



Open Research Online

Citation

Benatti, Francesca; Towheed, Shafquat; Blackburn-Daniels, Sally and Antonini, Alessio @TellMeWhatUReadingbot: the Multi-modal Strategy of the READ-IT Project for Collecting Experiences of Reading. In: Creative Intelligence: 35th ACM Conference on Hypertext and Social Media (HT '24), 10-13 Sep 2024, Poznan, Poland, ACM.

URL

<https://oro.open.ac.uk/99037/>

License

(CC-BY 4.0) Creative Commons: Attribution 4.0

<https://creativecommons.org/licenses/by/4.0/>

Policy

This document has been downloaded from Open Research Online, The Open University's repository of research publications. This version is being made available in accordance with Open Research Online policies available from [Open Research Online \(ORO\) Policies](#)

Versions

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding

@TellMeWhatUReadingbot: the Multi-modal Strategy of the READ-IT Project for Collecting Experiences of Reading

Francesca Benatti
Shafquat Towheed
francesca.benatti@open.ac.uk
shafquat.towheed@open.ac.uk
The Open University
Milton Keynes, UK

Sally Blackburn-Daniels
S.Blackburn-Daniels@tees.ac.uk
Teesside University
Hekla, UK

Alessio Antonini
alessio.antonini@open.ac.uk
Knowledge Media Institute
The Open University
Milton Keynes, UK

ABSTRACT

Hypertext readership has been placed under extensive scrutiny. At the same time, little to no work has been done on how hypertext systems can contribute to reading scholarship. This contribution reports on the experience of the EU Joint Programme Initiative for Cultural Heritage (JPI-CH) funded READ-IT project and the development of an interactive, multimodal system to crowdsourcing testimonies of reading experiences. The contribution describes articulating the research infrastructure aimed at federating research case studies across Europe in several languages and about the readership of different kinds in different periods. The infrastructure aims to broaden the understanding of reading in Europe. The infrastructure included a system to collect sources (testimonies of reading experiences) to be studied within and beyond the project. The infrastructure also provided a combination of online forms, postcards, and a chatbot to facilitate reflection on reading by members of the public. This contribution accounts for the experience of using this system in several countries and throughout a wide range of initiatives.

CCS CONCEPTS

• **Human-centered computing** → **Hypertext / hypermedia**; • **Applied computing** → **Arts and humanities**.

KEYWORDS

Reading Experience, Digital Humanities, Research Infrastructures, Chatbot

ACM Reference Format:

Francesca Benatti, Shafquat Towheed, Sally Blackburn-Daniels, and Alessio Antonini. 2024. @TellMeWhatUReadingbot: the Multi-modal Strategy of the READ-IT Project for Collecting Experiences of Reading. In *35th ACM Conference on Hypertext and Social Media (HT '24)*, September 10–13, 2024, Poznan, Poland. ACM, New York, NY, USA, 6 pages. <https://doi.org/10.1145/3648188.3675129>



This work is licensed under a Creative Commons Attribution International 4.0 License.

HT '24, September 10–13, 2024, Poznan, Poland
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0595-3/24/09
<https://doi.org/10.1145/3648188.3675129>

1 INTRODUCTION

Reading has been a defining activity for many human cultures and a powerful agent of change throughout history. During the last five decades, scholars in disciplines such as Book History, Literary Studies and Cultural History have worked to understand how reading occurs as an activity and process and how multiple factors shape it. These include external factors, such as socio-economic circumstances, and internal ones, such as the emotional and psychological aspects of the reader. Initially included in models such as Robert Darnton's 'communications circuit' [7], the study of reading changed with an increased focus on collecting empirical evidence from actual readers with the launch of the UK Reading Experience Database (UK RED) in 2006. UK-RED focused on historical readers from Great Britain and Ireland between 1450-1945, but the emergence of digital reading and the digitisation of cultural heritage since the launch of the project have introduced further complexities and made the need for a shared cross-disciplinary agenda even more urgent.

