/06/crowdsourcing_a.html.
- Humphrey, Tom. 'About Feedback Requests'. *Herbariaunited.Org Forum Index -> Herbarium Sheet Discussion*, December 2007.
<http://herbariaunited.org/forum/viewtopic.php?t=679>.
- . 'About Herbaria@home'. *Herbaria@home*. Accessed 29 March 2014.
<http://herbariaunited.org/ahcontent/AHabout/>.
- . 'Linking Collectors with Message Board Discussions'. *Herbariaunited.Org Forum Index -> Botanists*, 10 March 2010.
<http://herbariaunited.org/core/forum/viewtopic.php?t=3441>.
- . 'Request for Examples of Handwriting'. *Herbariaunited.Org Forum Index -> Botanists*, 10 March 2010.
<http://herbariaunited.org/core/forum/viewtopic.php?t=3442>.
- . 'Reviewing Sheets, Editing Collectors and Localities Etc.' *Herbariaunited.Org Forum Index -> Bugs and Suggestions*, 10 January 2010.
<http://herbariaunited.org/core/forum/viewtopic.php?t=2916>.
- . 'Updated Version of the Doc. Form, -- Please Help Test It!' *Herbariaunited.Org*, May 2008. <http://herbariaunited.org/core/forum/viewtopic.php?t=1040>.
- . 'Welcome!' *Herbariaunited.Org Forum Index -> News and Announcements*, 4 June 2006. <http://herbariaunited.org/core/forum/viewtopic.php?t=6>.
- International Organization for Standardization. 'ISO 9241-210:2010(En). Ergonomics of Human-System Interaction — Part 210: Human-Centred Design for Interactive Systems'. International Organization for Standardization, 2010.
<https://www.iso.org/obp/ui/#iso:std:iso:9241:-210:ed-1:vi:en>.
- Irons Walch, Victoria. 'Where History Begins. A Report on Historical Records Repositories in the United States'. Council of State Historical Records Coordinators, May 1998.
- Jacques, Jason T., and Per Ola Kristensson. 'Crowdsourcing a HIT: Measuring Workers' Pre-Task Interactions on Microtask Markets'. In *First AAAI Conference on Human Computation and Crowdsourcing*, 2013.
<http://www.aaai.org/ocs/index.php/HCOMP/HCOMP13/paper/view/7540/o>.
- Jennett, Charlene, Alexandra Eveleigh, Kathleen Mathieu, Zoya Ajani, and Anna L. Cox. 'Creativity in Citizen Cyber-Science: All for One and One for All'. Paris, France, 2013.
- JISC. 'Virtual Research Environment Programme', 2013.
<http://webarchive.nationalarchives.gov.uk/20140702233839/http://www.jisc.ac.uk/whatwedo/programmes/vre.aspx>.
- Johnson, Steve. 'Correcting Errors in Transcriptions'. *Interment.Net*, 17 April 2009.
<http://www.interment.net/column/2009/04/correcting-errors-in-transcriptions.html>.
- Judson, Michael. 'The Mobile App—What We've Learned'. *FamilySearch Blog*, 14 December 2012. <https://familysearch.org/blog/en/mobile-app-what-we-learned/>.
- Kathy. 'Database Data Dictionary'. *Old Weather Forum*, 11 June 2012.
<http://forum.oldweather.org/index.php?topic=2873.0>.
- Keynsham & Saltford Local History Society. 'Archives'. *Keynsham & Saltford Local History*

- Society*. Accessed 13 February 2015. <http://keysalthist.org.uk/archives.htm>.
- 'Kildare Launch of the "Letters of 1916"'. *Maynooth University*, 28 April 2014. <https://www.maynoothuniversity.ie/foras-feasa/news/kildare-launch-letters-1916>.
- Kittur, Aniket, and Robert E. Kraut. 'Harnessing the Wisdom of Crowds in Wikipedia: Quality Through Coordination'. In *Computer Supported Cooperative Work (CSCW)*, 37–46. San Diego, CA, USA, 2008.
- Kittur, Aniket, Jeffrey V. Nickerson, Michael Bernstein, Elizabeth Gerber, Aaron Shaw, John Zimmerman, Matt Lease, and John Horton. 'The Future of Crowd Work'. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, 1301–1318, 2013. <http://dl.acm.org/citation.cfm?id=2441923>.
- Kloetzer, Laure, Daniel Schneider, Charlene Jennett, Ioanna Iacovides, Alexandra Eveleigh, Anna Cox, and Margaret Gold. 'Learning by Volunteer Computing, Thinking and Gaming: What and How Are Volunteers Learning by Participating in Virtual Citizen Science?' In *Changing Configurations of Adult Education in Transitional Times*, 73–93. Berlin, Germany, 2013.
- Knupfer, Peter. 'The Coming of the Commons: An Important Message from H-Net'. *H-Announce*, 19 August 2013. <http://h-net.msu.edu/cgi-bin/logbrowse.pl?trx=vx&list=H-Announce&month=1308&week=c&msg=P5u9Jgjp/raThoo1UrHp3Q&user=&pw=>.
- Lambert, Susan, Marcus Winter, and Phil Blume. 'Getting to Where We Are Now'. *10most.Org.Uk*, 26 March 2014. <http://10most.org.uk/content/getting-where-we-are-now>.
- Latimer, Jack. 'Letter in the Attic: Lessons Learnt from the Project'. *My Brighton and Hove*, 25 February 2009. <http://www.mybrightonandhove.org.uk/page/letterintheatticlesons?path=op116p1543p>.
- 'Launching BillionGraves.Com'. *BillionGraves Blog*, 25 May 2011. <http://blog.billiongraves.com/2011/05/launching-billiongraves-com/>.
- Law, Edith, and Luis von Ahn. 'Input-Agreement: A New Mechanism for Collecting Data Using Human Computation Games'. In *Proceedings of the 27th International Conference on Human Factors in Computing Systems*, 1197–1206. Boston, MA, USA, 2009.
- Leadbeater, Charles, and Paul Miller. 'The Pro-Am Revolution: How Enthusiasts Are Changing Our Economy and Society'. Demos, 2004. <http://www.demos.co.uk/publications/proameconomy>.
- Lee, Stuart. 'Woruldthord Web Site Now Launched!' *Project Woruldthord*, 14 March 2011. <http://blogs.it.ox.ac.uk/woruldthord/2011/03/14/woruldthord-web-site-now-launched/>.
- Leon, Sharon. 'Scripto Alpha Launches with the Papers of the War Department'. *Scripto*, 17 March 2011. <http://scripto.org/scripto-alpha-launches-with-the-papers-of-the-war-department/>.
- Letters of 1916. 'Progress Update, April 2015'. *Letters 1916*, April 2015. <http://dh.tcd.ie/letters1916/progress-update/april-2015/>.
- 'Library of Congress Photos on Flickr - FAQ'. *Prints and Photographs Reading Room, Library of Congress*, 22 October 2010. http://www.loc.gov/rr/print/flickr_pilot_faq.html.
- Linn, Allison. 'Picture This: Microsoft Research Project Can Interpret, Caption Photos'. *Next at Microsoft*, 28 May 2015. <http://blogs.microsoft.com/next/2015/05/28/picture-this-microsoft-research-project-can-interpret-caption-photos/>.
- Lintott, Chris. '"Citizen" Science and "Real" Science'. *Zooniverse*, 29 December 2010. <http://blog.zooniverse.org/2010/12/29/citizen-science-and-real-science/>.

- . ‘Open Zooniverse: Beta Testers Required’. *Citsci-Discussion-L*, June 2015.
- . ‘PH1: A Planet in a Four-Star System’. *Planet Hunters*, 15 October 2012.
<http://blog.planethunters.org/2012/10/15/ph1-a-planet-in-a-four-star-system/>.
- . ‘Thanks to the Forum – and Farewell’. *Galaxy Zoo*, 9 July 2014.
<http://blog.galaxyzoo.org/2014/07/09/thanks-to-the-forum-and-farewell/>.
- . ‘Why the Zooniverse Is Easy to Use.’ *Zooniverse*, 17 February 2013.
<http://blog.zooniverse.org/2013/02/17/why-the-zooniverse-is-easy-to-use/>.
- Lintott, Chris, and The Citizen Science Alliance. ‘The Two-and-a-Bit Page Guide to Running a Zooniverse Project’. The Citizen Science Alliance, n.d.
- Lives of the First World War. ‘About’. *Lives of the First World War*, 2014.
<https://livesofthefirstworldwar.org/about>.
- London Lives. ‘About This Project’. *London Lives*, April 2012.
<http://www.londonlives.org/static/Project.jsp>.
- . ‘Biographies’. *London Lives Wiki*, 2010.
<http://www.hrionline.ac.uk/londonliveswiki/tiki-index.php?page=Biographies>.
- Luczak-Roesch, Markus, Ramine Tinati, Elena Simperl, Max Van Kleek, Nigel Shadbolt, and Robert Simpson. ‘Why Won’t Aliens Talk to Us? Content and Community Dynamics in Online Citizen Science’. In *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media*, 2014.
<http://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/download/8092/8136>.
- Lynch, Bernadette. ‘Whose Cake Is It Anyway? A Collaborative Investigation into Engagement and Participation in 12 Museums and Galleries in the UK’. Paul Hamlyn Foundation, 2011. <http://www.phf.org.uk/page.asp?id=1417>.
- Maher, Jacqui. ‘Hive: Open-Source Crowdsourcing Framework’. *Nytlabs Blog*, 9 December 2014. <http://blog.nytlabs.com/2014/12/09/hive-open-source-crowdsourcing-framework/>.
- Marshall, Jack. ‘Mobile Isn’t Killing the Desktop Internet’. *CMO Today, The Wall Street Journal*, 26 May 2015. <http://blogs.wsj.com/cmo/2015/05/26/mobile-isnt-killing-the-desktop-internet/>.
- Mary. ‘Calling Citizen Archivists to Crowdsourc Video Captions!’ *NARAtions*, 26 March 2014. <http://blogs.archives.gov/online-public-access/?p=9222>.
- . ‘What’s New On Amara?’ *NARAtions*, 24 February 2015.
<http://blogs.archives.gov/online-public-access/?p=9397>.
- Matheny, Kate. ‘Come Transcribe Our Items!’ *Digital Services @ the University of Alabama*, 9 February 2015.
<http://apps.lib.ua.edu/blogs/digitalservices/2015/02/09/transcribe-our-items/>.
- McCalman, Janet, Rebecca Kippen, Sandra Silcot, and Leonard Smith. ‘Building a Life Course Dataset from Australian Convict Records; Founders & Survivors: Australian Life Courses in Historical Context, 1803-1920’. In *Proceedings Workshop Population Reconstruction*. Amsterdam, 2014.
http://socialhistory.org/sites/default/files/docs/mccalman_kippen_silcot_smith_-_building_a_life_course_dataset.pdf.
- McClurken, Jeffrey W. ‘Digital History Reviews’. *Journal of American History*. Accessed 11 February 2015.
<http://www.journalofamericanhistory.org/submit/digitalhistoryreviews.html>.
- McCredie, Athol. ‘Help Wanted: Who, What, When?’ *Te Papa’s Blog*, 3 February 2015.
<http://blog.tepapa.govt.nz/2015/02/03/help-wanted-who-what-when/>.
- . ‘More Help Wanted: Photo-Detective #2’. *Te Papa’s Blog*, 7 February 2015.
<http://blog.tepapa.govt.nz/2015/02/07/more-help-wanted-photo-detective-2/>.
- . ‘You Can Play Photo-Detective #3’. *Te Papa’s Blog*. Accessed 4 March 2015.
<http://blog.tepapa.govt.nz/2015/02/19/you-can-play-photo-detective-3/>.

