Essay Review

Notes on the Reviewing of Learned Websites, Digital Resources, and Tools by Anna-Luna Post

and Andreas Weber

According to advocates of the “computational turn,” the job description of historians of science is about to change radically. Instead of rooting their research in carefully selected corpora of primary source material, and then using these sources to study how scientific knowledge has been produced, spread, and contested in a specific context, historians of the twenty-first century can capitalize on a variety of digital resources and tools to sketch a much bigger picture.¹ Doing research “digitally” has, as the wisdom goes, several advantages. First, it allows historians to give voice to (previously) marginalized groups in historiography. Second, digital tools allow historians of science to develop new grand narratives by basing their studies on fine-grained data mining, metric and visual analyses of large corpora of digitized primary source material. Third, the digital knowledge commons allow historians of science to work collaboratively. This enables a productive dialogue between professional historians, citizen scientists, and an interested wider public. Seen from this perspective, the computational turn entails the potential to pose and answer new research questions and to produce new kinds of historical narratives. Moreover, it can, as has been argued, also significantly change the way in which history of science scholarship is carried out.

As a result of this optimistic view, digital projects are gaining more prominence and drawing more funding than ever before. Yet unlike for journal articles and monographs—the traditional venues in which academic research is shared and debated—we currently have very limited ways of reflecting on the merits of either individual projects or the rise of computational methods in the history of science in general. In terms of reviewing, digital projects reside in a no-man’s-land, lacking the infrastructure used to assess the value of monographs or journal articles. But to cite George Sarton’s still-relevant “Notes on the Reviewing of Learned Books”:

Anna-Luna Post and Andreas Weber share the first authorship of this contribution.

Anna-Luna Post is a Ph.D. candidate at Utrecht University, where she studies the fame, reputation, and credibility of Galileo Galilei. Her research combines insights from cultural history and history of knowledge and science with a specific focus on early modern Italy and the Dutch Republic. She has written the introduction to the first English translation of Caspar Barlaeus’s oration *Mercator Sapiens* (Amsterdam, 2019). Trans 10, 3512 JK Utrecht, Netherlands; a.l.post@uu.nl.

Andreas Weber is an assistant professor at the University of Twente with a special interest in the long-term development of science and technology and digital biodiversity heritage collections. Most of his research focuses on the history of natural history and chemistry in colonial Indonesia and Europe in the long nineteenth century. He is also the editor of the Brill book series “Emergence of Natural History (ENH).” P.O. Box 217, 7500 AE Enschede, Netherlands; a.weber@utwente.nl.

¹ See, e.g., Jo Guldi and David Armitage, *The History Manifesto* (Cambridge: Cambridge Univ. Press, 2014), p. 93; and Manfred D. Laubichler, Jane Maienschein, and Jürgen Renn, “Computational Perspectives in the History of Science,” *Isis*, 2013, 104:119–130.

“learning cannot progress without appreciation or criticism.”² In line with its founder’s statement, *Isis* has therefore decided to start reviewing digital projects on a regular basis—and this essay offers a starting point for the assessment of such projects in the field of history of science.

Its aim is twofold. First, it presents a broad (albeit not exhaustive) overview of recent digital projects, which allows us to see the extent to which the computational turn has shaped the profession so far. From our brief sketch-map, a division into three categories of digital projects arises, which we believe will be useful as a framework for future reviewers. The first of these categories consists of initiatives that offer resources to historians of science (i.e., online archives of manuscripts and illustrations). The second category is made up of tools that enable metric analysis of word frequencies in large corpora, network visualization, or transcription and annotation of sources. The third category comprises those initiatives that can best be called digital knowledge commons. They allow users to interact with primary source material and digitized artifacts and to share their insights online for the benefit of the community as a whole.

The second aim is to present more general guidelines that may serve as the basis for future reviews of individual projects. Each of the three categories comes with its own set of strengths and weaknesses, which merit particular attention. Yet overlap between the categories also exists, and many digital projects are actually hybrid in form. For this reason, we believe that it is possible not only to come up with specific touchstones for each category but also to present more general guidelines that may serve as the basis for informed reviews of learned digital projects. Before doing so, we discuss the three categories in somewhat greater detail.

RESOURCES

Over the last decades, a number of primary sources have been made available online to historians of science. Sources presented on the Web include manuscript collections of, among many others, Isaac Newton, Charles Darwin, Alfred Russel Wallace, and Albert Einstein. A comprehensive collection of nature-related publications and manuscripts by different authors is offered by the Biodiversity Heritage Library (BHL). Some websites also give access to collections of instruments, specimen collections, and visual sources: the Instruments & Innovation section of the Science History Institute in Philadelphia and the data portal of the Natural History Museum in London are good examples. Many of these websites provide access to digitized books, periodicals, archives, and object collections in the form of scans, images, or full text, and the historian is free to pose his or her own questions to the data. Owing to the accessibility and searchability of such digital collections, primary source material and scientific artifacts can easily be studied from university offices, from coffee shops, and even from home. In many cases, digital access to such resources saves historians of science time and travel costs.