Reading Europe Advance Data Investigation Tool (READ-IT 2018-2021) was a collaborative research project on the cultural heritage of reading across Europe. READ-IT was funded by the Joint Programming Initiative for Cultural Heritage (JPI-CH) and included partners from four European countries (Institute of Czech Literature, Academy of Sciences, Prague; The Open University, UK, including the SME IN2; Utrecht University-DH Lab, Netherlands; CNRS-IRISA, Rennes and Le Mans Université-3LAM, France). The project also featured associate partners in France (University of Tours), Germany (Johannes Gutenberg University, Mainz), Italy (Università degli Studi di Milano) and the UK (Queen Mary University of London). The project integrated a multidisciplinary perspective connecting books and reading history, literature, computer sciences, information sciences and digital humanities. The backbone of the integration was its research infrastructure designed to collect and curate a vast corpus of readers' testimonies in multiple languages from the 18th to the 21st century [3]. The READ-IT infrastructure addressed challenges set by a large variety of reading representations emerging from a wide range of historical and contemporary sources, methods and questions.

The READ-IT infrastructure aimed to create a user-friendly, open-access, semantically enriched investigation tool to identify and share evidence about readers. The vision was of a federation of small research case studies integrated and interoperable thanks to using a common ontology for sources and findings. A key component was a module dedicated to collecting and curating sources,

text or images about reading. The project strategy included engagement events during which the experiences of contemporary readers were collected, to be aggregated into the project repository. The public engagement initiatives ranged from the sharing and study of personal reading archives and diaries, to interviews and ad hoc notes elicited through social media, a QR enabled set of postcards (translated into six European languages) and a chatbot. Both data collection and public engagement strategies for the READ-IT project had to be adjusted in March 2020 due to the COVID-19 pandemic, which ended any opportunities for face-to-face events during the remainder of the funded period. Face to face events were replaced with postal and online campaigns, with the chatbot and postcard packs deployed to reach people remotely. Despite this unexpected challenge, the result was an agile, richer multi-modal approach, which this presentation illustrates in the following sections. Previous publications have addressed the development of the two core ontologies and the bespoke annotation tool developed by READ-IT but they have not yet addressed the approach and technology behind the sourcing of reading experiences. Indeed, source curation is hardly a hot topic, particularly for historical and cultural heritage. However, little scholarship is yet available on how to engage with contemporary readers through modalities that are engaging and fit with the current practices. As such, this contribution aims to (1) report on the positive experience of READ-IT and provide a comprehensive vision of its overall technological infrastructure and (2) reflect on lessons learned and potential further development of its multi-modal engagement approach with readers. The rest of the paper is structured as follows. Section two provides background on reading studies. Section three describes the overall research infrastructure with a specific focus on source crowdsourcing. Section four reports on the results of using the infrastructure and chatbot as part of several research initiatives and the most relevant lessons learned. The presented insights report the perspective of reading scholars who used the technology first-hand in their activities. Finally, section five provides concluding remarks about the benefits and the potential future directions for exploring the use of our approach to the broader range of experiential studies.

2 ON THE STUDY OF READING

The history of reading has emerged in the last three decades as a lively field within the larger discipline of Book History. For a long time in the early and mid twentieth century, literary studies were focused on the opinions and experiences of the author. This approach to literary history in the West remained dominant until reader response and reception theories shifted the scholarly lens to the experiences of the reader [10] [11] [8] [5]. UK RED, a volunteer-led open-access database which collated documentation of the reading experiences of British subjects between 1450 and 1945, significantly changed the ways in which reading is being studied [14]. Contributions of reading experiences submitted to the database crossed class, gender, and education boundaries, allowing a number of scholars to begin to focus on discovering the reading habits and processes of ordinary, non-specialist readers from the past [13] [6].

2.1 Source of Reading Experience

The possible sources of reading testimonies vary depending on the period and location under scrutiny. For European readers, they include written sources such as letters and diaries, surveys and questionnaires, and oral sources such as interviews. For the most recent two decades, hypertext sources such as social media and blogs offer potentially rich pickings [2] [12].

Both these types of sources present challenges for researchers. Older sources such as letters and diaries are often unavailable in machine-readable format and require extensive perusing in libraries and archives to tease out any mention of reading. Newer digital sources can overwhelm the researcher through their sheer abundance and difficulty collating information from disparate platforms [4] [9]. Privacy and data protection may also need to be accounted for when surveying the writings of living persons.