- McDaniel, W. Caleb. 'Open Notebook History'. *W. Caleb McDaniel*, 22 May 2013.
<http://wcm1.web.rice.edu/open-notebook-history.html>.
- McGonigal, Jane. 'Engagement Economy: The Future of Massively Scaled Collaboration and Participation'. Palo Alto: Institute for the Future, 2008.
- MeredithMJ. 'Looking for MCN 2009 Co-Presenter, Topic: Crowdsourcing'. *MCN-L*, 9 April 2009. <https://www.mail-archive.com/mcn-l@mcn.edu/msg00162.html>.
- Miller, Grant. 'Saturday Status – Operation War Diary'. *Daily Zooniverse*, 25 January 2014.
<http://daily.zooniverse.org/2014/01/25/saturday-status-operation-war-diary/>.
- Mitra, Tanushree, and Eric Gilbert. 'The Language That Gets People to Give: Phrases That Predict Success on Kickstarter'. CSCW, 2014.
<http://comp.social.gatech.edu/papers/csw14.crowdfunding.mitra.pdf>.
- Moretti, Franco. 'Changes'. *Public Books*, 1 March 2014.
<http://www.publicbooks.org/briefs/changes>.
- Muller, Michael J, and Sandra Kogan. 'Grounded Theory Method in HCI and CSCW'. Cambridge, MA: IBM Center for Social Software, 2010.
- Museums, Libraries and Archives Council. 'Generic Learning Outcomes'. *Inspiring Learning*, 2008.
<http://www.inspiringlearningforall.gov.uk/toolstemplates/genericlearning/>.
- National Archives of Australia. 'ArchIVE – Homepage'. *ArchIVE*. Accessed 18 June 2014.
<http://transcribe.naa.gov.au/>.
- Nauta, Paul G. 'Massive Online US Obituaries Project Will Help Find Your Ancestors'. *FamilySearch Blog*, 1 October 2014. <https://familysearch.org/blog/en/massive-online-obituaries-project-find-ancestors/>.
- 'NEH Grant Details: Resurrecting Early Christian Lives: Digging in Papyri in a Digital Age'. *National Endowment for the Humanities*, 12 February 2014.
<https://securegrants.neh.gov/publicquery/main.aspx?f=1&gn=HJ-50178-14>.
- Ngatuvai, Spencer. '2014 Indexing Year in Review'. *FamilySearch Blog*, 2 January 2015.
<https://familysearch.org/blog/en/2014-indexing-year-review/>.
- Nielsen, Jakob. '10 Usability Heuristics for User Interface Design', 1995.
<http://www.nngroup.com/articles/ten-usability-heuristics/>.
- . 'Finding Usability Problems through Heuristic Evaluation'. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 373–380. ACM, 1992.
<http://dl.acm.org/citation.cfm?id=142834>.
- . 'Guerrilla HCI: Using Discount Usability Engineering to Penetrate the Intimidation Barrier'. *Nielsen Norman Group*, 1994.
http://www.useit.com/papers/guerrilla_hci.html.
- . 'How to Conduct a Heuristic Evaluation'. *Nielsen Norman Group*, January 1995.
<http://www.nngroup.com/articles/how-to-conduct-a-heuristic-evaluation/>.
- Nielsen, Jakob, and Rolf Molich. 'Heuristic Evaluation of User Interfaces'. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 249–256. CHI '90. New York, NY, USA: ACM, 1990. doi:10.1145/97243.97281.
- Noordegraaf, Julia, Angela Bartholomew, and Alexandra Eveleigh. 'Modeling Crowdsourcing for Cultural Heritage'. In *Museums and the Web 2014*, edited by Nancy Proctor and Rich Cherry. Museums and the Web, 2014.
<http://mw2014.museumsandtheweb.com/paper/modeling-crowdsourcing-for-cultural-heritage/>.
- Norman, Donald A., and Andrew Ortony. 'Designers and Users: Two Perspectives on Emotion and Design'. In *Proc. of the Symposium on Foundations of Interaction Design at the Interaction Design Institute*. Ivrea, Italy, 2003.
- Nov, Oded, Ofer Arazy, and David Anderson. 'Technology-Mediated Citizen Science Participation: A Motivational Model'. In *Proceedings of the AAAI International Conference on Weblogs and Social Media*. Barcelona, Spain, 2011.

- nQuire It. 'About NQuire It'. *NQuire It*. Accessed 23 February 2015. <http://www.nquire-it.org/#/about>.
- O'Donnell, Jim. Operation War Diary stats. Interview by Mia Ridge, 29 June 2014.
- Old Weather Moderators, Editors and Volunteers. 'Compendium of Maritime and Weather Information'. Accessed 25 June 2014. <http://www.naval-history.net/OWShips-ForumMat.htm>.
- Oomen, Johan, and Lora Aroyo. 'Crowdsourcing in the Cultural Heritage Domain: Opportunities and Challenges'. In *Proceedings of the 5th International Conference on Communities and Technologies*, 138–149, 2011. <http://www.cs.vu.nl/~marieke/OomenAroyoCT2011.pdf>.
- Oomen, Johan, Maarten Brinkerink, Lieke Heijmans, and Thijs van Exel. 'Emerging Institutional Practices: Reflections on Crowdsourcing and Collaborative Storytelling'. In *Museums and the Web 2010: Proceedings*, edited by Jennifer Trant and David Bearman. Archives & Museum Informatics, 2010. <http://www.museumsandtheweb.com/mw2010/papers/oomen/oomen.html>.
- Operation War Diary. 'About Operation War Diary'. *Operation War Diary*, 2014. <http://www.operationwardiary.org/#/about>.
- . 'Operation War Diary'. *Operation War Diary*. Accessed 6 June 2015. <http://www.operationwardiary.org/>.
- O'Reilly, Tim. 'What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software'. *O'Reilly*, 30 September 2005. <http://www.oreilly.com/pub/a/web2/archive/what-is-web-2.0.html>.
- OSA Archivum. 'About Us'. *Open Society Archives Archivum*. Accessed 6 June 2015. <http://www.osaarchivum.org/about-us>.
- 'Ownership and Control'. *WikiTree*, 18 August 2014. http://www.wikitree.com/wiki/Ownership_and_control.
- Palmer, Carole L., Lauren C. Tefreau, and Carrie M. Pirmann. 'Scholarly Information Practices in the Online Environment: Themes from the Literature and Implications for Library Service Development'. OCLC Research, 2009. www.oclc.org/programs/publications/reports/2009-02.pdf.
- Papers of the War Department. 'Become a Transcription Associate'. *Papers of the War Department*. Accessed 29 May 2015. <http://wardepartmentpapers.org/transcribe.php>.
- 'Planet Hunters. VI. An Independent Characterization of KOI-351 and Several Long'. Accessed 4 September 2014. <http://adsabs.harvard.edu/abs/2013arXiv1310.5912S>.
- 'Post-Quench Galaxies: Literature Review'. Accessed 23 June 2014. <http://postquench.blogspot.co.uk/>.
- Powell, Kimberly. 'World Archives Project vs. FamilySearch Indexing'. *About.Com Genealogy*, 8 September 2008. <http://genealogy.about.com/b/2008/09/09/world-archives-project-vs-familysearch-indexing.htm>.
- Preist, Chris, Elaine Massung, and David Coyle. 'Competing or Aiming to Be Average?: Normification as a Means of Engaging Digital Volunteers'. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 1222–33. Baltimore, MD, USA.: ACM Press, 2014. doi:10.1145/2531602.2531615.
- Prescott, Andrew. 'Dennis the Paywall Menace Stalks the Archives'. *Digital Riffs*, 2 February 2014. <http://digitalriffs.blogspot.co.uk/2014/02/dennis-paywall-menace-stalks-archives.html>.
- Prestopnik, Nathan R., and Kevin Crowston. 'Citizen Science System Assemblages: Understanding the Technologies That Support Crowdsourced Science', 7–10. Toronto, Canada, 2012.
- 'Project Description'. *Wir Waren so Frei*. Accessed 15 February 2015. <https://www.wir->

- waren-so-frei.de/index.php/About/Index.
- Quinn, Alexander J., and Benjamin B. Bederson. 'Human Computation: A Survey and Taxonomy of a Growing Field'. In *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems*, 1403–1412, 2011.
<http://dl.acm.org/citation.cfm?id=1979148>.
- Raddick, M. Jordan, Georgia Bracey, Karen Carney, Geza Gyuk, Kirk Borne, John Wallin, and Suzanne Jacoby. 'Citizen Science: Status and Research Directions for the Coming Decade'. In *Astro2010: The Astronomy and Astrophysics Decadal Survey*, Vol. 2010, 2009.
<http://www8.nationalacademies.org/astro2010/DetailFileDisplay.aspx?id=454>.
- rainydaygoods. 'Ravelers Rocked the Search Party!' *Unraveled*, 20 July 2010.
<http://blog.ravelry.com/2010/07/20/ravelers-rocked-the-search-party/>.
- Randell, Brian, and Louis R. Mills. 'Getting Started in Genealogy and Family History'. *GENUKI*, December 2013. <http://www.genuki.org.uk/gs/>.
- Randi. 'How to Look at Log Pages before and after the One You Are Working On'. *Old Weather Forum » Questions and Answers » The Logs and FAQ*, 12 February 2013.
<http://forum.oldweather.org/index.php?topic=3598.0>.
- Ravelry. 'About Our Site'. *Ravelry*. Accessed 6 June 2015. <http://www.ravelry.com/about>.
- Raymond, Allan. 'RootsWeb: FreeBMD-Admins-L FreeBMD *** 300 Million *** Records Milestone Achieved'. *FreeBMD-Admins-L Archives*, 25 February 2014.
<http://archiver.rootsweb.ancestry.com/th/read/FreeBMD-Admins/2014-02/1393290478>.
- Reed, Jason, Whitney Rodriguez, and Angelique Rickhoff. 'A Framework for Defining and Describing Key Design Features of Virtual Citizen Science Projects'. In *Proceedings of the 2012 IConference*, 623–625. ACM, 2012.
<http://dl.acm.org/citation.cfm?id=2132314>.
- 'Reputation on ISpot'. *ISpot*. Accessed 13 June 2014. <http://www.ispotnature.org/help-reputation>.
- 'Research'. *The UCL Space Syntax Laboratory*. Accessed 17 April 2015.
<http://www.bartlett.ucl.ac.uk/space-syntax/research>.
- Research Councils UK. 'RCUK Policy on Open Access'. *Research Councils UK*. Accessed 31 May 2015. <http://www.rcuk.ac.uk/research/openaccess/policy/>.
- Reside, Doug. 'Crowdsourcing Performing Arts History with NYPL's ENSEMBLE', 2014.
<http://dharchive.org/paper/DH2014/Paper-131.xml>.
- Ridge, Mia. 'Creating a Digital History Commons through Crowdsourcing and Participant Digitisation'. Herrenhausen Palace, Hanover, Germany, 2013.
<http://www.miaridge.com/herrenhausen/>.
- Rogers, Katina. '#alt-Ac in Context'. *Media Commons*. Accessed 3 June 2015.
<http://mediacommons.futureofthebook.org/alt-ac/alt-ac-context>.
- Romeo, Fiona, and Lucinda Blaser. 'Bringing Citizen Scientists and Historians Together'. In *Museums and the Web 2011: Proceedings*, edited by Jennifer Trant and David Bearman. Toronto, Canada: Archives & Museum Informatics, 2011.
http://www.museumsandtheweb.com/mw2011/papers/bringing_citizen_scientists_and_historians_tog.
- Rose-Sandler, Trish. 'Citizen Science Uses Art to Unlock Scientific Knowledge'. *Biodiversity Heritage Library*, 31 March 2015. <http://blog.biodiversitylibrary.org/2015/03/citizen-science-uses-art-to-unlock.html>.
- . 'Crowdsourcing Your Cultural Heritage Collections: Considerations When Choosing a Platform'. presented at the Visual Resources Association conference, Denver, 13 March 2015. <http://www.slideshare.net/trosesandler/crowdsourcing-your-metadata>.
- Ross, Jim. 'Why Are Contextual Inquiries So Difficult?' *UXmatters*, 4 June 2012.

