

Digital History: The Raw and the Cooked¹

Daniel J. Cohen

Essays on the future of history by Josh Brown, Ed Gallagher, Alan B. Howard and Carl Smith (with colleagues Brian Dennis and Jonathan Smith), inspired by experiences online and off, helpfully remind us that there is much more to being a historian than the reading and writing of articles and books. The technology of the web allows us to highlight other ways that we can present, interpret and discuss the past, and enable others to do the same. We can focus on images rather than texts to literally create a vision of a historical moment (Brown); we can construct new tools and methods for engaging historical artifacts (Smith *et al.*); we can serve as archivists and exhibitors (Howard); we can enhance our teacher–student and student–student interactions (Gallagher). Although the continued pre-eminence of the written word—especially as it relates to tenure—militates against it, the web may elevate these other forms of historical practice, or at the very least make them more prominent and effective.

While these essays present a generally coherent, and I believe favourable, overview of the current state of online history and its future possibilities, they also reveal an important and currently unresolved tension in digital work that may not be apparent to historians new to the subject matter: the extent to which online exhibitions, archives and learning aids gain advantage from organization or structure. With a nod to Claude Lévi-Strauss, let us call this the difference between ‘raw’ and ‘cooked’ digital history. Raw digital history comprises documents, information and communications that are heterogeneous and that have little, if any, organization. Cooked digital history takes such historical materials and adds helpful markings and a measure of homogeneity. An offline cognate for this difference between raw and cooked digital history would be the distinction between various archives’ unruly boxes of notes, manuscripts and letters by Benjamin Franklin and the carefully vetted, transcribed, formatted,

annotated and indexed volumes of Yale University's *Papers of Benjamin Franklin*.

It would appear self-evident that the cooked is greatly preferable to the raw, that there is much to be gained by structure. No one likes a disorganized class discussion, or a book without a table of contents, or chapters without identifiable and coherent topics. Many of these same scruples naturally remain when we move online. Surely we prefer websites with decent navigation, clearly defined sections and graphical consistency. Furthermore, some of the more advanced possibilities for online history that I outline in 'History and the Second Decade of the Web', such as the interoperability of the Distributed Digital Library of Mathematical Monographs, can only occur within a fairly rigid organizational system. Sites that use technologies for structuring historical materials, such as the Extensible Markup Language (XML) for text or Geographic Information Systems (GIS) for locations, enable useful searches as well as novel methods for displaying such information. There is no doubt that it would be wonderful if everyone who posted a historical document online included XML 'tagging' for the fifteen basic elements of the Dublin Core Metadata Initiative, including the name of the creator, the title of the document, and the date of its creation (<http://dublincore.org/documents/dces/>). Of even greater potential is Bruce Robertson *et al.*'s Historical Event Markup and Linking Project (<http://www.heml.org/>), which, if adopted by a large number of historians putting materials online, would produce a more easily digestible and functional history web—one that could, for instance, create time-lines instantly from scattered resources. It is easier to find, relate, analyse and cite historical resources when they are structured in this fashion.

Unfortunately, cooked digital materials come with significant costs to both producers and consumers. Those who have created highly organized web archives or exhibits, and especially those who have experimented with markup languages such as XML, know how time-consuming and expensive they can be. Even with the assistance of software tools and partial automation there can be significant labour costs involved in the construction of a database, graphical environment or well-formed set of documents. There are other pragmatic questions to be asked: Is it possible to get a critical mass of historians—who, like most academics, are argumentative and contradictory and who generally lack the requisite funds to hire technical help—to adopt any common scheme for structuring online materials? And are new digital views of history, such as instantly created time-lines, worth the cost of such organization? Historians are not used to doing cost-benefit analysis, but this may be necessary in the future of online history. Moreover, it is

important to remember that in the first five years of the web Yahoo! was one of the dominant websites because it provided a cooked view of the medium, adding the librarian's touch of classification and order to a confusing hotchpotch of sites. Now Google's rapid search of the raw mass of disorganized, heterogeneous web pages has replaced, by a tremendous margin, Yahoo's tidy directory as the leading referrer of web visitors. The brute force of computer algorithms has proven far more useful than any human cataloguing.

Looking at revealing server logs and usage patterns for the web projects I have worked on, it is clear that consumers of digital history have reasons for accessing and using online materials that producers of digital history cannot fully anticipate or understand. For example, only 30 per cent of the web traffic to The September 11 Digital Archive in 2003 came first to our home page, which provides an overview, guide and news about the collection (what we, the producers, consider most important); 70 per cent went directly to specific materials in the archive, such as a photo, story or video, and they often did so without using our search tools (mostly using Google instead) or by following our site's organizational cues (which were based on our internal typology of the contributions: still images, moving images, email and so on). In addition, while most visitors who came to the site in 2003 via a search engine had typed in 'September 11' or '9/11', 228 visitors were interested instead in 'teen slang', 421 were searching for information on the 'USS Comfort' (one of the Navy's hospital ships), and 157 were simply looking for a 'map of lower Manhattan'. There were thousands of visitors who came to our archive for reasons that had absolutely nothing to do with 11 September 2001.

The value a user gains from the cooking of a digital history project may be questioned in other ways. As Josh Brown notes candidly and shrewdly in his account of the construction of a remarkable virtual version of P. T. Barnum's Lost Museum, adding even a modicum of structure has the potential to enshrine certain choices, paths, interpretations and associations. Scholars who structure historical documents with markup languages such as XML make choices—often quite good choices, but choices none the less—about which elements of a document are most important. But future readers of those documents may have other interests or concerns, or may have other ways of scanning them. There are also thorny questions related to 'granularity' or detail. When you are digitizing a 1909 map of Chicago, how many coordinates do you highlight, and which ones? A hundred, ten, a thousand? Similarly, to what extent do you structure or restrict online discussions by students, or guide their ventures into historical websites? How much do you allow others to attach comments, critiques or other

materials to your site? In general, how much organization should we provide as historians online, and how much control should we cede to our students, our colleagues, and the current and future audiences for our digital histories?

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>