3 READ-IT MULTIMODAL DATA COLLECTION

The study of reading may involve a wide range of sources, ranging from ‘legacy’ documents such as diaries, interviews, letters, and marginalia (notes at the margin) to contemporary ones like social media comments, forum threads and ratings. A second-order requirement about sources was the ability to collect, and curate collections elicited through engagement activities ranging from private documents to crowdsourcing campaigns with the ambition to enable their reuse in different case studies. This last point was crucial for addressing a key objective of the project of building a sustainable, long-lasting infrastructure that, through time, would collect and federate a wide spectrum of small case studies that together would support a broad study of reading in Europe.

3.1 READ-IT Research Infrastructure

The source management module is part of the larger infrastructure illustrated in Figure 1. The infrastructure includes a semantic layer, source curation and content curation modules.

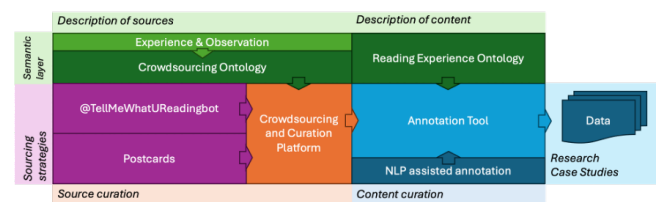


Figure 1: Components of the READ-IT research infrastructure.

The semantic layer provided the unified research language describing the reading experiences and the modality through which the experience evidence had been produced. For example, Irish poet Thomas Moore (1779-1852) kept a diary throughout most of his life where he noted down, among his busy social calendar, the books he was currently reading. He often mentioned whether he had purchased a book, borrowed it or received it as a gift. Similarly, he usually observed whether he was reading scholarly treatises alone in a library in order to research one of his publications or

reciting aloud the latest novels to entertain his family. Contemporary readers of digital media such as webcomics may mention sneaking a peek at their favourite series during class or hiding their mobile screens from parents or siblings who may object to their choices. All of these experiences can be called ‘reading’ but differ significantly in facets such as the location of reading, its duration, the length of the texts being read, the medium of reading and the ultimate purpose of the reading activity. The semantic layers have been the object of two publications reporting on the Reading Experience Ontology (REO) [2], the ontology design pattern Experience & Observation (E&O) and Crowdsourcing Ontology [1]. Relevant to this specific contribution is the Crowdsourcing Ontology used as the data model of the Crowdsourcing and Curation Platform and the interchange data schema of the chatbot.

E&O and the Crowdsourcing Ontology provide the language to harmonise the multimodal strategies and the different types of sources. Referring to the previous examples, Thomas Moore did not expect his diary to be published and sometimes expressed critical views on well-known authors and books. Webcomics readers generally leave comments anonymously or pseudonymously but do so on web pages that can be read by a worldwide audience. The two ontologies enable both types of sources to be annotated through the use of a shared semantic framework, which, in turn, enables cross-cultural studies. As such, the data collected through the crowdsourcing platform and the chatbot are pre-annotated based on the specific strategy and crowdsourcing campaign. For instance, the information about a specific question or the testimony being expressed in public rather than in private provides a critically important context to interpret the sources. The dedicated paper fully explains how this approach supports data interoperability and reuse at an epistemic level by making operational cues about the potential research value (and limits) of specific sources [1] based on the collection strategy and type of prompting used to elicit experience evidence from readers.

The content curation layer provided an ad hoc web annotation tool (see Figure 2) to classify text fragments based on REO concepts. The system included a rule-based NLP pre-annotator capable of identifying relevant sections about the reading experience in large documents. The annotation tool includes a SPARQL endpoint to extract Linked Open Data in REO CIDOC-CRM format (an implementation of REO as a CIDOC module).

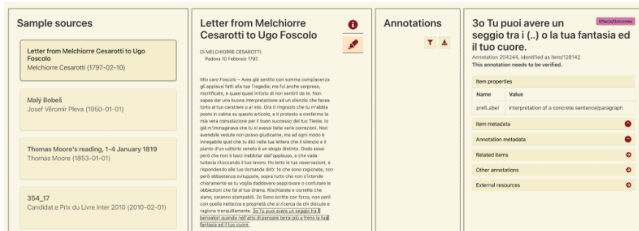


Figure 2: The annotation tool provides the ability to annotate text fragments using the REO concepts. In the screenshot, a letter to Italian poet Ugo Foscolo (1778-1827) and one fragment have been annotated as a reading effect.