- <http://www.uxmatters.com/mt/archives/2012/06/why-are-contextual-inquiries-so-difficult.php>.
- Rotman, Dana, Jennifer Preece, Jennifer Hammock, Kezee Procita, Derek Hansen, Cynthia Parr, Darcy Lewis, and David Jacobs. 'Dynamic Changes in Motivation in Collaborative Citizen-Science Projects'. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, 217–226. Seattle, 2012. doi:10.1145/2145204.2145238.
- Royal Historical Society. 'Membership - Notes for Applicants'. *Royal Historical Society*. Accessed 21 May 2015. <http://royalhistsoc.org/membership/members/membership-notes-for-applicants/>.
- 'RunCoCo: How to Run a Community Collection Online'. Oxford: Oxford University Computing Services (OUCS), 2011. <http://projects.oucs.ox.ac.uk/runcoco/>.
- Rutner, Jennifer, and Roger C Schonfeld. 'Supporting the Changing Research Practices of Historians'. Ithaka S+R, 7 December 2012. <http://www.sr.ithaka.org/research-publications/supporting-changing-research-practices-historians>.
- Safire, William. 'On Language - Fat Tail'. *The Times Magazine*, 5 February 2009. <http://www.nytimes.com/2009/02/08/magazine/o8wwln-safire-t.html>.
- Sample Ward, Amy. 'Crowdsourcing vs Community-Sourcing: What's the Difference and the Opportunity?' *Amy Sample Ward's Version of NPTEch*, 18 May 2011. <http://amysampleward.org/2011/05/18/crowdsourcing-vs-community-sourcing-whats-the-difference-and-the-opportunity/>.
- Schaffner, Jennifer. 'The Metadata Is the Interface: Better Description for Better Discovery of Archives and Special Collections'. OCLC Research, May 2009. <http://www.oclc.org/programs/publications/reports/2009-06.pdf>.
- Schwamb, Megan E. 'Planet Hunters's First Circumbinary Planet- A True Team Effort'. *Planet Hunters*, 15 October 2012. <http://blog.planethunters.org/2012/10/15/planet-hunterss-first-circumbinary-planet-a-true-team-effort/>.
- Scott, Kerry, and Various commenters. 'Why Is There So Much Drama On Find A Grave?' *Clue Wagon*, 13 April 2011. <http://www.cluewagon.com/2011/04/why-is-there-so-much-drama-on-find-a-grave/>.
- Shaw, Al. 'No Windows. One Exit. Free Drinks: Casino-Driven Design for Crowdsourcing'. *ProPublica*, 20 March 2013. <http://www.propublica.org/nerds/item/casino-driven-design>.
- Sheppard, Alice. 'Peas in the Universe, Goodwill and a History of Zooite Collaboration on the Peas Project'. *Galaxy Zoo*, 7 July 2009. <http://blog.galaxyzoo.org/2009/07/07/peas-in-the-universe-goodwill-and-a-history-of-zooite-collaboration-on-the-peas-project/>.
- Shoup, Eric. 'Ancestry.Com Focuses on Core Offerings', 4 June 2014. <http://blogs.ancestry.com/ancestry/2014/06/04/ancestry-com-focuses-on-core-offerings/>.
- Shulman, Lee S. 'Taking Learning Seriously'. *Change*, 1999.
- SI Transcription Ctr. ' '.@mia_out Great Q! We Update by e-Mail & Blogpost (& New Section, Shhh) - Today's Numbers: 722 Projects Comp and/or 91,135 Pages Completed! Microblog. @TranscribeSI, 15 May 2015. <https://twitter.com/TranscribeSI/status/599234883030921216>.
- . ' '.@mia_out June 2013 Beta Then July 2014 (Though Always Iterating/Improving) #thanksforasking. Microblog. @TranscribeSI, 27 February 2015. <https://twitter.com/TranscribeSI/status/571133725406302208>.
- Simpson, Robert. 'One Million Volunteers'. *Zooniverse*, 14 February 2014. <http://blog.zooniverse.org/2014/02/14/one-million-volunteers/>.
- Sinclair, Stéfan. 'Voyeur Tools & Old Bailey'. *Criminal Intent*, 29 May 2011. <http://criminalintent.org/voyeur-tools/>.

- Smith, Arfon. 'Making the Zooniverse Open Source'. *Zooniverse*, 18 February 2013. <http://blog.zooniverse.org/2013/02/18/making-the-zooniverse-open-source/>.
- . 'Zooniverse, GitHub and the Future'. *Zooniverse*, 13 August 2013. <http://blog.zooniverse.org/2013/08/13/zooniverse-github-and-the-future/>.
- Smith, Luke. 'Inside Team War Diary...'. *Operation War Diary*, 8 April 2014. <http://blog.operationwardiary.org/2014/04/08/inside-team-war-diary/>.
- Smith Rumsey, Abby. 'Full Report: Scholarly Communication Institute 7 Report: Spatial Technologies and the Humanities'. Scholarly Communication Institute, 2009. <http://uvasci.digress.it/full-draft-report/>.
- Smithsonian National Museum of Natural History. '#FWTrueLove Transcription Challenge'. *Field Book Project*, 13 February 2015. <http://nsmnh.typepad.com/fieldbooks/2015/02/fwtruelove-transcription-challenge.html>.
- Smithsonian Transcription Center. '#MyTCdiscovery'. *Facebook*, 27 April 2015. <https://www.facebook.com/SmithsonianTranscriptionCenter/photos/a.341352889336961.1073741829.339997839472466/517754628363452/?type=1>.
- Snoek, Cees G. M., Bauke Freiburg, Johan Oomen, and Roeland Ordelman. 'Crowdsourcing Rock N' Roll Multimedia Retrieval'. In *Proceedings of the International Conference on Multimedia*, 1535–38. New York, NY, USA: ACM, 2010. <http://doi.acm.org/10.1145/1873951.1874278>.
- Springer, Michelle, Beth Dulabahn, Phil Michel, Barbara Natanson, David Reser, David Woodward, and Helena Zinkham. 'For the Common Good: The Library of Congress Flickr Pilot Project'. Library of Congress, 2008. http://www.loc.gov/rr/print/flickr_report_final.pdf.
- State Library of Queensland. 'Text Correct'. *State Library of Queensland*. Accessed 12 May 2014. <http://www.slq.qld.gov.au/about-us/pitch-in/text-correct>.
- Stokmans, Mia. 'MAO-Model of Audience Development: Some Theoretical Elaborations and Practical Consequences'. In *Proceedings on CD-Rom of the 8th International Conference on Arts and Cultural Management*, 1–11. Montréal, Québec, Canada, 2005. http://neumann.hec.ca/aimac2005/PDF_Text/Stockmans_Mia.pdf.
- Stone, Edward. 'Happy Birthday Art Detective!'. *The Public Catalogue Foundation*, 19 March 2015. http://www.thepcf.org.uk/what_we_do/48/reference/951.
- Sukumaran, Abhay, Stephanie Vezich, Melanie McHugh, and Clifford Nass. 'Normative Influences on Thoughtful Online Participation'. In *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems*, 3401–3410. Vancouver, BC, Canada: ACM, 2011. <http://dl.acm.org/citation.cfm?id=1979450>.
- 'TaDiRAH - Taxonomy of Digital Research Activities in the Humanities (v. 0.5, 02/2014)'. *TaDiRAH - Taxonomy of Digital Research Activities in the Humanities*, February 2014. <https://github.com/dhtaxonomy/TaDiRAH>.
- TEI Consortium. 'TEI: FAQ'. *Text Encoding Initiative Consortium*, 29 June 2014. <http://www.tei-c.org/About/faq.xml>.
- The CASE programme. 'Evidence of What Works: Evaluated Projects to Drive up Engagement'. London: Culture and Sport Evidence (CASE) programme, January 2011. http://www.culture.gov.uk/images/research/evidence_of_what_works.pdf.
- 'The Great War Archive Flickr Group Rules'. *Flickr*. Accessed 15 February 2015. <https://www.flickr.com/groups/greatwararchive/rules/>.
- The University of Melbourne. 'Is There a Tasmanian Convict Settler in Your Family Tree? Participants Sought for Study into Australia's Founding Population'. *The University of Melbourne*, 9 February 2009. <http://archive.uninews.unimelb.edu.au/view-54040.html>.
- The U.S. National Archives and Records Administration. 'The Soundex Indexing System'. *The U.S. National Archives and Records Administration*, 30 May 2007.

- <http://www.archives.gov/research/census/soundex.html>.
- The Zooniverse. 'It's Good to Talk!' *Daily Zooniverse*, 10 April 2014.
<http://daily.zooniverse.org/2014/04/10/its-good-to-talk/>.
- The Zooniverse, and Commenters. '1 Million Classifications'. *Seafloor Explorer*, 10 December 2012. <http://blog.seafoorexplorer.org/2012/12/10/1-million-classifications/>.
- Theimer, Kate. 'Why We Need to Find a Term to Replace "Citizen Archivist"'. *ArchivesNext*, 9 April 2010. <http://www.archivesnext.com/?p=1214>.
- Thomas, William G., Patrick D. Jones, and Andrew Witmer. 'History Harvests: What Happens When Students Collect and Digitize the People's History?' *Perspectives on History*, January 2013. <https://www.historians.org/publications-and-directories/perspectives-on-history/january-2013/history-harvests>.
- Tice, Richard. 'Doing Temple Work at Home'. *Ensign*, July 1990.
<https://www.lds.org/ensign/1990/07/doing-temple-work-at-home?lang=eng>.
- 'Transcribers' Knowledge Base'. *FreeBMD*, 9 February 2014.
http://www.freebmd.org.uk/vol_faq.html#13j.
- 'Transcrib'. *Drupal.Org*, 31 July 2012.
https://www.drupal.org/project/transcribe_distribution.
- 'Transcription Conventions'. *Scriptorium: Medieval and Early Modern Manuscripts Online*. Accessed 24 February 2015.
<http://scriptorium.english.cam.ac.uk/handwriting/materials/conventions/>.
- tranScriptorium. 'Objectives'. *TranScriptorium*. Accessed 15 May 2015.
<http://transcriptorium.eu/pagina/objectives/>.
- Trant, Jennifer. 'Tagging, Folksonomy and Art Museums: Results of Steve.Museum's Research'. *Archives & Museum Informatics*, 7 January 2009.
<http://conference.archimuse.com/files/trantSteveResearchReport2008.pdf>.
- Trouille, Laura. 'Galaxy Zoo Quench – Experience the Full Scientific Process'. *Galaxy Zoo*, 10 July 2013. <http://blog.galaxyzoo.org/2013/07/10/galaxy-zoo-quench-experience-the-full-scientific-process/>.
- Twells, Alison. 'Community History'. *Making History*, 2008.
http://www.history.ac.uk/makinghistory/resources/articles/community_history.html.
- UK Data Service. 'Open Access Data'. *UK Data Service*. Accessed 16 May 2015.
<http://ukdataservice.ac.uk/get-data/open-data.aspx>.
- University of Minnesota Libraries, Andrew W. 'A Multi-Dimensional Framework for Academic Support: A Final Report', June 2006. <http://purl.umn.edu/5540>.
- Unsworth, John. 'Scholarly Primitives: What Methods Do Humanities Researchers Have in Common, and How Might Our Tools Reflect This?' presented at the symposium on 'Humanities Computing: formal methods, experimental practice', King's College, London, 13 May 2000. <http://people.lis.illinois.edu/~unsworth/Kings.5-00/primitives.html>.
- Various. 'AHRC Crowd Sourcing Study: Scoping Seminar'. King's College London, 18 October 2012. <http://crowds.cerch.kcl.ac.uk/>.
- . 'Flairedusers'. *Reddit AskHistorians*. Accessed 7 April 2014.
<http://www.reddit.com/r/AskHistorians/wiki/flairedusers>.
- . 'The Panel of Historians VIII: AskHistorians'. *Reddit AskHistorians*, March 2014.
http://www.reddit.com/r/AskHistorians/comments/1zifm6/the_panel_of_historians_viii/.
- Veridian. 'Crowdsourced User Text Correction (UTC)'. *Veridian*. Accessed 21 July 2014.
<http://www.veridiansoftware.com/knowledge-base/crowdsourced-user-text-correction/>.
- 'Volunteers' FAQ'. *Gutenberg*, 7 December 2014.

