Rethinking History
Vol. 8, No. 2, June 2004, pp. 293–301

Routledge
Taylor & Francis Group

History and the Second Decade of the Web[1]

Daniel J. Cohen

More than ten years of experience with the web has allowed us to understand what the medium does well and what it does poorly, and how we may improve online historical efforts so they capitalize on the web's strengths while avoiding its weaknesses. This essay explores three possible ways to advance digital history: interaction between historians and their subjects, interoperation of dispersed historical archives, and the analysis of online resources using computational methods. Thinking about such possibilities raises important, age-old questions about how we should preserve and chronicle the past.

Keywords: Digital history; Data-mining; Interoperability; XML; Archives

With the World Wide Web now in its second decade of existence, and with the euphoria of the dot-com era now well behind us, it is a good time to reflect on what historians have been able to do in this still immature medium and how in the next decade of its existence we can better make use of it. Generally a conservative bunch in terms of the adoption of new technology (if not in political inclination), historians have, mostly for better but surely on occasion for worse, incorporated the medium into their work over the past ten years. For the most part using the web—it seems appropriate in our post-bubble sobriety to drop the grandiose alliterative phrase as well as the capitalization—has meant the posting of materials for courses, exhibitions, independent work and collaborations, as well as the communication of news and views from all corners of the discipline. Websites have flourished on almost every conceivable historical topic, created by historians from within and beyond the academy.

To be sure, this represents a tremendous step forward. As many have noted, the marginal cost of reaching another person with these electronic

ISSN 1364-2529 (print)/ISSN 1470-1154 (online)
DOI: 10.1080/1364252041000168395O

materials is almost zero, which represents a great advance over the physical world of paper (Negroponte 1996). With the web a hundred interested colleagues can access an online essay as cheaply and simply as a dozen. Moreover, the structure of the web allows interested parties to access that essay from any internet-connected computer, at any time, and to search the text for phrases or keywords. Creator and audience alike may link this piece to others on the web, catalogue and copy it, and even print it if so desired. By now most of us are familiar with these significant advantages that come with digital, rather than physical, historical resources.

One senses, however, that the medium of the web has not been exploited to its fullest if the best we can say about historians' use of this highly advanced computer network is that it has become a giant, global fax machine, faithfully reproducing and distributing copies of historical documents (primary and secondary), related commentaries and professional missives. And because of the openness of the medium—and the always tenuous relationship between the professoriate and the large population of lay historians and the general public interested in history—many historians have found the web to be a mixed blessing: prolific but unmediated, powerful but untamed, open to all but taken seriously by few. When my colleagues Roy Rosenzweig, Michael O'Malley and Andrew McMichael first surveyed the historical web in 1995, they were impressed with the number of websites on FDR: 49. Today that number stands at 628,000—and one would venture to guess that not all of them are compelling, accurate, and the product of the kind of deliberative thought and research that our profession honours (McMichael *et al.* 1996). Online history has not—and indeed may never—rival the gold standard of the book, replete with its physical and (hopefully) intellectual heft, peer review, and centuries of technical improvements such as the footnote, table of contents and index (Grafton 1997; O'Malley 2000).

Yet historians will never know if this apparent inequality of media is inherent in the natures of the web and print unless we work harder to capitalize on the advantages of the web and lessen its disadvantages, in the process creating new forms of history that can only exist online. Surely it is ridiculous to ask this new medium to reproduce the exact successes of paper, just as it would be silly to denigrate Ken Burns' *Huey Long* for being less comprehensive than the biography by T. Harry Williams. Both worthy examinations of a crucial figure in American history, these two efforts had different goals that were commensurate with the strengths of their intended medium. Furthermore, just because the majority of websites on FDR are not worth the electrons they vibrate should not condemn by association the thoughtfully or creatively produced minority.

So what have we learned about the strengths and weaknesses of this young medium, and how can we maximize the web's advantages and minimize its disadvantages to create the best forms of online history? It is clear that the web is very good at connecting large numbers of people regardless of physical location; allowing for synchronous and asynchronous communications between those people and in general facilitating collaboration; enabling the storage and transmission of huge amounts of information; encoding information in such a way as to make it rapidly searchable by machines; removing barriers to publication and providing space for unbounded publications and parallel publications of a single entity (such as translations); and allowing for revision and updating. The web's disadvantages include its lack of stability and persistence; its difficulty in displaying human-friendly, legible formats for text, especially for the long-form narratives that have been the hallmark of history for generations; its failure to provide reliable clues that help one discern the real from the fake and the good from the bad; and its lack of the kinds of checks that in the physical world lead to an often beneficial winnowing of available information.