3.2 Crowdsourcing and Curation Platform

The Crowdsourcing and Curation Platform (platform) was included in the original project proposal as a web portal for uploading media and creating curated collections to be used within research use cases. As part of the engagement strategy developed in the following period, the platform was complemented by a system of digital and material postcards that would collect answers to a selection of questions about reading. Material postcards (see Figure 3) included a QRcode linked to the platform to upload a picture of the filled postcards (see Figure 4 when not collected, mailed and digitised by the research team (e.g., for a study with prisoners or during workshops), see Figure 3.



Figure 3: A READ-IT postcard.



Figure 4: The QR code leads to the crowdsourcing platform, and contributions explored are curated into collections (masonry exploration page).

3.3 Chatbot Functionalities

The @TellMeWhatUReadingbot is a chatbot for Telegram and the project website. It was developed using the Flow.AI infrastructure and translated into six other languages (French, Dutch, Italian, German, Czech and Croatian). Its development followed an iterative process of further refining involving testing within the project team and in public events. For instance, the chatbot was showcased

during the leading conference on book studies and several national book fairs in Europe. The chatbot was also adopted (beyond the consortium’s scope) by a nationally-funded research project on Reading in Croatia to complement the face-to-face interviews with readers with follow-ups. The chatbot has three main functionalities:

- (1) *Introduction* to the project and relevant information. The intro is run just once on Telegram for each phone number. This component also provides basic instructions about best interacting with the chatbot, like using specific keywords or expressions to navigate its functionalities.
- (2) *Question-answer* flow delivers a question about reading and engages in a simulated conversation with the user about enriching his/her answer. Questions are sourced from the active crowdsourcing campaigns on the sources platform (filtered by closing data and language). Users can either answer the presented question or ask for an alternative. The simulated conversation follows the first answer with standard expressions like “very interesting, can you tell me more?”. This functionality reconstructs and logs the answer in the platform annotated with the specific campaign, the prompting question, language, date and time.
- (3) *Follow-up* flow engages after the answer is completed and checks if the user wishes to continue with another answer or if the chatbot can get in touch another day with another question. This system also recommends a blog article on the topic of the project. If the user agrees to the follow-up, the chatbot sends a message 24 hours later by triggering the question-answer flow. This mechanism works for Telegram users only but not in the project website implementation.

The Flow.AI system provides a workflow-based interface that incorporates a bag-of-words Natural Language Processing capability to interpret intentions (user answers) based on a set of example answers and a scripting environment to manipulate data and interact with the external APIs of the crowdsourcing platform (see Figure 5).

The chatbot used a similar approach to postcards but re-adapted to multiple engagements providing, e.g., the missing context about the research (see Figure 6). The initial set of questions was changed during deployment to allow multiple significant answers. For instance, a question about “the most influential book of your life” was reformulated from the question posed on the postcards: “a book that was significant in your life”. Exciting and unexpected was the need to adapt the chatbot’s behaviour based on its translation. While we expected the linguistic tone to be language-dependent, we did not expect the need to change the timing of messages. The speed of messaging (and pauses) had to be tailored alongside the communication registry; for instance, the English version is significantly slower than the Croatian version.

The last aspect worth noticing was the (overall) positive response to the simulated communication. While it was transparent to all users that the chatbot was just reacting with a random standard reply, this resulted in more articulated answers and the perception of a useful and intimate engagement.