- https://www.gutenberg.org/wiki/Gutenberg:Volunteers%27_FAQ.
- 'Volunteers Needed for Massive Smithsonian Digitization Project'. *Newsdesk Newsroom of the Smithsonian*, 12 August 2014. <http://newsdesk.si.edu/releases/volunteers-needed-massive-smithsonian-digitization-project>.
- von Massenbach, Camilla. 'RootsWeb: FreeBMD-Admins-L FreeBMD Project'. *FreeBMD-Admins-L Archives*, 25 September 1998. <http://archiver.rootsweb.ancestry.com/th/read/FreeBMD-Admins/1998-09/0906727973>.
- 'Wedding Fashion'. *Victoria and Albert Museum*, 2011. <http://web.archive.org/web/20110818221004/http://www.vam.ac.uk/things-to-do/wedding-fashion/home>.
- Weinberger, David. '[2b2k] Citizen Scientists'. *Joho the Blog*, 28 December 2010. <http://www.hyperorg.com/blogger/2010/12/28/2b2k-citizen-scientists/>.
- 'Welcome to the Arago Volunteer Researcher Program'. *National Postal Museum*. Accessed 26 February 2015. <http://postalmuseum.si.edu/Arago/Researchers.html>.
- Wenger-Trayner, Etienne. 'Communities of Practice: A Brief Introduction'. *Wenger-Trayner*. Accessed 3 June 2014. <http://wenger-trayner.com/theory/>.
- 'What Is 18thConnect?' *18thConnect*. Accessed 4 September 2014. <http://www.18thconnect.org/about/>.
- 'What Is NameX?' *NameX*. Accessed 2 February 2015. <http://namevariants.co.uk/introduction.aspx>.
- 'What Is Public Engagement?' *National Co-Ordinating Centre for Public Engagement*. Accessed 5 March 2013. <https://www.publicengagement.ac.uk/what>.
- 'What next for CLIP?' *CLIP - Crew List Index Project*, August 2013. <http://www.crewlist.org.uk/aboutclip.html>.
- Whitenton, Kathryn. 'Minimize Cognitive Load to Maximize Usability'. *Nielsen Norman Group*, 22 December 2013. <http://www.nngroup.com/articles/minimize-cognitive-load/>.
- Whittle Wrigley, Heather. 'Technology Helps FamilySearch Volunteers Hit Major Milestone'. *Liahona*, December 2009. <https://www.lds.org/liahona/2009/12/news-of-the-church?lang=eng>.
- 'Who Is behind Find A Grave?' *Find A Grave*. Accessed 14 February 2015. <http://www.findagrave.com/cgi-bin/fg.cgi?page=whois>.
- 'WieWasWie Project Informatie'. *VeleHanden*. Accessed 1 August 2014. http://velehanden.nl/projecten/bekijk/details/project/wiewaswie_bvr.
- Wiggins, Andrea, and Kevin Crowston. 'From Conservation to Crowdsourcing: A Typology of Citizen Science'. In *Proceedings of the 44th Annual Hawaii International Conference on System Sciences*. Koloa, Hawaii, 2011. doi:10.1109/HICSS.2011.207.
- Wikimedia contributors. 'Commons:Project Scope'. *Wikimedia Commons*, 2015. https://commons.wikimedia.org/wiki/Commons:Project_scope.
- Wikipedia contributors. 'Wikimedia Commons'. *Wikipedia, the Free Encyclopedia*, 2013. https://en.wikipedia.org/w/index.php?title=Wikimedia_Commons&oldid=581836800.
- Wikisource. 'What Is Wikisource?' *Wikisource*, 10 July 2014. https://en.wikisource.org/wiki/Wikisource:What_is_Wikisource%3F.
- Wilcox, Jill. 'About MarineLives'. *MarineLives Project Manual*, 14 May 2014. http://marinelives.wikispot.org/About_MarineLives.
- Wilderman, Candie C. 'Models of Community Science: Design Lessons from the Field'. In *Citizen Science Toolkit Conference*, edited by C. McEver, R. Bonney, J. Dickinson, S. Kelling, K. Rosenberg, and J. L. Shirk. Cornell Laboratory of Ornithology, Ithaca, New York: National Science Foundation, 2007. <http://birds.cornell.edu/citscitoolkit/conference/proceeding->

- pdfs/Wilderman%202007%20CS%20Conference.pdf.
- Willett, Kyle. 'New Paper: Galaxy Zoo and Machine Learning'. *Galaxy Zoo*. Accessed 31 March 2015. <http://blog.galaxyzoo.org/2015/03/31/new-paper-galaxy-zoo-and-machine-learning/>.
- Willever-Farr, Heather L., and Andrea Forte. 'Family Matters: Control and Conflict in Online Family History Production'. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work and Social Computing*. Baltimore, MD, USA.: ACM, 2014. <http://dl.acm.org/citation.cfm?id=2531737>.
- Winnips, Koos, and Catherine McLoughlin. 'Six WWW Based Learner Supports You Can Build.' In *Proceedings*. Tampere, Finland, 2001. <http://eric.ed.gov/?id=ED466231>.
- Wolfe, Jen. 'Contributing in Code'. *The University of Iowa Libraries*, 18 November 2014. <http://blog.lib.uiowa.edu/drp/2014/11/18/contributing-in-code/>.
- . 'Crowdsourcing Continued'. *The University of Iowa Libraries*, 16 March 2012. <http://blog.lib.uiowa.edu/drp/2012/03/16/crowdsourcing-continued/>.
- . 'DIY History Celebrates 50,000th Transcription!' *The University of Iowa Libraries*, 11 November 2014. <http://blog.lib.uiowa.edu/drp/2014/11/11/diyh5ok/>.
- . 'UI Libraries Launches New Crowdsourcing Site with Manuscript Cookbooks and More'. *The University of Iowa Libraries*, 9 October 2012. <http://blog.lib.uiowa.edu/drp/2012/10/09/diyhistory/>.
- Yakel, Elizabeth, Polly Reynolds, Seth E. Shaw, Jeremy York, and Magia Krause. 'Polar Bear Expedition Digital Collections: Enhancing Online Use through Digital Curation'. Chapel Hill, North Carolina, 2007.
- York, Amber. 'Possible New Species? "Convict Worm"'. *Seafloor Explorer*, 14 September 2012. <http://blog.seafloorexplorer.org/2012/09/14/possible-new-species-convict-worm/>.
- Young, Kerri. 'Year of the Bay Project Officer Evaluation'. San Francisco: HistoryPin, May 2014.
- Young, Stephen C. 'Working Together'. *Ensign*, March 1996. <https://www.lds.org/ensign/1996/03/working-together?lang=eng>.
- Zamora, Amanda. 'Crowdsourcing Campaign Spending: What We Learned From Free the Files'. *ProPublica*, 12 December 2012. <http://www.propublica.org/article/crowdsourcing-campaign-spending-what-we-learned-from-free-the-files>.

Appendix A: Websites and projects reviewed

'Africa' Flickr set, The National Archives UK,
<https://www.flickr.com/photos/nationalarchives/collections/72157625827328771/>

#TranscribeTuesday,
<https://twitter.com/search?f=realtime&q=%23transcribeTuesday&src=typd>

10 Most Wanted, <http://10most.org.uk/>

100 Toys (and Their Stories) that Define Our Childhood,
<http://web.archive.org/web/20120723072736/http://www.childrensmuseum.org/100toys>

1000Memories, <http://web.archive.org/web/20121123001926/http://1000memories.com/>

1001 Stories of Denmark, http://www.kulturarv.dk/1001fortaellinger/en_GB

18th Century Common, <http://www.18thcenturycommon.org/>

18thConnect, <http://www.18thconnect.org/>

A Vision of Britain Through Time, <http://www.visionofbritain.org.uk/>

Academia.edu, <http://academia.edu/>

Accurator, <http://accurator.nl/>

Acumen, University of Alabama Libraries, <http://acumen.lib.ua.edu/home>

Addressing History, <http://addressinghistory.edina.ac.uk/>

AHA Communities, <http://communities.historians.org>

Ajapaik, <http://ajapaik.ee/>

Aluka, <http://www.aluka.org/>

Amara, <http://www.amara.org/en/>

American Air Museum in Britain, <http://www.americanairmuseum.com/contribute>

Ancestral Atlas, <http://www.ancestralatlas.com>

Ancestry Content Publisher Beta, <http://publish.ancestry.com/>

Ancestry Shoebox, <http://shoebox.ancestry.com/>

Ancestry World Archives Program (AWAP), <http://community.ancestry.com/awap>

Ancestry, <http://www.ancestry.com/>

Ancient Lives, <http://www.ancientlives.org/>

AncientFaces, <http://www.ancientfaces.com/>

Anglo-Scottish Migration,
http://web.archive.org/web/20131118083936/http://wiki.angloscottishmigration.humanities.manchester.ac.uk:80/index.php/Main_Page

AnnoMarket, <https://annomarket.com/>

Annotated Books Online, <http://www.annotatedbooksonline.com/>

Arago, <http://arago.si.edu/>

Archaeology Community Co-Production of Research Data (ACCORD),

<https://accordproject.wordpress.com/>
 Archival Great Lakes Ship List Transcription Project, <http://www.shipwrecks.net/shipwreck/projects/listproject/>
 arcHIVE National Archives of Australia, <http://transcribe.naa.gov.au/>
 Archives of American Gardens, <http://www.gardens.si.edu/collections-research/aag.html>
 Archives Portal Europe, <http://www.archivesportaleurope.net/>
 Archives2Share State Records NSW, <https://www.flickr.com/groups/archives2share/rules/>
 Art Detective, Public Catalogue Foundation (PCF), <http://www.thepcf.org.uk/artdetective/>
 Arthur Schnitzler digital, <http://www.arthur-schnitzler.org/>
 Artigo, <http://www.artigo.org/>
 Arts and Humanities Data Service (AHDS), <http://www.ahds.ac.uk/>
 Atlas of Living Australia DigiVol, <http://volunteer.ala.org.au/>
 AustESE, <http://www.austese.net/>
 Australasian Heritage Software Database,
<https://web.archive.org/web/20150402192137/http://www.ourdigitalheritage.org/>
 Bat Detective, <http://www.batdetective.org/>
 BBC - WW2 People's War, <http://www.bbc.co.uk/history/ww2peopleswar/>
 Berkeley Prosopography Services,
<https://wikihub.berkeley.edu/display/istds/Berkeley+Prosopography+Services+Wiki+Home>
 Beside the Seaside, National Maritime Museum, <http://www.rmg.co.uk/whats-on/exhibitions/past/beside-the-seaside/home>
 Big Seaweed Search, <http://www.nhm.ac.uk/nature-online/british-natural-history/seaweeds-survey/>
 Billion Graves, <http://billiongraves.com/>
 Biodiversity Heritage Library, <http://www.biodiversitylibrary.org/>
 Bone Commons, <http://alexandriaarchive.org/bonecommons/>
 Book Traces, <http://www.booktraces.org/>
 Bracero History Archive, <http://braceroarchive.org/>
 Britain From Above, <http://www.britainfromabove.org.uk/>
 British Library Georeferencer, <http://www.bl.uk/maps/>
 British Library UK Soundmap, <http://sounds.bl.uk/Sound-Maps/UK-Soundmap>
 British Newspaper Archive, <http://www.britishnewspaperarchive.co.uk/>
 British Postal Museum and Archive wiki,
<http://web.archive.org/web/20100324090955/http://postalheritage.org.uk/wiki>
 British Record Society, <http://www.britishrecordsociety.org>
 Building Inspector, New York Public Library, <http://buildinginspector.nypl.org/>
 BYU Historic Journals Project,
<http://web.archive.org/web/20130927050552/http://journals.byu.edu/>