However, the high costs involved in creating and maintaining such digital repositories have led to foregrounding publications and collections of “Great Men of Science” from the West. As their works and correspondences were often already easily available in different media, the added value of these projects for the profession as a whole is not always clear. Do these sites really encourage the development of new types of research questions, or do they simply offer a way of organizing and providing access to primary source material different from printed resources? To make available online, in particular, the voices of women, of non-European actors, or of actors whose voices have not been recorded on paper is a major challenge digital projects face—but one that would have enormous impact when overcome.³

² George Sarton, “Second Preface to Volume 41 of *Isis*: Notes on the Reviewing of Learned Books,” *Isis*, 1950, 41:149–158, on p. 149.

³ See, e.g., Kapil Raj, *Relocating Modern Science: Circulation and the Construction of Knowledge in South Asia and Europe, 1650–1900* (Basingstoke: Palgrave Macmillan, 2007).

An example of a project that does open the road to new research questions is the Casebooks Project, which—as we are writing this in August 2018—presents almost fifty thousand digitized medical cases from the records of the early modern astrologers Simon Forman and Richard Napier. Along with information on Forman and Napier themselves, the Casebooks Project provides visitors with data on doctor visits, various kinds of ailments, practices of record keeping, and patients' responses to medical practitioners.⁴ In using Forman's and Napier's collections as a starting point for the digital reconstruction of daily practices of early modern medicine, the Casebooks Project is a prime example of offering meaningful access to relevant collections in which marginalized actors also have a voice. Unlike many other resources, which simply give access to digitized source material, the Casebooks Project encourages and prompts historians to reflect critically on the role of physicians, on their patients, and on daily medical practice not only in the seventeenth century but also more generally in the historiography of medicine and science.

TOOLS

The second category comprises tools set up to enable historians of science to carry out new forms of analysis in the context of digitized archives and collections. Although many of these tools have not been developed specifically for historians of science, they are applicable to research questions raised in the field. In contrast to resource-oriented websites, in many cases historians have to supply their own data sets. The tools then allow them to study and analyze their data in new ways.

The network analysis tool Nodogoat, for instance, allows users to curate, explore, and visualize large data sets in a Web-based environment.⁵ In the context of the history of science, such a tool can, among other things, be used to visualize the development of correspondence networks over time and space. Historians can also use the tool to study when, and why, a certain scholar was excluded from networks of learned exchange.⁶

Other tools focus on disclosing the content of digitized source material. Projects that rely on automated handwriting and image recognition are an important subcategory in this domain. In the context of the READ (Recognition and Enrichment of Archival Documents) project, researchers have managed to produce a searchable full-text transcription of the botanical manuscript *Historia de las plantas* by the Spanish naturalist Bernardo de Cienfuegos (1580–1640).⁷ The multimodality of natural history collections has also triggered researchers to use more visually oriented systems such as the MONK system.⁸ MONK is a handwriting and image recognition system that actively learns from the input of experts and volunteers to recognize and connect words (e.g., place-names, species names) and visual structures (e.g., ears) in digital collections.⁹ Thanks to live training, historians and citizen scientists can have a strong impact with regard to the content of digitized collections that is made searchable and interlinked and how that is being accomplished. Precisely because of its user-centeredness, MONK has the potential to provide his-

⁴ For an in-depth discussion of the project see Lauren Kassell, "Paper Technologies, Digital Technologies: Working with Early Modern Records," in *The Edinburgh Companion to the Critical Medical Humanities*, ed. Anne Whitehead et al. (Edinburgh: Edinburgh Univ. Press, 2016), pp. 120–135.

⁵ <http://nodegoat.net/> (accessed 22 June 2018).

⁶ See Tobias Winnerling's fascinating project "Getting Forgotten within the World of Learning, 1700–2000."

⁷ <https://read.transkribus.eu/> (accessed 22 June 2018); and A. H. Toselli, L. A. Leiva, I. Bordes-Cabrera, C. Hernandez-Tornero, V. Bosch, and E. Vidal, "Transcribing a Seventeenth-Century Botanical Manuscript: Longitudinal Evaluation of Document Layout Detection and Interactive Transcription," *Digital Scholarship in the Humanities*, 2018, 33:172–202.

⁸ A. Weber, M. Ameryan, K. Wolstencroft, L. Stork, M. Heerlien, and L. Schomaker, "Towards a Digital Infrastructure for Illustrated Handwritten Archives," *Digital Cultural Heritage*, ed. Marinos Ioannides (New York: Springer, 2018), pp. 155–166. See also www.makingsenseproject.org.

⁹ L. Schomaker, "Design Considerations for a Large-Scale Image-Based Text Search Engine in Historical Manuscript Collections," *Information Technology*, 2016, 58(2):80–88.

torians of science with important groundwork for their historical analyses of natural historical and other digitized collections.