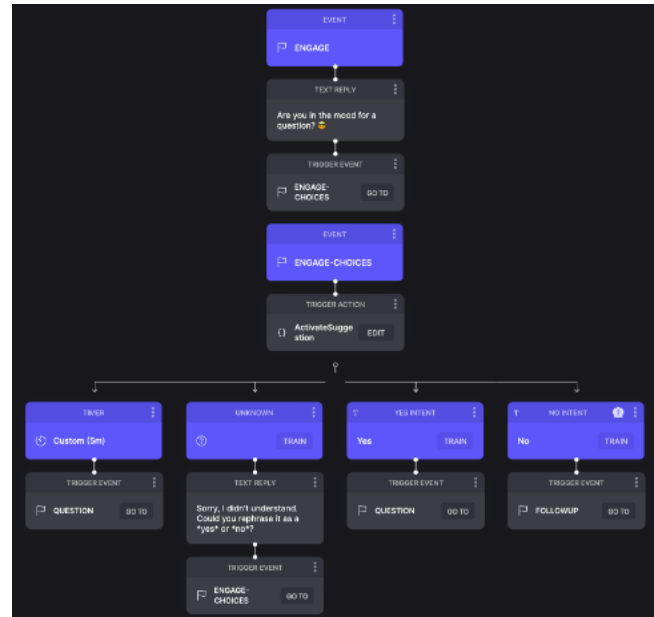


Figure 5: The development environment of Flow.ai, and the engagement flow triggered after introductions and 24 hours for users that opt in for a follow-up.

4 APPLICATION

One of the objectives of READ-IT was to provide resources that could enable a pan-European perspective on the study of reading. It was, therefore, essential to enable the collection of reading testimonies in multiple languages and their annotation. The @TellMe-WhatUReadingbot forms part of this multilingual approach and sits within an articulated public engagement strategy with the postcards and the crowdsourcing platform. In the following section, we provide examples of crowdsourcing for each type of use case we encountered, but first, we describe the use of the infrastructure and postcards, and how the chatbots were used as blueprints to instantiate ephemeral hypertext.

Figure 7 outlines a typical use of the infrastructure. Beyond the production of linked open data, the first benefit of the work on the Semantic Layer (see Figure 1) was to support the setup of research case studies, particularly in formulating research questions using the same definitions of reading across disciplines. The process can be quite articulated, but what is relevant for this contribution are the transitions from Phase 2 and 3 and 3 and 4. The implementation phase instantiates the two modalities of data collection (cards and chatbot) templates with the user questions designed based on the specific use case. Technically, the infrastructure takes the questions of active campaigns and makes the relative user questions available for data collection. Between phases 3 and 4, the collected data (the chatbot conversations and postcard replies) are made available to research teams to support their case studies and re-used by external groups or refurbished in teaching modules or other purposes [3].

The two transitions of phases produce two different types of hypertext between researchers and users and researchers and other groups re-using data.

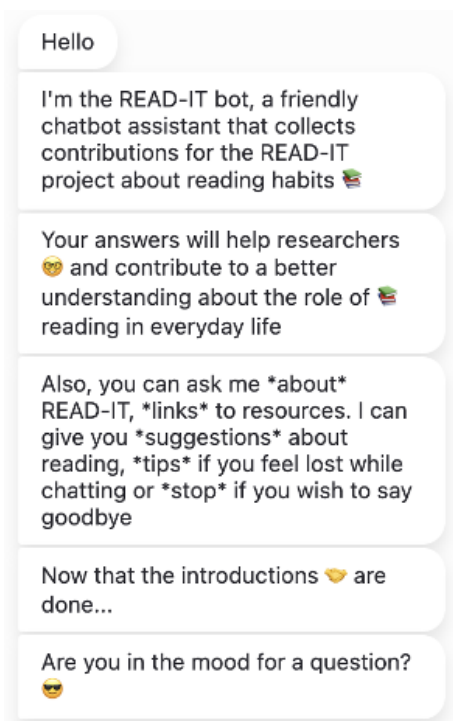


Figure 6: Chatbot’s introduction from the website popup chat that bootstraps the delivery of one of the active research questions.

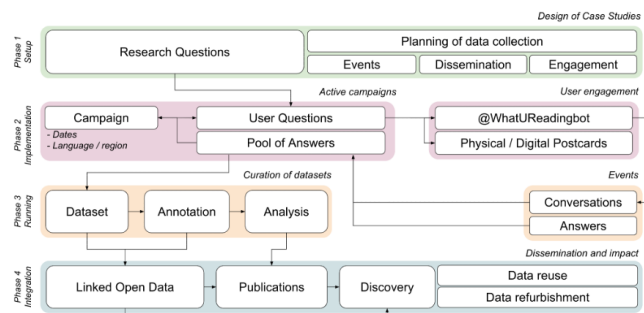


Figure 7: A typical workflow of activities is articulated in four phases, from designing research case studies to implementing and planning data collection campaigns and disseminating results.