California Digital Newspaper Collection, <http://cdnc.ucr.edu/cgi-bin/cdnc>

Cambridge Group for the History of Population and Social Structure, <http://www.campop.geog.cam.ac.uk/>

Canadian Writing Research Collaboratory (CWRC), <http://www.cwrc.ca/en/>

Carnamah Historical Society & Museum, <http://www.virtualvolunteering.com.au/volunteer/>

Cartoteca, <http://cartotecadigital.icc.cat/cdm/>

Cell Slider, <http://www.cellslider.net/>

CHALICE, <http://chalice.blogs.edina.ac.uk/>

Children of the Lodz Ghetto, <http://www.ushmm.org/online/lodzchildren/>

Chimp & See, <http://chimpandsee.org/>

Chronicling America, <http://chroniclingamerica.loc.gov/>

Church Detective, <http://www.churchdetective.org.uk/>

Citizen Archivist Dashboard, National Archives (NARA) , <http://www.archives.gov/citizen-archivist/>

Citizen Science Alliance, <http://www.citizensciencealliance.org/>

Citizendium, <http://en.citizendium.org/>

CitizenSort, <http://web.archive.org/web/20150511030747/http://citizensort.org/>

Civil War Pathways Read-In, <http://pathways.omeka.net/about>

Clergy of the Church of England Database, <http://theclergydatabase.org.uk/>

Clickworkers, <http://nasaclickworkers.com/>

Clio, <http://www.theclio.com/web/>

Co-Curate, <https://co-curate.ncl.ac.uk/>

Collaborative European Digital Archive Infrastructure (CENDARI), <http://www.cendari.eu/virtual-research-environment/>

Collaborative for Historical Information and Analysis (CHIA), <http://chia.pitt.edu/index.htm>

Collections Crowdsourcing Beta, V&A Museum, <http://collections.vam.ac.uk/crowdsourcing/>

Colored Conventions, <http://coloredconventions.org/>

Comm@NET, <http://commanet.org/>

Commodity Histories, <http://www.commodityhistories.org/>

Commons In A Box (CBOX), <http://commonsinabox.org/>

Commons Machinery, <http://commonsmachinery.se/>

Community Maps, http://www.communitymaps.org.uk/version6_1/includes/CommunityMaps.php

Community Sites software, <http://www.communitysites.co.uk/>

Community-driven Curation for Taxonomic Databases (ComTax), <http://taxoncuration.myspecies.info/>

Connected Histories, <http://www.connectedhistories.org/>
 Cornwall Online Parish Clerks, <http://www.cornwall-opc.org>
 Correct, Gallica, <http://www.reseau-correct.fr/>
 Crew List Index Project (CLIP), <http://www.crewlist.org.uk/>
 Critical Commons, <http://www.criticalcommons.org/>
 Crowd-Ed Plugin, <http://gsbodine.github.io/crowd-ed/>
 Crowd-Ed, <http://crowd-ed.org/>
 Crowdcrafting, <http://crowdcrafting.org/>
 Crowdmap the Crusades, <http://dhcrowdscribe.com/crowdmap-the-crusades/>
 CrowdMap, Ushahidi, <http://www.ushahidi.com/product/crowdmap/>
 CUNY Academic Commons, <http://commons.gc.cuny.edu/>
 Curarium, <https://www.curarium.com/>
 Curatescape, <http://curatescape.org/>
 Curio, <http://www.crowdcurio.com/>
 Cymru1900Wales, <http://www.cymru1900wales.org/>
 Cynefin, <http://cynefin.archiveswales.org.uk/>
 Danish Demographic Database, http://ddd.dda.dk/ddd_en.htm
 David Rumsey Historical Map Collection, <http://www.davidrumsey.com>
 Dead Fred, <http://deadfred.com/>
 Describe Me, <http://describeme.museumvictoria.com.au/>
 Dickens Journal Online, <http://www.djo.org.uk/>
 DigiPal, <http://www.digipal.eu/>
 Digital Archive of Japan's 2011 Disasters, <http://www.jdarchive.org/>
 Digital Exposure of English Place-names (DEEP), <http://englishplacenames.cerch.kcl.ac.uk>
 DigitalKoot, <http://www.digitalkoot.fi/>
 DigitalMellini, <http://www.getty.edu/research/mellini/>
 DigiVol Australian Museum, <http://australianmuseum.net.au/digivol>
 Direct User Repository Access (DURA), <http://jisc-dura.blogspot.com/>
 Discovering Anzacs, <http://discoveringanzacs.naa.gov.au/>
 Disk Detective, <http://www.diskdetective.org/>
 Distributed Proofreaders, <http://www.pgdp.net/>
 DIYHistory, <http://diyhistory.lib.uiowa.edu/>
 Dotspotting, <http://dotspotting.org/>
 Dukesfield Documents, <http://www.dukesfield.org.uk/research/dukesfield-documents/>
 Duolingo, <https://www.duolingo.com/>
 e-artexte, <http://e-artexte.ca/>

Early Modern OCR Project (eMOP), <http://emop.tamu.edu/>
 Early Transatlantic Writing Project, <http://transatlanticwriting.com/>
 East London History, <http://eastlondonhistory.com/>
 Edinburgh Collected, <https://www.edinburghcollected.org/>
 Edinburgh Geoparser, http://www.ltg.ed.ac.uk/clusters/Edinburgh_Geoparser/
 Editors' Notes, <http://editorsnotes.org/>
 Electronic Cultural Atlas Initiative (ECAI), <http://ecai.org/>
 Emigrants, <http://web.archive.org/web/20140125085218/http://emigrants.net.au/>
 Empire Faith War, <http://www.empirefaithwar.com/>
 English Handwriting Course, <http://scriptorium.english.cam.ac.uk/handwriting/course/>
 Ensemble, New York Public Library, <http://ensemble.nypl.org/>
 European History Primary Sources (EHPS), <http://primary-sources.eui.eu/>
 Europeana 1914-1918, <http://www.europeana1914-1918.eu/en>
 Europeana 1989, <http://www.europeana1989.eu/en/>
 Expose - Mijn Mooiste Landschap, <http://editie2.kmmexpose.nl/>
 Family and Community Historical Research Society (FACHRS), <http://www.fachrs.com/>
 Family History Transcription Project, State Library of North Carolina,
<https://www.flickr.com/photos/statelibrarync/sets/72157627124710723>
 FamilySearch Indexing, <https://familysearch.org/indexing/>
 FamilySearch Memories, <https://familysearch.org/photos/>
 FamilySearch, <https://familysearch.org/>
 Field Notes of Junius Henderson,
http://en.wikisource.org/wiki/Field_Notes_of_Junius_Henderson
 Field Notes of Laurence M. Klauber,
<http://web.archive.org/web/20141018121033/http://fromthepage.bpoc.org:80/>
 Figshare, <http://figshare.com/>
 Find A Grave, <http://www.findagrave.com/>
 Find-A-Record, <https://www.findarecord.com/>
 Findmypast, <http://findmypast.com/>
 Flickr Commons Tagr, <http://whatsthatpicture.com/flickr/commons/>
 Flickr Commons, <https://www.flickr.com/commons/>
 Floating Forests, <http://www.floatingforests.org/>
 Fold3, <http://www.fold3.com/>
 FoldIt, <http://fold.it>
 Folgerpedia, http://folgerpedia.folger.edu/Main_Page
 Founders and Survivors, <http://foundersandsurvivors.org/>
 Free the Files, <http://www.propublica.org/series/free-the-files>

FreeBMD, <http://www.freebmd.org.uk/>
 FreeCen, <http://www.freecen.org.uk/>
 FreeReg, <http://www.freereg.org.uk/> (new version <http://freereg2.freereg.org.uk/>)
 From Fights to Rights Transcription Project,
<http://scrcdigital.swem.wm.edu/collections/show/1>
 FromThePage, <http://beta.fromthepage.com/>
 Galaxy Zoo 1 data release, <http://data.galaxyzoo.org/>
 Galaxy Zoo Forum, <http://galaxyzooforum.org>
 Galaxy Zoo, <http://www.galaxyzoo.org/>
 Gazetteer of British Place Names, Association of British Counties,
<http://www.gazetteer.co.uk/section4.html>
 GeneaGraves, <http://en.geneanet.org/geneagraves>
 genealogy.com, <http://www.genealogy.com/>
 Geneanet, <http://en.geneanet.org/>
 Geni, <http://www.geni.com/>
 Genius Loci, <http://geniusloci.chydenius.fi/>
 GENUKI, <http://www.genuki.org.uk/>
 GenWiki, http://wiki-en.genealogy.net/Main_Page
 Geograph, <http://www.geograph.org.uk/>
 GeoKey, <http://geokey.org.uk/>
 GeoNames, <http://www.geonames.org/>
 GeoTag-X, <http://geotagx.org/>
 Google Ancient Places, <https://googleancientplaces.wordpress.com/>
 Google Earth, <http://earth.google.com/>
 Google Maps Engine, <https://mapsengine.google.com>
 Google Maps, <http://maps.google.co.uk/>
 Google Scholar, <http://scholar.google.co.uk/>
 Gough Map, <http://www.goughmap.org/>
 Gravestone Photographic Resource, <http://www.gravestonephotos.com>
 Great War Forum, <http://1914-1918.invisionzone.com/forums/>
 Guardian MPs Expenses, <http://mps-expenses.guardian.co.uk>
 Guardian Witness, <https://witness.theguardian.com/>
 Guild of One-Name Studies, <http://one-name.org/>
 Hanslope and District Historical Society, <http://www.mkheritage.co.uk/hdhs/index.html>
 Help from Home, <http://helpfromhome.org/>
 Herbaria@Home, <http://herbariaunited.org/atHome/>
 Herts Memories, <http://www.hertsmemories.org.uk/>

Higgs Hunters, <http://www.higgshunters.org/>
 HistoGrafica, <http://web.archive.org/web/20120404081644/http://www.histografica.com/>
 HistoGraph, <https://cubrik3.eng.it:8443/SMILA/hoef/index.html>
 Histopolis, <http://www.histopolis.com/>
 Historic Digimap, EDINA,
http://edina.ac.uk/digimap/description/historic_overview.shtml
 Historic Graves, <http://historicgraves.com/>
 Historical Place Names Microservice, <http://tautosaka.iliti.it/en/unitedgeo/>
 History Commons, <http://historycommons.org/>
 History of Computing in Learning and Education, <http://www.hcle.org>
 History Working Papers Project, <http://www.historyworkingpapers.org/>
 HistoryPin, <https://www.historypin.org/>
 Home Front Legacy 1914-18, <http://www.homefrontlegacy.org.uk/wp/>
 Hurricane Digital Memory Bank, <http://www.hurricanearchive.org/>
 Hydra-in-a-Box, <http://projecthydra.org/>
 hypercities, <http://www.hypercities.com/>
 Image Recognition Check, Public Catalogue Foundation (PCF),
<http://taggerimagerecognition.thepcf.org.uk/pcf/results/index.php>
 Immigrant Ancestors Project, <http://immigrants.byu.edu/>
 Improve Google Translate, <https://translate.google.com/community>
 iNaturalist, <http://www.inaturalist.org>
 interment.net, <http://www.interment.net>
 International Amateur Scanning League, <https://groups.google.com/forum/#!forum/iasl-discuss>
 Internet Archive, <http://archive.org/>
 Internet History Sourcebooks Project, <http://www.fordham.edu/halsall/>
 Invisible Australians, <http://www.invisibleaustralians.org/>
 Ireland Reaching Out, <http://www.irelandxo.com/>
 Islamic History Commons, <http://islamichistorycommons.org/>
 Islamic Manuscripts at Michigan, <http://www.lib.umich.edu/islamic/>
 iSpot, <http://www.ispotnature.org/>
 Itinera Nova, <http://freyja.uni-koeln.de:8585/in/home>
 Jesuit Libraries Provenance Project,
<https://www.flickr.com/photos/jesuitlibrariesprovenanceproject/>
 Jewish Women on the map,
<http://web.archive.org/web/20130113073456/http://jwa.org/onthemap>
 Jones' Icones, <http://www.jonesicones.com/identify/>
 JRC Fuzzy Gazetteer, <http://dma.jrc.it/services/fuzzyg/>