Examining this list of pluses and minuses, it should not be surprising that many in our profession have regarded the web with suspicion. We spend much of our time as readers, and we like our documents stable, authentic, persistent and legible. Moreover, when even web designers who have come out of academia such as Sarah Horton and Patrick J. Lynch note that it is best to break text on the web into 'small chunks', defenders of the traditional historical essay and book have had a right to be worried (Lynch & Horton 2002). In addition, high expectations for both 'e-books' and new, non-linear forms of digital writing ('hypertext') have thus far proven unwarranted. Most of us still like our text linear and coherent, and in crisp black fonts on a white page.

Although technologists constantly praise computers as 'universal machines', and while it is true that they are remarkably flexible and powerful, this does not mean that we should use them for every kind of historical effort or production. We can truly begin to think about history in the second decade of the web, and to postulate new, more appropriate and ideally more enlightening forms of online history when we begin to look beyond the distribution of documents (including essays, articles, exhibits, news and messages) and consider instead the collection, interrelation and exploration of those documents. Much of the problem here is conceptual: from the beginning of the web historians have largely discussed *nouns* such as web pages and websites, rather than *verbs* such as searching, sorting, gathering and communicating.

We can reorient ourselves by remembering that the web is a subset of

the *inter*net, its very name representing the way this computer network shuttles information *between* and *among* people, rather than just a publishing medium that goes from point A (historians) to point B (an interested audience). This character of the web is already being exploited well by websites that have furthered professional communication. For instance, H-Net (http://www.h-net.msu.edu) has online discussion groups on over a hundred topics that one can access via the website or through email. Every day historians ask each other for assistance with their research and analyse new work in their fields. In contrast to paper media, the internet seems ideally suited for this kind of vibrant, daily exchange.

There are other forms of interactivity, or two-way communication, on the web that are less developed but that have the potential to create new kinds of history on the web. What about communications not just among historians but between historians and their subjects? For historians working on topics in the post-Second World War era, the web can be a tremendous resource for reaching those across the globe who have recollections or materials that could further their investigations in the present and contribute to archives for future historians.

Two web projects I have worked on, the ECHO Project (http://echo.gmu.edu) and the September 11 Digital Archive (http://911digitalarchive.org), have explored this possibility of using the web not only to present the past but also to collect it. Both sites use flexible databases and the upload capacity of the web (in addition to its vast capacity to download, or distribute, documents) to acquire materials from far-flung historical subjects such as scientists and engineers (in the case of ECHO) and those who witnessed and were affected by 9/11 (in the case of the September 11 Digital Archive). It is so far an imperfect science, to be sure, with pitfalls and insecurities, but the pay-off can be tremendous when a project that collects history online is successful. The nature and extent of what one can gather, while certainly different from a traditional oral history project or museum effort, seems just as enlightening and important as a future historical resource, and likely will grow more so as an increasing percentage of our communications and expressions occur in digital media.

That has certainly been the case with the September 11 Digital Archive, which the Library of Congress recently accessioned—the first major digital acquisition by the institution. Started in January 2002 with a rudimentary website by myself and colleagues at the Center for History and New Media at George Mason University and our associates at the American Social History Project/Center for Media and Learning at the Graduate Center of the City University of New York, and funded by the Alfred P. Sloan Foundation, the archive now contains over 130,000 digital objects from

the tragic events of that day and its aftermath, including stories, email, photos, artwork, poetry, audio recordings and digital video. The materials have been donated via the website and through less sophisticated means such as postal mail and the telephone (digitally connected to our servers) by over 30,000 individual contributors. The result is an extremely varied and vital archive of the experiences, thoughts and emotions of a broad spectrum of contributors from every state and from many nations. The archive contains the recollections and images of emergency workers and those who survived Ground Zero in New York and the Pentagon, the sounds of voicemail and the chatter of instant messages, as well as more impressionistic responses and subsequent reflections. While the September 11 Digital Archive was by far the largest and most comprehensive online effort to record this history, we were not alone in this endeavour; at the same time that we were permitting submissions through our website, dozens of other individuals and institutions were engaging in similar projects on the web, creating parallel collections that historians in decades to come will hopefully find useful when they try to understand the true meanings of 11 September 2001.