The first is the instantiation of the data collection and conversation patterns with research input, open for a while in a specific location, language group, and a given setting. It is a form of ephemeral hypertext that is available until the researcher is content with the quantity or quality of interactions collected.

The second type of hypertext is the curated collection of data, including the ephemeral hypertext, the collected interactions and analysis in the form of annotations and publications. This hypertext provides a different and more subtle form of interaction focused on

supporting the discovery and exploration of sources and creating an assemblage of different works into new collections that would support research. This last interaction was supported by the infrastructure itself, which allowed both the creation of collections based on pieces of different data collection campaigns and the extraction and overriding of annotations from the annotation platform.

4.1 Use cases

Reading Groups. The @TellMeWhatUReadingbot was trialled within the reading groups run by The Verbal Arts Centre, a charity in Northern Ireland that employs storytelling and shared reading to reach marginalised communities. The @TellMeWhatUReadingbot was deployed with a group of 10-11-year-old pupils from a school in Derry/Londonderry. The chatbot was found to positively affect young students’ discussion and engagement and was praised for its novelty compared to other traditional approaches.

The READ-IT postcards were employed with a reading group in HMP Hull, a prison located in Great Britain. The group facilitator appreciated the offline approach provided by the postcards since UK prisoners are not allowed to access web-enabled devices. The trial proved to the READ-IT researchers the importance of avoiding assumptions on attitudes to reading. A number of prisoners, for example, could not respond to questions on their favourite book or memory of reading, as they had never experienced reading as positive. Both postcards and the @TellMeWhatUReadingbot were subsequently expanded with questions that allowed participants to report negative or problematic reading experiences.

Research projects. The Croatian version of the @TellMeWhatUReadingbot was tested with the postcards and crowdsourcing portal by the PoKUS project, University of Zagreb, which studied memories of reading in Croatia. The project staff employed the chatbot as an ice-breaking introductory tool, which was seen as a useful preparation before embarking upon more in-depth interviews. The READ-IT ontologies helped the PoKUS project expand its interview questions and connect to other European projects focused on reading.

Book fairs. The Italian version of the @TellMeWhatUReadingbot was trialled with participants at the BookCity Milano book fair in collaboration with Università Statale Milan. The participants were mostly undergraduate students in Book Studies and were intrigued by the bot and the crowdsourcing portal. Due to the small number of participants who had installed the Telegram app, the postcards on the crowdsourcing portal gathered a greater number of responses. The availability of multiple Italian-language questions was praised for its inclusivity.

4.2 Lessons Learned

The @TellMeWhatUReadingbot originated during the COVID-19 pandemic, which shut down several READ-IT engagement activities. The chatbot filled an unplanned gap, enabling READ-IT to maintain engagement and data collection during challenging times. The iterative approach to development also enabled Book History researchers to refine pertinent and engaging questions that could be asked from as wide a public as possible, continuing the expansion of reading studies research beyond scholars and researchers pioneered in previous projects [6].

Feedback from users and scholars on the @TellMeWhatUReadingbot has generally been positive. The young participants in the Verbal Arts Centre pilot were intrigued by its novelty and anonymity. The BookCity Milano students appreciated its multilingual capabilities, while the PoKUS scholars highlighted its potential to initiate discussions. Its limitations were also pointed out, such as the inability of the chatbot to tailor its follow-up questions to the context of longer, more in-depth interviews.

Researchers were initially unsure of how the @TellMeWhatUReadingbot would integrate with the other project tools but discovered through pilots and trials that it can complement the other established READ-IT data-gathering channels. The chatbot enables data collection through short, rapid interactions wherever and whenever participants choose, contrasting with more structured approaches such as interviews and book fairs. The multilingual platform provides all READ-IT core and affiliated partners with tools to be employed within their local contexts. Adopting the same questions and the underlying READ-IT ontologies enable the chatbot data to be integrated with those from other sources and transnational, transhistorical projects, moving scholars one step closer to understanding the history and present of reading in Europe.