Kete Horowhenua, <http://horowhenua.kete.net.nz>
 Kete software, <http://www.kete.net.nz/>
 Know Your Place, <http://maps.bristol.gov.uk/knowyourplace/>
 Letter in the Attic, My Brighton and Hove,
http://www.mybrightonandhove.org.uk/category_id__1543.aspx
 Letters of 1916, <http://dh.tcd.ie/letters1916/>
 LibriVox, <https://librivox.org/>
 Linked Jazz, <https://linkedjazz.org/>
 Listening Experience Database, <http://led.kmi.open.ac.uk/>
 LitLong, <http://litlong.org/>
 Lives of the First World War, <https://livesofthefirstworldwar.org/>
 Living with Railroads, <http://www.historypin.com/project/42-railroads/>
 Llŷn Archaeology and History Society, <http://www.crwydro.co.uk>
 Localwiki, <https://localwiki.org/>
 Locating London's Past, <http://locatinglondon.org/>
 London Lives, <http://londonlives.org/>
 London Street Views 1840, <http://crowd.museumoflondon.org.uk/lsv1840/>
 Look back at Braddock, <http://braddockheritage.org/>
 Look Back Maps, <http://lookbackmaps.net/>
 Lost Plays, http://www.lostplays.org/index.php/Main_Page
 Lost Visions Illustration Archive, <http://illustrationarchive.cardiff.ac.uk/>
 LostCousins, <http://lostcousins.com/>
 Louisville Leader Transcription Project,
<http://transcribe.library.louisville.edu/omeka/items/browse?collection=10>
 Madison, Hive, New York Times R&D Lab,
<http://www.nytlabs.com/projects/madison.html>
 Making History Searchable: Transcribe It!, Dolph Briscoe Center for American History,
<http://web.archive.org/web/20140409024523/http://www.cah.utexas.edu/projects/transcriber.php>
 Manchester and Lancashire Family History Society (MLFHS),
http://www.mlfhs.org.uk/projects/proj_list_public.php
 Map the Museum, <http://mapthemuseum.org.uk>
 Map Warper, New York Public Library, <http://maps.nypl.org/warper/>
 Mapping Texts, <http://mappingtexts.org/>
 Maptcha, <http://maptcha.org/>
 Marine Lives, <http://marinelives.org/>
 Mark My Bird, <http://hridigital.shef.ac.uk/mark-my-bird>
 Martha Berry Digital Archive (MBDA), <https://mbda.berry.edu/participate>

Mechanical Turk, <https://www.mturk.com/>
 MediaCommons, <http://mediacommons.futureofthebook.org/>
 MediaWiki, <https://www.mediawiki.org/>
 Medici Archive BIA, <http://bia.medici.org>
 Medieval Electronic Scholarly Alliance (MESA), <http://www.mesa-medieval.org/>
 meiosis.org.uk, <http://www.meiosis.org.uk/>
 MicroPasts, <http://micropasts.org/>
 Microsoft Academic Search, <http://academic.research.microsoft.com/>
 Milton Keynes Heritage Association,
<http://www.mkheritage.co.uk/mkha/mkha/members/members.html>
 Minnesota Geospatial Commons, <https://gisdata.mn.gov/>
 MLA Commons, <https://commons.mla.org/>
 Mobiang, <http://mobiang.appspot.com/>
 Mocavo, <https://www.mocavo.com/>
 Monasterium.net, <http://monasterium.net/mom/>
 MShed Your Stories,
<http://web.archive.org/web/20140208002353/http://mshed.org/explore-contribute/your-stories/>
 Mukurtu, <http://www.mukurtu.org/>
 Museum Metadata Games, <http://museumgam.es/>
 Museum of London - Surrey Canal Project,
<http://crowd.museumoflondon.org.uk/surreycanal/>
 My Brighton and Hove, <http://www.mybrightonandhove.org.uk/>
 My Leicestershire History, <http://specialcollections.le.ac.uk/cdm/myleicestershirehistory>
 My-Parish, <http://my-parish.org/>
 MyHeritage, <http://www.myheritage.com/>
 MyHistoricLA,
<http://web.archive.org/web/20141025125947/http://www.myhistoricla.org:80/>
 myKive, <http://www.mykive.org/>
 National Archives Citizen Archivist Research,
<https://www.flickr.com/groups/citizenarchivist/>
 National Archives Transcription Pilot Project ,
<http://web.archive.org/web/20140711032140/http://transcribe.archives.gov/>
 National Historical Geographic Information System (NHGIS), <https://nhgis.org/>
 Naturalis, <http://www.naturalis.nl/en/museum/livescience/crowd-sourcing>
 Nature Locator, <http://naturelocator.org/>
 Naval-History.Net, <http://naval-history.net/>
 Neatline, <http://neatline.org/>

Networked Infrastructure for Nineteenth-Century Electronic Scholarship (NINES),
<http://www.nines.org/>
 newruskinarchives, <http://newruskinarchives.org.uk/>
 Nominatim, <http://nominatim.openstreetmap.org>
 Notes from Nature, <http://www.notesfromnature.org/>
 Now & Then, www.nowandthen.net.au/
 NQuire-It, <http://www.nquire-it.org/>
 NYC Space/Time Directory, <http://spacetime.nypl.org/>
 OED Appeals, Oxford English Dictionary, <http://public.oed.com/appeals>
 Ohio History Service Corps Community Collection,
<http://www.ohiohistoryhost.org/ohiomemory/about/community>
 Old Bailey Online, <http://www.oldbaileyonline.org/>
 Old Maps Online, <http://www.oldmapsonline.org>
 Old Weather, <http://www.oldweather.org/>
 OldSF, <http://www.oldsf.org/>
 Olney and District Historical Society (ODHS), <http://www.mkheritage.org.uk/odhs/>
 Omeka, <http://omeka.org/>
 OneGreatFamily, <http://www.onegreatfamily.com/>
 Online Names, <http://web.archive.org/web/20140124075314/http://onlinenames.net.au/>
 Online Parish Clerks,
<http://web.archive.org/web/20130114221811/http://www.onlineparishclerks.org.uk/>
 Open Book Publishers, <http://www.openbookpublishers.com/>
 Open Context, <http://opencontext.org/>
 Open Domesday, <http://opendomesday.org/>
 Open Library of Humanities, <https://www.openlibhums.org/>
 Open Museum, <http://www.openmuseum.org/>
 Open Plaques, <http://openplaques.org/>
 Open Society Archives (OSA), <http://www.osaarchivum.org/>
 openICPSR, Inter-university Consortium for Political and Social Research,
<https://www.openicpsr.org/>
 OpenStreetMap, <http://openstreetmap.org/>
 OpenText.org, <http://opentext.org/>
 Operation War Diary, <http://operationwardiary.org/>
 Orchid Observers, <http://www.orchidobservers.org/>
 Oregon State Special Collections & Archives Research Center,
<http://scarc.library.oregonstate.edu/omeka/items/browse>
 Origins, <http://www.origins.net/>
 Oxfordshire Family History Society, <http://www.ofhs.org.uk/>

Papers of the War Department, <http://wardepartmentpapers.org/>
 Parallel Archive, <http://www.parallelarchive.org/>
 Pararchive Project, <http://pararchive.com/>
 Parish Chest, <http://www.parishchest.com>
 Parish Locator, <http://www.parloc.pwp.blueyonder.co.uk/parishlocator.html>
 Parish Register Transcription Society, <http://www.prtsoc.org.uk/>
 Pastigo, <http://web.archive.org/web/20130205021406/http://www.pastigo.com/>
 Pastmapper, <http://pastmapper.com/>
 Pelagios, <http://pelagios-project.blogspot.co.uk/>
 Penguin Watch, <http://www.penguinwatch.org>
 PennTags, <http://tags.library.upenn.edu/>
 People and the Post Digital Memory Book, Smithsonian National Postal Museum,
<http://memorybook.si.edu/>
 Pepys Diary, <http://pepysdiary.com/>
 PhillyHistory, <http://www.phillyhistory.org/>
 Phrase Detectives, <http://anawiki.essex.ac.uk/phrasedetectives/index.php>
 Picture Australia, <http://web.archive.org/web/2011001034702/http://pictureaustralia.org/>
 Planet Hunters, <http://www.planethunters.org/>
 Plankton Portal, <http://www.planktonportal.org/>
 Play it again, <http://playitagainproject.org/>
 Pleiades, <http://pleiades.stoa.org/>
 Polar Bear Expedition,
<https://web.archive.org/web/20070611202024/http://polarbears.si.umich.edu/>
 Politiets Registerblade, <http://www.politietsregisterblade.dk/>
 Potlatch2 for MapQuest OpenStreetMap, <http://wiki.openstreetmap.org/wiki/MapQuest>
 Project Budburst, <http://budburst.org/>
 Project Gado, <http://projectgado.org/>
 Project Gutenberg, <https://www.gutenberg.org/>
 Project Woruldhord, <http://projects.oucs.ox.ac.uk/woruldhord/>
 Prosop, <http://www.prosop.org/>
 Prosopography of Anglo-Saxon England (PASE), <http://www.pase.ac.uk/>
 Provenance Online Project, <https://www.flickr.com/photos/58558794@No7/>
 Public Record Office Victoria PROV Wiki, <http://wiki.prov.vic.gov.au/>
 Public Researcher Wiki, National Archives (NARA),
<http://web.archive.org/web/20130807103345/http://www.ourarchives.wikispaces.net/>
 Putting Art on the Map, Imperial War Museums and HistoryPin,
<https://www.historypin.org/project/41-putting-art-on-the-map/>

Pybossa, <http://pybossa.com/>
 Qualitative Data Repository (QDR), <https://qdr.syr.edu/>
 Quench, Galaxy Zoo, <http://quench.galaxyzoo.org/>
 Random Acts of Genealogical Kindness, <http://www.raogk.org/>
 Ravelry, <http://www.ravelry.com/>
 Reading Experience Database, <http://www.open.ac.uk/Arts/RED/>
 reCAPTCHA,
<https://web.archive.org/web/20140115093426/http://www.google.com/recaptcha>
 Recogito , <http://pelagios.org/recogito/>
 Red een Portret, <http://redeenportret.nl/>
 reddit AskHistorians, <http://www.reddit.com/r/AskHistorians/>
 Register of One-Place Studies, <http://www.register-of-one-place-studies.org.uk/index.html>
 Remember Me, <http://rememberme.ushmm.org/>
 Representing Re-Formation, <http://representingreformation.net/>
 Researchobject.org, <http://www.researchobject.org/>
 Reviews in History, <http://www.history.ac.uk/reviews/>
 Roots2Share, <http://www.roots2share.org/>
 RootswEB Mailing Lists, <http://lists.rootswEB.ancestry.com>
 RootswEB, <http://rootswEB.com/>
 Royal BC Museum Transcribe, <http://transcribe.royalbcmuseum.bc.ca/>
 RunCoCo, <http://projects.oucs.ox.ac.uk/runcoco/>
 Sapelli, <https://www.ucl.ac.uk/excites/software/sapelli>
 Scalar, <http://scalar.usc.edu/scalar/>
 Science Gossip, <http://www.sciencegossip.org/>
 Science Museum Object Wiki,
<http://web.archive.org/web/20140310083339/http://objectwiki.sciencemuseum.org.uk:80/wiki/>
 SciStarter, <http://scistarter.com/>
 ScotlandsPlaces, <http://www.scotlandsplaces.gov.uk/>
 Scripto, <http://scripto.org/>
 Seafloor Explorer, <http://www.seafloorexplorer.org/>
 Sejrs Sedler, <http://www.sejrssedler.dk/>
 Self Archiving Legacy Toolkit (SALT),
<http://stanfordluminaryarchives.googlepages.com/salt>
 Sepia Town, <http://www.sepiatown.com/>
 Share History, <http://sharehistory.org/>
 Share Your Story, Agricultural Innovation and Heritage Archive, National Museum of
 American History, <http://americanhistory.si.edu/agheritage/share-your-story>