Of course, tools such as Nodegoat or handwriting and image recognition systems such as MONK are not free of tensions. Since they are often developed in the context of specific research projects, their long-term availability and maintenance can be guaranteed only by applying new forms of funding mechanisms, which often involve private parties. Moreover, the design of tools often depends on the interests of a specific group of users at a given moment in time, which in turn influences future research. Ideally, in order to guarantee their sustainability over time, tools should be designed with enough flexibility that they can co-evolve with historians of science and their research. Then the high initial investment is not lost but may be turned to yield new profits, even after the original project has come to an end.

DIGITAL KNOWLEDGE COMMONS

Digital knowledge commons comprise projects that allow those interested to contribute to them and provide the interface to enable such collaboration. The data—which may vary from early modern annotated books to nineteenth-century ships' logs and illustrations in Victorian books on nature—is provided by the project, and users are invited to help “unlock” the material by annotating, tagging, or extracting elements of the collection.¹⁰ In the Old Weather Project, for instance, users are asked to extract information on the weather from ships' logs, thus compiling information useful for the reconstruction of climate change over the centuries.

In the same fashion, users transcribe and translate early modern marginalia in the Annotated Books Online project to unveil early modern reading practices. Especially useful is the project's collaborative character: users interact not only with the material but also with each other, as they can comment on earlier transcriptions or translations and replace them with improved versions if necessary. This characteristic is shared by the Science in the Making project of the Royal Society, which presents users with a broad spectrum of material relating to that body.¹¹ The comment function allows researchers from all over the world to collaborate, compare, and discuss material. It is an important step away from sharing only published results and provides a useful model for the exchange of knowledge and expertise in earlier stages of research, prior to publication.

While the interactive character of these projects is a great strength, there are three important challenges for this type of project. First, their quality is difficult to guarantee, as they open up modes of collaboration with the lay public. Second, who can claim the results of these projects? The project coordinators or the contributors? Do results “count” if they remain unpublished in more traditional venues? A better conceptual understanding of the value of the results of these projects is therefore needed. A third and related challenge is the strong tension between the collaborative character of these projects and the growing emphasis on individual and institutional achievements. We hope that the latter will not triumph over the former, as we believe that these projects' potential for sharing knowledge and ideas before publication is an important asset.

FINAL REMARKS

This account shows that different categories of projects each have their own strengths and weaknesses and that they each make good on different promises of the computational turn. Online archives have the potential to make new source material available, which might lead

¹⁰ See, e.g., the Annotated Books Online project, <http://www.annotatedbooksonline.com> (accessed 13 June 2018); Old Weather, <https://www.oldweather.org> (accessed 13 June 2018) (ships' logs); and Science Gossip, <https://www.sciencegossip.org/#/> (accessed 13 June 2018) (illustrations in Victorian books on nature).

¹¹ <https://makingscience.royalsociety.org/s/ts/page/welcome> (accessed 22 June 2018).

to new questions and new historical narratives. Digital tools enable us to pose new questions and to do research differently. The same is true for digital knowledge commons, which allow researchers to collaborate fruitfully before publication. In the preceding sections, we have outlined the strengths, weaknesses, and challenges for the three categories we distinguish. We hope these will function as the first basis for future reviews of digital projects. In the remainder of this essay, we present several further guiding principles that we believe are indispensable for such reviews, whether published in *Isis* or elsewhere.

Each website, and each category of websites, deserves to be judged on its own merits, with attention to its specific, individual goals and aims. Reviewers would do well to pay attention to what source material is used, how it is presented, what the goals of the “author(s)” of the website are, and whether those goals are fulfilled. Similarly, reviewers might discuss the accessibility of a website, its interface (just as one might briefly discuss the table of contents and the index of a book), its relation to other projects, and what the credentials of its “author(s)” are.

These reviewing principles do not differ greatly from those put forward almost seventy years ago by George Sarton. This suggests that we might see websites as simply a new way of advancing learned knowledge: rather than publishing research results in books, we now publish important parts of our findings online. Yet, since the computational turn seeks and promises to reshape historical research, rather than merely to present its results through a new medium, we also think it is useful to advance three more general guidelines for judging digital projects. In our view, these are helpful touchstones to assess the extent to which projects contribute to this reshaping of the field. Eventually, they may even come to serve as useful handles for the design of new digital research projects. The principles are related to the three main promises outlined in the introduction to this review, as well as to what we believe is the most important added value of the computational turn. We believe that this lies in the capability to link different projects and create a sustainable, interoperable landscape of digital initiatives. The three main guidelines therefore read as follows:

- Does the project stimulate users to ask new research questions; does it present new ways of finding knowledge; does it open the road to new source material?
- Does the project enable meaningful collaboration or exchange between scholars, possibly before publication?
- Is the project set up in a sustainable and transparent manner, allowing scholars to study, use, or enrich its content and code in order to create an interoperable landscape of digital initiatives?

Together, we think, these principles will help overcome the risks involved in the computational turn and make good, to the largest possible extent, on its promises.