5 CONCLUSIONS

Hypertext readership has been placed under extensive scrutiny. At the same time, little to no work has been done about how hypertext systems could contribute to reading scholarship. Intuitively, any interaction can bring to light some aspects of what readers think, want, like, believe, or just decide. In this contribution, we presented the experience of using a multimodal approach designed to provide an intimate experience to readers through a postcard and a chatbot. The chatbot specifically provided a very simple way to choose what theme interested readers by cycling between questions designed by researchers and digging into their experience by playing a simulated conversation with an interested chatbot. This system created a new type of reader, the researchers, who had access to anonymised conversation transcripts.

The multimodal approach to crowdsourcing provided a more comprehensive range of opportunities and new issues about the harmonisation of sources. It is worth highlighting how the modality through which experiences are elicited matters for their interpretation. In this regard, the different modalities had to be documented in relation to the questions asked, the time, location, and method for the benefit of the reader-researcher who did not directly collect the testimony and may use a mix-bag of differently sourced data.

As a last remark, there is still an untapped potential for the HT systems and the different practices of the community developed around, e.g., debugging and testing narratives by looking into readers' journeys. This contribution reported a particular take on the topic resulting from an incidental need: the COVID-19 pandemic. A thorough survey of these practices through these lenses would greatly benefit and support the adoption of the HT paradigm in book studies of contemporary readers and embed a most needed structural monitor-support mechanism to read into part of HT systems.

ACKNOWLEDGMENTS

This work was partially supported by Reading Europe – Advanced Data Investigation Tool (READIT), a project funded by the JPI Cultural Heritage under the European Union Horizon 2020 Research and Innovation programme (grant agreement No 699523).

REFERENCES

- [1] Alessio Antonini, Alessandro Adamou, Mari Carmen Suárez-Figueroa, and Francesca Benatti. 2023. Experiential Observations: An Ontology Pattern-Based Study on Capturing the Potential Content within Evidences of Experiences. *ACM Journal on Computing and Cultural Heritage* 16, 3 (2023), 1–30.
- [2] Alessio Antonini, Mari Carmen Suárez-Figueroa, Alessandro Adamou, Francesca Benatti, François Vignale, Guillaume Gravier, and Lucia Lupi. 2021. Understanding the phenomenology of reading through modelling. *Semantic Web* 12, 2 (2021), 191–217.
- [3] Francesca Benatti, François Vignale, Alessio Antonini, and Edmund King. 2023. Reading in Europe—Challenges and lessons learned from the case studies of the READ-IT project. *Digital Scholarship in the Humanities* 38, 2 (2023), 477–481.
- [4] Karen Bourrier and Mike Thelwall. 2020. The Social Lives of Books: Reading Victorian Literature on Goodreads. *Journal of Cultural Analytics* 5, 1 (2020), 1–34.
- [5] Roger Chartier. 1995. *Forms and meanings: texts, performances, and audiences from codex to computer*. University of Pennsylvania Press.
- [6] Stephen Colclough and Edmund GC King. 2019. Readers: Books and Biography. *A Companion to the History of the Book* (2019), 157–171.
- [7] Robert Darnton. 1982. What is the History of Books? *Daedalus* (1982), 65–83.
- [8] Kate Flint. 1995. *The Woman Reader*. Clarendon Press.
- [9] Danielle Fuller and DeNel Rehberg Sedo. 2023. *Reading Bestsellers: Recommendation Culture and the Multimodal Reader*. Cambridge University Press.
- [10] Wolfgang Iser. 1974. *The Implied Reader: Patterns of Communication in Prose Fiction from Bunyan to Beckett*. Johns Hopkins University Press.
- [11] Hans Robert Jauss. 1982. *Toward an Aesthetic of Reception*. University of Minnesota Press.
- [12] Federico Ponzola. 2021. *Digital Social Reading*. MIT Press.
- [13] Jonathan Rose. 2008. *The Intellectual Life of the British Working Classes*. Yale University Press.
- [14] The Open University. [n. d.]. *UK Reading Experience Database*. <https://www.open.ac.uk/Arts/reading/UK/about.php>

Received 20 February 2024; revised 12 March 2024; accepted 5 June 2024