Shropshire's Heritage Heroes, <http://www.heritageheroes.org.uk/>

Smithsonian Digital Volunteers: Transcription Center, <https://transcription.si.edu/>

Smithsonian Field guide for men in the armed forces,
<http://archive.org/stream/fieldcollectorsmoosmit#page/n5/mode/2up>

Snapshot Supernova,
<https://stargazing2015.zooniverse.org/#/projects/zooniverse/Snapshot%20Supernova>

Social Networks and Archival Context (SNAC), <http://socialarchive.iath.virginia.edu/>

Society of Genealogists, <http://www.sog.org.uk/>

Soldier Studies, <http://www.soldierstudies.org/>

St Fagans Archive, National Museum Wales,
<https://www.flickr.com/photos/museumwales/sets/72157645102673236/>

St. Helier Memories, <http://www.heliermemories.org.uk/>

Standards for Networking Ancient Prosopographies: Data and Relations in Greco-Roman Names (SNAP:DRGN), <http://snapdrgn.net/>

Stardust@home, <http://stardustathome.ssl.berkeley.edu/>

State Library NSW Georeferencer, <http://nsw.georeferencer.com/random>

State Library NSW Transcripts, <http://transcripts.sl.nsw.gov.au/>

State Library Queensland Digital Volunteers,
<https://www.flickr.com/groups/archives2share/>

steve.museum/, <http://web.archive.org/web/20150530183936/http://www.steve.museum/>

Stockholms Stad,
<http://www2.ssa.stockholm.se/Bildregistrering/Mantalsregister/Default.aspx>

Stories From Main Street, <http://storiesfrommainstreet.org/>

Strandlines, <http://www.strandlines.net/>

Stream of Digital Archives (SODA), <http://soda.naa.gov.au/>

Suda on Line, <http://www.stoa.org/sol/>

Sydney Benevolent Asylum, <http://www.sydneybenevolentasylum.com/>

System for Annotation and Linkage of Sources in Arts and Humanities (SALSAH),
<http://www.salsah.org/>

tagasauris, <http://tagasauris.com/>

Tags en Uitleg, https://velehanden.nl/projecten/bekijk/details/project/mai_tagging

TAPAS Project, <http://tapasproject.org/>

Tate Art Maps, <http://artmaps.tate.org.uk/>

TED Open Translation Project, <http://www.ted.com/about/programs-initiatives/ted-open-translation-project>

Tenbury Regal Memory Reel, <http://www.regaltenbury.org.uk/memory-reel/>

Texas Manuscript Cultures, <http://txmscultures.writingstore.com/>

Text-Image Linking Environment (TILE), <http://mith.umd.edu/tile/>

TextGrid, <http://www.textgrid.de>

The Agas Map of Early Modern London, <http://mapoflondon.uvic.ca/>

The Civil War in Letters, <http://publications.newberry.org/civilwarletters/>

The Digital Inn, <http://arkivverket.no/eng/Digitalarkivet/About-the-Digital-Archives/The-Digital-Inn>

The Digitally Encoded Census Information and Mapping Archive (DECIMA), <http://maps-hist.chass.utoronto.ca/decima/>

The Great War Archive, <http://www.oucs.ox.ac.uk/wwlit/gwa>

The Historic Gazetteer, <http://gov.genealogy.net/search/index>

The Historical Gazetteer of England's Place-Names (DEEP), <http://placenames.org.uk/>

The History Harvest, <http://historyharvest.unl.edu/>

The National Archives Community, <https://community.nationalarchives.gov.uk/>

The Remembering Site, <http://www.thereforgettingsite.org/>

The Shoebox by PSNH, <http://psnhshoebox.com/>

The Theory and Practice of Social Machines (SOCIAM), <http://sociam.org/>

The Victorian Web, <http://www.victorianweb.org/>

Tiltfactor Metadata Games, <http://www.tiltfactor.org/game/metadata-games/>

Time Shutter, <http://www.timeshutter.com/>

TNA Your Archives, <http://yourarchives.nationalarchives.gov.uk/>

Tombstone Transcription Project, <http://usgwtombstones.org/>

Transcribable, <https://github.com/propublica/transcribable>

Transcribe Bentham, <http://www.ucl.ac.uk/transcribe-bentham/>

Transcribe Tuesday, Horniman Museum and Gardens Flickr Group, <https://www.flickr.com/groups/transcribetuesday/>

Transcribing the Past, <https://transcribe.ou.edu/>

Transcribo, <http://transcribo.org>

Transcription for Paleographical and Editorial Notation (T-PEN), <http://t-pen.org/>

tranScriptorium, <http://transcriptorium.eu>

Trove Newspapers, <http://trove.nla.gov.au/newspaper/>

True North Mapping Minnesotas History, <http://www.mngeo.state.mn.us/ghol/>

TypeWright, www.i8thconnect.org/typewright/

UK Census Online, <http://www.freecen.org.uk/>

UK Data Archive, <http://www.essex.ac.uk/depts/ukda.aspx>

UK Data Service, <http://ukdataservice.ac.uk/>

Unlock, Edina, <http://edina.ac.uk/unlock/>

UrCrowdsource, <http://urcrowdsource.org/omeka/>

USGenWeb Free Census Project, <http://www.usgwcensus.org/>

USGS North American Bird Phenology Program, <http://www.pwrc.usgs.gov/bpp/>

VCH Explore, Victoria County History, <http://www.victoriacountyhistory.ac.uk/explore/>

Velehanden, <http://velehanden.nl/>

Veridian Digital Library Software, <http://www.veridiansoftware.com/>

Victoria County History, <http://www.victoriacountyhistory.ac.uk/>

Victorian Professions, <http://www.victorianprofessions.ox.ac.uk/>

Viewshare, <http://viewshare.org/>

Virginia Memory Making History - Transcribe, <http://www.virginiamemory.com/transcribe/>

Volunteer Match, <http://www.volunteermatch.org>

Voyant Tools, <http://voyeurtools.org/>

Waisda?, <http://waisda.nl/>

Weather Detective, <http://www.weatherdetective.net.au/>

Weaving History, <http://www.weavinghistory.org/>

Wedding Fashion, V&A Museum, <http://web.archive.org/web/20110312043926/http://www.vam.ac.uk/things-to-do/wedding-fashion/home>

Welsh Newspapers Online, <http://papuraunewyddcymru.llgc.org.uk/en/home>

Welsh Voices of the Great War Online, <http://www.peoplescollection.wales/collections/377191>

WeRelate, <http://www.werelate.org>

What Was There, <http://www.whatwasthere.com/>

What's on the Menu Geotagger, New York Public Library, <http://menusgeo.herokuapp.com/>

What's on the Menu, New York Public Library, <http://menus.nypl.org/>

What's the Score, <http://www.whats-the-score.org/>

Whitney Research Group Wiki, http://wiki.whitneygen.org/wrg/index.php/Main_Page

WieWasWie, <https://www.wiewaswie.nl/>

Wiki for the Anglo-American Legal Tradition Website (WAALT), <http://aalt.law.uh.edu/>

Wiki Loves Monuments, <http://www.wikilovesmonuments.org/>

Wiki::Score, <http://www.wiki-score.org/doku.php>

Wikimaps Warper, <https://warper.wmflabs.org/>

Wikimedia Commons, <https://commons.wikimedia.org/>

WikiProject NARA, http://en.wikisource.org/wiki/Wikisource:WikiProject_NARA

Wikisource, https://en.wikisource.org/wiki/Main_Page

WikiTree, <http://wikitree.com/>

WikiWar, <http://wikiwar.net/>

Wiltshire OPC, <http://www.wiltshire-opc.org.uk/genealogy/>

Wir Waren So Frei, <https://www.wir-waren-so-frei.de/>

Women of Scotland, <http://www.womenofscotland.org.uk/>
 Women's Archive of Wales, <http://www.womensarchivewales.org>
 World Archives Project, <http://community.ancestry.com/wap/download.aspx>
 World Beach Project, V&A Museum, <http://www.vam.ac.uk/content/articles/w/world-beach-project/>
 World Memory Project, <http://www.worldmemoryproject.org/>
 World Service Radio Archive Prototype, <http://worldservice.prototyping.bbc.co.uk/>
 WorldConnect, <http://wc.rootsweb.ancestry.com/cgi-bin/igm.cgi>
 WorldGenWeb, <http://www.worldgenweb.org/>
 Worm Watch, <http://www.wormwatchlab.org/>
 WWI 3rd Bn CEF War Diary Transcriptions, <http://qormuseum.org/2012/11/21/update-on-our-transcription-project-for-3rd-bn-war-diaries/>
 Yad Vashem The Central Database of Shoah Victims' Names, <http://www.yadvashem.org/>
 Year of the Bay, <http://yearofthebay.org>
 You! Enhance Access to History (YEAH), <http://www.apex-project.eu/index.php/en/news-events/182-copy-of-pawliczek-article-248>
 Your Paintings Tagger, Public Catalogue Foundation (PCF), <http://tagger.thepcf.org.uk/>
 youryongestreet, http://web.archive.org/web/*/http://youryongestreet.omeka.net/
 YUMA / Annotorious, <http://annotorious.github.io/>
 Zoo Tools, <http://tools.zooniverse.org/>
 Zooniverse Dating,
<http://web.archive.org/web/20131209200826/http://dating.zooniverse.org/>
 Zooniverse Letters, <http://letters.zooniverse.org>
 Zooniverse, <https://www.zooniverse.org/>
 Zotero Commons, <https://www.zotero.org/support/commons>

Appendix B: Interview participants

Interviewees' experience location, age, years of experience with historical research, training, and the most relevant categories to describe their practice within the definitions used in this thesis.

'Practices' refers to their main historical activities rather than their specific research interests. 'Training' summarises participants' own assessments of their training. Participants marked with a * were also part of collaborative community history projects.

