

RESTORATION AND REPURPOSING OF DH LEGACY PROJECTS

The institutional funding system of Digital Humanities (DH) is usually devoted to the creation of new projects, creating a recurring problem of unsupported legacy projects whose material cost of upkeep depends on the voluntary contributions of institutions and individuals. The lack of resources to invest in “remedial” actions [10] pushes DH projects towards outdatedness. Additional funding success delays this process by introducing extra resources but, simultaneously, it fast-forwards obsolescence by advancing the field.

Indeed, the impact of a DH project can be considered as the ability to establish as common knowledge what was once innovative and cutting-edge by fulfilling its research questions. In this scenario, managing successful DH projects requires addressing competing issues related to the preservation of their integrity [1] (i.e. consistency of data, questions and vision) and of their role and purpose (i.e. their use in the field).

The management of legacy systems has been widely studied from a technical perspective [2,3,4,5,6,7,8], e.g. cost/value [9], approach to integration [1,9], change of use [1] and archiving [3,4]. Rather than presenting technical solutions, this contribution focuses on the rationale for defining an approach through human, financial and political perspectives [2].

The issue of legacy is not one of data formats but principally a cultural one that we analyse from two distinct approaches:

1. **The “restoration” approach**, implementing remedial actions [10] that “update” the project to new contexts to preserve its function and role (e.g. extending its data structure to address new questions)
2. **The repurposing approach** [1], implementing actions that rethink the value of the project by finding it new purposes, functions and roles in new contexts (e.g. defining new questions to be addressed with its existing data).

At stake in the two approaches are the **integrity** and **identity** of the project. A project’s integrity is the logical and historical connection between its origin, output and outcomes. A project’s identity is the meaning or role it has within the community of people involved. In this scenario, we argue that addressing a project’s legacy should take into account:

1. **The project's vision**, research questions and target “knowledge gap”; the project creators' motivations and aims; the funding bodies' goals and priorities
2. **The project's practices**, orchestration of people, organisations and tools, operational limits and constraints
3. **The project's knowledge**, research data and outputs, correlated research activities, answers produced and outcomes (e.g. new projects, scholarly research, education, impact on the field).

The contribution then discusses a real case, the UK Reading Experience Database (RED), as emblematic of the challenges of managing legacy DH projects. RED has had a long history, repeated funding successes and significant visibility in Book History scholarly literature. It was devised by Simon Eliot in 1993, first implemented in 1996, published on the web in 2007 and finally closed to new submissions in 2018 [11]. RED's vision was to advance research in the history of reading by establishing a new methodology based on empirical evidence [12]. RED's practice established strong synergies between researchers, students and volunteers for the distributed acquisition and curation of evidence of reading. The RED contribution form's structured approach to knowledge encouraged data inputters to pay close attention to the contexts and agents involved in the reading experience: who was reading, what was read, and when and where the act of reading took place [13].

RED data has been successfully converted from a legacy custom relational database to linked open data. Still, RED is a legacy project because of what its data expresses about reading experiences: a now-outdated vision established more than twenty years ago, which has now become embedded in the DH community through successful activities, follow-on projects, publications, research and training initiatives [14,15,16,17]. RED is both a vast database and the centre a wide network of collaborations; therefore addressing its legacy is not a trivial decision.

1. **The Repurposing of RED:** Repurposing RED's vision means, for instance, rethinking its role from research infrastructure to an educational resource. Consequently, RED's practices could be reframed as a playground for DH students, providing an environment for training and annotation evaluation. The knowledge produced and encompassed in RED could document the history of DH methods or become a training set for machine-learning algorithms.

2. **The Restoration of RED.** Restoring RED's vision means, for instance, incorporating in RED new approaches currently required by funding bodies (e.g. collaboration with data science) and current research priorities within Book History, changed considerably since 1993. Consequently, the study of sources could be combined with machine-learning and natural-language processing tools not included in the original structure of RED. Finally, new research questions such as the effects of reading [18] and the multi-modality of reading on new media [19] could be addressed.

With a repurposing approach, the integrity of the dataset could be preserved by relinquishing RED's role as a research project. With a restoration approach, RED's role as a research project could be preserved through the entire re-curation of its data, the complete re-development of the tool ecosystem to include automatic steps and the entire reassessment of its value as a research resource in light of current DH and Book History research agendas. Unsurprisingly, to keep a project's role we must face the cost of adapting to the new context, while to keep its form, we must search for a new purpose.

On a more general level, there is a question about how to preserve the "human legacy" of RED, e.g. the network of collaborations, student volunteers and contributors engaged. A DH project is a Cultural Artefact, and therefore its historical context can guide the re-tuning of its role as the context changes or the search for new purposes compatible with the values and vision of the social system of the project.

As a final remark, we hope these questions can elicit a broader discussion about new and future DH projects and how we could design for their legacy, e.g. the new Reading Europe Advanced Data Investigation Tool.

Alessio Antonini (alessio.antonini@open.ac.uk), The Open University, United Kingdom, Francesca Benatti (francesca.benatti@open.ac.uk), The Open University, United Kingdom and Edmund King, The Open University, United Kingdom