This concurrence also suggests a second, relatively unexplored possibility for history in the next decade of the web: the potential for 'interoperability'. A word most frequently used by 'web services' software developers who work on integrating websites and databases for decentralized businesses, and associated with opaque acronyms and new information architectures such as the international standard XML and Microsoft's NET initiative, interoperability holds promise for historical researchers interested in creating and accessing vast archives of documents that exist in different places on the web. For example, for my research on the history of nineteenth-century mathematics I have been following with great interest the virtual integration of several major archives of digitized mathematical texts by the Cornell University Library, the University of Michigan Library and the State and University Library Göttingen, called the Distributed Digital Library of Mathematical Monographs (http://www.library.cornell.edu/mathbooks/). This novel effort, facilitated by these new technologies that allow dispersed web servers to respond in tandem to requests for information, shows the power of interoperability. Certainly one could have gone to each library's website and conducted a separate search for books or phrases. However, the combined archive, created through seamless, hidden communications between these websites, is greater than the sum of its parts, and not only as a time saver. The Distributed Digital Library makes it far easier to get an overview of passages on a specific topic, such as non-Euclidean geometry, and to spot trends in the development of that subject during the

nineteenth century. It represents both a quantitative and a qualitative advance over prior resources in the history of mathematics.

My own effort to explore the potential of interoperability has been an attempt to virtually combine the often disregarded educational products of tens of thousands of academics. By using an algorithm to analyse documents found through behind-the-scenes communications with Google's enormous database, my Syllabus Finder (http://chnm.gmu.edu/tools/syllabi) searches the web for pages that look like syllabi and, when successful, saves them in a special format. It can then present those syllabi to interested seekers, figure out which school, college or university a course is being taught at, and, more experimentally, attempt to extract assigned books and other notable features of a syllabus. In a sense the Syllabus Finder creates a single, overarching course directory (though obviously it includes only syllabi that are posted to the web), with the potential to show how many courses are being taught on a specific topic, which books or articles are popular or influential, and which types of assignments faculty like to assign. My hope is that ultimately it will provide a unique window into the state of higher education. With over 250,000 syllabi collected so far (including over 10,000 history syllabi) it is not unreasonable to say that it is, by several orders of magnitude, the largest archive of course materials ever collected—and all of this was done in an automated fashion following my initial programming.

This collection of a quarter of a million syllabi, which could only have been acquired in a medium where text is machine-readable, searchable and easily reproduced, suggests another possibility for history in the second decade of the web: a digital cognate of 'close reading' that computer scientists call 'data-mining'. Data-mining involves complex analyses of digital materials to find meaningful patterns. In a way it is a more advanced version of what some classicists and literary scholars have been doing for years when they have manually counted the frequency of certain words, or compared the various uses of those words, in a text or set of texts.

On the web the speed with which one is able to do this sort of textual analysis can enable both quick assessments of historical collections as well as more substantive investigations. When Michael Kazin used search tools to scan the September 11 Digital Archive for the frequency of words such as 'patriotic' and 'freedom', he came to some important, if initial, conclusions about the American reaction to the terrorist attacks. Kazin discovered that fewer Americans than one might imagine saw 9/11 in terms of nationalism, radical Islam versus the values of the West, or any other abstract framework. Instead, most saw the events in far more personal and local terms: the loss of a friend, the effect on a town or community, the

impact on their family or jobs (Kazin 2003). Using similar electronic techniques, researchers are already probing the Syllabus Finder database to investigate a variety of themes, such as the changing role of race and ethnicity in undergraduate education, the growth of distance learning, how course materials on the web differ from offline materials, and how globalization is affecting the teaching of history and international relations. Future enquiries may use an extremely powerful computer science method called 'regular expressions', which one can use to search documents for all kinds of text patterns, not just keyword or phrase matches.