Code name	Age	Location	Years research	Training	Practices
Alison	30s	UK	12	Academic	Academic
Anne	30s	UK	11	Academic	Academic
Barbara	50s	US	25+	Academic	Academic and alt-ac role
Bob	50s	Europe	20	Self-taught	Family
Charles	70s	UK	5+	Self-taught	Local*
Claire	70s	UK	14	Academic	Family (former academic)
Daniel	30s	UK	5	Academic	Academic
Dorothy	70s	Aus	20	Self-taught	Local
Doris	70s	UK	20	Work-related	Biographical (family)
Ed	40s	UK	15	Self-taught	Family
Emily	30s	Europe	11	Academic	Family, local
Frank	20s	UK	5	Work-related	Local (professional)
Gina	30s	UK	8	Self-taught	Family, local
Helen	30s	UK	10	Academic	Academic, alt-ac role
John	70s	UK	50	Academic	Academic, local*
Karen	70s	Aus	25	Self-taught	Family*
Kathryn	60s	UK	45	Academic	Academic
Larry	60s	Aus	50	Self-taught	Family*
Martin	50s	Aus	30	Self-taught	Family
Nell	70s	Aus	5	Self-taught	Family
Oliver	40s	UK	4+	Self-taught	Family
Peter	60s	Aus	25	Self-taught	Family*
Rosemary	60s	UK	5+	Academic	Academic
Sarah	60s	UK	25+	Work-related	Biographical*
Theresa	30s	UK	15	Academic	Alt-ac role
Tom	60s	UK	18	Academic	Local*
Ursula	40s	Europe	10	Academic	Academic
Wendy	50s	US	15	Academic	Academic
Yvonne	30s	US	13	Academic	Academic

Appendix C: Interview and survey questions

This Appendix contains three research instruments. The first, *Interview questions (faculty, family and local historians)*, is the script for the semi-structured interviews with faculty, family and local historians. The second, Questionnaire for project stakeholders, was used in a small number of informational interviews. The third, *Online survey: Crowdsourcing and public participation in digital history*, contains questions posted in an online survey in June 2014.

Interview questions (faculty, family and local historians)

Thank you for agreeing to this interview... As you will have seen in the 'information for participants', this interview will take about an hour. We can stop at any time if necessary, and you can ask questions at any point. [If in work space, then explain observation process and discuss suitable task]. Do you have any questions now?

We'll start with some demographic questions, which are entirely optional, they will just help me group the results as I do the analysis.

Demographic information

What decade does your age fall into? [e.g. 30s, 40s.]

Gender [guess for self]:

Location [e.g. town or county]:

How comfortable are you with using computers? [Not at all, not very, ok, quite comfortable, very]

Highest education completed? Subjects studied?

What is your main occupation? [Is history part of your main occupation?]

Can you summarise in one sentence the type of historical research you do? [e.g. how would you describe it to someone at a party. And what's your definition of history?]

[For genealogists/family historians:] would you consider yourself a genealogist or a family historian? [e.g. Is your main interest in putting together names and dates, or understanding lives and collecting stories?]

How did you learn to do historical research? [e.g. formal training, friends or family members, reading websites]?

How much experience have you doing historical research [e.g. years practice since training]?

[If need an easy warm-up question...] How did your interest in history start? [What got you interested in history?]

Do you have any questions before we go on?

Ok, thanks for that. Now we'll move onto the more exploratory, conversational part of the

interview...

Background for their historical research

Can you expand a little more on the types of historical research you do? e.g. What sort of history would you consider your work to be? This might include: types of material/sources, schools of thought/approach, period/place, discipline, motivation [family research, academia, etc], motivation for/purpose of research [publication, personal archive, etc]?

Research processes

Thinking about times when you start researching a new line of inquiry e.g. new person, topic, place, or for new publication, could you give me an overview of the main stages you go through?

Thinking about a typical research session [e.g. one mentioned earlier], can you take me through the types of resources and tools you might use? I'm particularly interested in when and how you use resources e.g. databases, archives, books and tools e.g. methods of recording, checking, note-taking, copying, etc.

Are there any other things/resources you use in your research? [e.g. tools, books, databases, websites, catalogues, etc]

Are there any things you've tried or used in the past but stopped using?

How do you manage and organise your research data? What information do you keep about resources you've used? e.g. information to find it again, cite it, record its content; metadata about sources, hand-written or typed annotations or transcriptions, links, screenshots, print-outs...

Assessing resources

Thinking about times when you've discovered a new resource, whether a book, website, database, etc - how do you work out whether you'd use it? What qualities are important to you? [Particularly focus on questions of trust, authorship, authority; how do you determine the provenance, reliability and probable accuracy of digital resources?]

Is there a difference between resources you'd use and those you'd rely on or cite?

What might make you suspicious of a resource?

When you hear of a new digital resource, do you apply any additional checks, or variations on the methods already mentioned?

If you hear of a resource created by amateur historians or open to contributions from members of the general public, do you consider using it? [e.g. Wikipedia, ancestry.com, Transcribe Bentham] If so, you apply any additional checks, or variations on methods already mentioned?

What about blog posts or other social media?

Do you assess individual resources or posters [as in individual people] on a crowdsourced site, or judge the site as a whole?

[For biographical researchers of various sorts] What 'colour', information other than basic facts and dates do you look for, or record if you come across it?

Do you have any other comments on the resources you use in research, whether physical, digital; archival or amateur?

Use of mapping, geospatial techniques

On a different note, do you work with mapping, place names, or geospatial content, at all? If so, expand... [e.g. short-term tasks vs long-term goals?]

What's your goal in using geo data/mapping?

Can you give me examples of how you use place names or geographic locations in your research?

If you're using an old street, parish or place name, do you use historical maps, text indexes or other sources to cross-reference it?

At what scale or level of precision do you work? Does it ever vary?

In what format is your mapped/gelocated information stored? With other material?

Have you encountered issues with changes in place names changing, duplicate names, places disappearing, boundaries changing? Can you give me an example?

Do you record uncertainty or contingency in your geo-location or place names? e.g. do you record it differently, record sources of geo-spatial information, level of precision?

Are there things you would like to be able to do with mapping or place names that you can't currently do?

In an ideal world, what additional geospatial skills would you have?

In an ideal world, what things [tools, software, data] would you have access to?

Any final comments or thoughts on place?

Sharing research data

Who do you share your work with? What types of data [e.g. raw data, draft notes, interpreted results]. What determines when or how you do? If you do share, how?

Have you had any good results as a consequence of sharing?

On the flipside, who don't you share with? What types of data wouldn't you share? Is there anyone you wouldn't want to see your data?

Have you had any bad results as a consequence of sharing your research data?

Thinking about use of online resources, has there ever been a time when someone else's changes or corrections have affected you in a positive or negative way? [e.g. corrected transcriptions or place references]

Would you share your research data on a crowdsourcing/public site? Why, why not?
Under what conditions?

Do other researchers share their work with you? What determines when or how they do?

Are you part of any collaborative research [project or task]? If so, do you share any raw data? How do you manage sharing 'draft' data while its reliability is still uncertain? [e.g. labelling it by directory or file name, in notes, conversation, etc?]

How much does credit/attribution/authorship matter to you? For your own work, or for others?

Final comments, thoughts

Coming to the end... any final comments or thoughts? Anything you want to return to?

Is there a researcher with a contrasting approach [or whatever has come up in the interview] that you'd suggest I talk to?

Thank you very much for your time, it's been really helpful.

Questionnaire for project stakeholders

Thanks for agreeing to take part in my research! Your answers will help me understand more about the challenges and solutions for creating effective crowdsourcing and participatory projects in digital and history. If you've got thoughts to share that aren't covered by my questions, feel free to add any comments at the end.

It'd be really helpful if you answered as many questions as you can. If you need to check with others about particular answers, you can always fill in those responses in a follow-up email, and if you aren't sure how to answer a specific question, have a go anyway and include a note in your response so I can clarify further for you. Your honest evaluation will be really useful, so if there are any questions you'd prefer not to be quoted on (i.e. being identifiable as coming from you or about your project), put '[not for identifiable quoting]' at the start of your question.

It'd be great if you could get your responses back to me at m.ridge@open.ac.uk within a week. Even a partial response will be helpful, and you can always supply more information later. If you get stuck or if anything is unclear, please email me xxx@open.ac.uk

About the project

Tell me a bit about your project(s): (a link is fine)

Is participation/collaboration the main purpose of the site? Yes/no

What forms of participation, crowdsourcing, content enhancement or use, etc, does the project offer?

What's your role on this project?

What's the main goal your organisation is hoping to achieve with this project?

How does it relate to your organisational mission?

Do any partner organisations have differing goals for this project? What impact do you think that might have had?

Participants on your project

Who are the target participants for this project? (If there are other, less active audiences, how do they differ?)

To what extent did/do you have an existing relationship with these audiences?

Can participants take on different tasks as their skills or knowledge increase? Yes/no If yes, can you give examples?

Do participants have a platform for communicating with each other? (e.g. site forum, social media) Yes/no If yes, can you give examples?

Did any participants take on extra tasks or share information beyond the original project scope? Yes/no If yes, were you able to respond or use it as you would wish?

Did any participants undertake tasks or share content beyond those designed into the

project? Yes/no If yes, can you give examples?

If you promote volunteers to positions of responsibility, how does that usually happen?
Can you give an example of a conversation or event that prompted you to think of someone for a particular role?

Evaluating the project

How do/will you measure success?

To what extent do you feel the project achieved its goals (or is likely to)?

To what extent did the project successfully engage its target communities? Why and/or why not?

Have your participants or the project outcomes surprised you in any way?

Are there any aspects (e.g. functionality, tasks, content, communication) you think were particularly successful?

What unexpected barriers to participation do you think potential participants encountered?

And what would you do differently with the benefit of hindsight?

Further background information about the project

Which factors were important during the interface and task design process (or platform selection if you used an existing site)? e.g. technical, abilities of platform, organisational concerns, research on users, type of content?

Did you 'seed' the site with successful contributions before launch?

What kinds of user testing or review did you conduct before public release (if any)?

If participation requires registration, why is that so?

How did you market the project to potential participants? For bonus points: were you able to publicise it via traditional or social media? How did you select those channels? What difference do you think that made?

Does the project have any in-person events, exhibitions or other activities?

Do you feel the project allowed adequate resources for communication and outreach with participants?

What quality control or data validation methods does your project use?

Finally, is there anyone else I should talk to about the project?

Online survey: Crowdsourcing and public participation in digital history

The purpose of this survey is to gather data about crowdsourcing or participatory history projects for my PhD research, which is asking 'How do academic and family/local historians evaluate, use and contribute to crowdsourced resources, particularly geo-located historical materials?'.

Your answers will help me:

- * ensure that my coverage of projects is comprehensive,
- * learn more about how crowdsourcing participants learn new skills or get interested in a discipline, and
- * understand more about the challenges of creating effective crowdsourcing and participatory projects in digital and history.

You can find out more about my PhD research project at <http://www.miaridge.com/my-phd-research/> and you can contact me via <http://www.miaridge.com/contact/> or @mia_out. If you have any additional queries, you can contact my supervisors, Dr Elton Barker or Dr Deborah Brunton at The Open University.

Please note that by taking part in this survey, you are consenting to my holding and analysing the information you provide for use in my PhD thesis and/or any subsequent publications or presentations.

The three 'compulsory' fields are easy and take about a minute to answer. Four of my questions generally require more thought so the time to complete them will vary. This survey closes on June 20, 2014.

* Required

About your crowdsourcing or participatory history project

Project name *

Project URL *

Please tick the categories that describe the core purpose or content of your project *
If you're not sure, tick the box anyway or leave a comment and I'll work it out

- Crowdsourcing with the public (i.e. through an open call)
- For scholarly research
- Crowdsourcing with experts or small groups (sometimes called nichesourcing or community-sourcing)
- Using historical material (including documents, images, objects)
- For digitisation or record enhancement
- For public engagement
- To collect new material
- Using geospatial material (such as maps and location information)
- As a pilot project to learn more about crowdsourcing
- Other:

What impact did the project have on participants?

Can project participants learn or improve skills or experience through the project? If so, how?

This includes everything from unexpected learning opportunities or those designed into the projects goals.

Have project participants surprised you in any way? If so, how?

For example, did they take on additional tasks, produce creative responses to the material, learn new skills for the project?

What challenges did your project face?

What was the biggest challenge your project faced?

Please also feel free to comment on how you overcame it (or tried to)

Looking back on everything you've learnt from your project, what would you do differently in hindsight?

Any other comments?

If I can contact you to find out more about your project, please enter your details

below

Your email address (optional)

Your name and/title (optional)