As with interactivity and interoperability, early efforts at data-mining hint at a future of doing and using history on the web that may have more to do with providing materials and structures for historical research than with the presentation of a finished secondary source such as an electronic book or exhibit website. Done correctly, and taking advantage of the medium's unparalleled ability to store, scan and interrelate documents, the web may enable new investigative tactics and resources. The 130,000 objects in the September 11 Digital Archive are not like any other 130,000 objects in the Library of Congress; we have acquired them, and can explore them, in ways we are only beginning to consider. Such collections and their associated electronic methods have the potential (an appropriately cautious word at this point) to supplement, though not replace, traditional historical research in physical archives.

Finally, thinking about doing history on the web, just like thinking about history in any other medium, should ideally force us to revisit age-old questions about what history is and how we should engage in it. For some years I taught Herodotus and Thucydides to freshmen and perennially enjoyed (and felt professionally reinvigorated by) their sense of the differences between these two progenitors of the idea of history. Purveyor of a mostly tidied, marching chronology that aspired to the truth about what seemed to him (and to most historians since) to be the most devastating events to befall ancient Greece, and with a particular focus on remarkable figures such as Pericles and Alcibiades, Thucydides' history of the Peloponnesian War captured the hearts and minds of most in my classes. Students readily identified his compelling narrative of warfare and diplomacy as 'history'. Other students found themselves drawn to Herodotus' idiosyncratic tangents and more expansive view of ancient culture beyond the battles between the Greeks and the Persians. This minority found Herodotus' sense of what belonged in the historical record refreshing: the strange rituals of non-Greek cultures, the sentiments of common people in addition to leading figures, competing and contradictory accounts. As Herodotus told his audience, he was saving and recounting it all because in

the future people might have different notions of what or who is important: 'I will go forward in my account, covering alike the small and great cities of mankind. For of those that were great in earlier times most have now become small, and those that were great in my time were small in the time before. Since, then, I know that man's good fortune never abides in the same place, I will make mention of both alike' (Herodotus 1987, p. 35). Not having to worry about fitting his account into a limited number of pages, a rich description of what he had learned from a wide variety of sources, Greek and otherwise, seemed to Herodotus to be the sensible course to take.

Although the medium is still in its infancy, it appears that the web may be more propitious for history in the inclusive and wide-ranging mode of Herodotus than in the resolute mode of Thucydides. On the web historians can supplement their narratives with virtually unbounded collections of sources, notes, graphs, charts, images and links that even a profligate publisher could not hope to fit into a book. In addition, given the open access of the web it seems appropriate to cast the widest possible net (so to speak) in projects like the September 11 Digital Archive, rather than focus on figures such as government leaders who will likely dominate coverage in print. The massive capacity of the web means that historians can push beyond the selectivity of paper collections to create more comprehensive archives with multiple viewpoints and multiple formats (including audio and video as well as text). These archives, hopefully partially making up for their lack of a curator's touch with their size, scope and immediacy, will in turn require more sophisticated tools for future research. If carefully developed, such collections—ideally interoperable with others of their ilk—may provide historians not only of this generation, but also of generations to come, with the means to understand the past better and more deeply.

Note

[1] Routledge cannot be held responsible for the content or accuracy of the urls linked to from the online version of this article, which can be found at http://www.tandf.co.uk/journals/titles/13642529.asp

References

Grafton, A. (1997) *The Footnote: A Curious History*, Harvard University Press, Cambridge, MA.

Herodotus (1987) *The History*, trans. David Grene, University of Chicago Press, Chicago, IL.

Kazin, M. (2003) '12/12 and 9/11: tales of power and tales of experience in contemporary history', *History News Network*, 11 September, http://hnn.us/articles/1675.html.

Lynch, P. J. & Horton, S. (2002) *Web Style Guide* (2nd edn), Yale University Press, New Haven, CT.

McMichael, A., O'Malley, M. & Rosenzweig, R. (1996) 'Historians and the web: a guide', *AHA Perspectives*, January, pp. 11–16.

Negroponte, N. (1996) *Being Digital*, Vintage, New York.

O'Malley, M. (2000) 'Building effective course sites: some thoughts on design for academic work', *Inventio*, Spring, http://www.doit.gmu.edu/inventio/spring00/omalley/momalley_1.html.