

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

---

Documentary Editing: Journal of the Association  
for Documentary Editing (1979-2011)

Documentary Editing, Association for


---

2008

The Pitfalls of Digital History: Review of *Digital History: A Guide to Gathering, Preserving, and Presenting the Past on the Web*. Daniel Cohen and Roy Rosenzweig.

David Spiech

Follow this and additional works at: <http://digitalcommons.unl.edu/docedit>

 Part of the [Digital Humanities Commons](#), [Other Arts and Humanities Commons](#), [Reading and Language Commons](#), and the [Technical and Professional Writing Commons](#)

---

Spiech, David, "The Pitfalls of Digital History: Review of *Digital History: A Guide to Gathering, Preserving, and Presenting the Past on the Web*. Daniel Cohen and Roy Rosenzweig." (2008). *Documentary Editing: Journal of the Association for Documentary Editing (1979-2011)*. Paper 59.

<http://digitalcommons.unl.edu/docedit/59>

This Article is brought to you for free and open access by the Documentary Editing, Association for at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in *Documentary Editing: Journal of the Association for Documentary Editing (1979-2011)* by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

# The Pitfalls of *Digital History*

David Spiech

*Digital History: A Guide to Gathering, Preserving, and Presenting the Past on the Web.* Daniel Cohen and Roy Rosenzweig. Philadelphia: University of Pennsylvania Press. 2006 (paper). 316 pages. ISBN 978-0-8122-1923-4. \$28.95. Online at <http://chnm.gmu.edu/digitalhistory/>.

With *Digital History*, Dan Cohen and Roy Rosenzweig have attempted to boost scholarly authors and editors into the wild world of web publishing, even those who have tried to maintain a conservative academic distrust of electronic media. They write engagingly and frankly, addressing the reader as a colleague who knows historical material well but needs comprehensive background about every facet of digital access and presentation.

Not surprisingly, as historians, they provide details from the short history of computerized research, data analysis, and Internet publishing. What makes this volume valuable is that they go beyond the historical background and give practical, somewhat current advice about managing digital publishing projects and preserving digital materials. Obviously drawing on their own experience with the Center for History and New Media (CHNM), the authors are not shy about treading in controversial areas such as copyright law and digital archiving. The most helpful parts of this work are found in the practical details that can be known only from actually trying to use or present online historical materials. They carefully counter the traditional historian's fears that contextual information will be lost, as well as overstated claims for the permanence and broad distribution of digital materials.

For the academic historian who may be familiar only with the front end of web and database applications, they provide an appendix with an introduction to database programming and XML markup. They

skim quickly over SGML and TEI (Chapter 3: Becoming Digital), while making clear that the most useful future work will require TEI compliance and the use of some form of contextual markup, such as XML.

By bringing the reader up to speed on various cultural and technological developments pertaining to the World Wide web and digital media yet stopping right at the brink of complex XML implementations, the authors succeed in presenting web publishing as a worthwhile venture for the independent or underfunded historian.

One of the unique aspects of this volume is that the authors have, in a sense, followed their own advice by publishing the volume itself online in a straightforward HTML presentation. However, it has suffered some of the pitfalls that the authors themselves warn about regarding format changes, conversion to web presentation, and the stability of Internet links.

### **Online Edition Compared to Print Edition**

The University of Pennsylvania Press agreed to allow Cohen and Rosenzweig to post their book online, with free access. However, the online edition contains errors that are not in the print edition, possibly from rekeying and reformatting, or perhaps the online text is from a preprint version of the text.

The online edition features a home page that functions like a preface, as well as an “About the Authors” page that has not been updated since the death of Roy Rosenzweig. The online edition also includes a “Buy the Book” page with links to the University of Pennsylvania Press, Amazon, and Barnes & Noble, which unhelpfully shows that the press’s page is out of date: it gives the publication date as 2005 rather than 2006, the number of pages as 325 rather than 316, and the number of illustrations as 45 rather than 43.

The online edition uses no chapter numbers, and some chapter and section headings are different. Every section online begins with a nice faux-Renaissance-style illuminated drop-cap that is not in the print edition. Notes in the online edition are presented as footnotes that cross-reference to the permanent “Links” page, whereas notes in the print edition are endnotes. Along with the placement of the table of contents as a hyperlinked sidebar, these differences represent improvements made to accommodate the web presentation.

Some clues suggest most of the online edition was implemented after the print edition was completed, or else done separately with less editorial skill. All of the illustrations online are presumably the “original” images, since they sometimes contain more information than in the print edition. However, they vary in sometimes showing less than the images in the print edition, sometimes showing completely different images, and sometimes showing a different screen-capture frame.

For example, some screenshots of webpages contain different embedded images (Figures 5, 7, 9), indicating that they were captured at a different time than the print images. Usually this is irrelevant to the point made in the text or the caption, but sometimes the discrepancy causes an egregious error in presentation.

For Figure 11, the screenshot is from the wrong day, so that the online image doesn’t match the caption. To compound the problem, the text links to a note that links to a cached PDF version of the site that is different from the image accompanying the online text and the print image! Therefore, the correct image found in the print edition is completely lost to the online reader.

Figure 25 presents two images for comparison, yet the online text presents the wrong first image, so that comparison is more difficult for the online reader. The purpose of the example is to show how a specific text can be reformatted using design principles. Yet, because the authors have substituted a screenshot image with a different text for the initial appearance, the point is lost. This might seem trivial unless you had seen the print edition, which makes the point clearly. That discrepancy reveals that the online version was done carelessly.

For Figure 42, the text discusses the image of a user agreement, but the online image shows a different version from the print image. Notably, all of the captions online must have been rekeyed, because they contain many errors and edits compared to the print captions. Capitalization errors, typos, and transpositions are typical, but some errors are more problematic.

For example, Figure 1 includes a copyright statement in the print edition, but this statement is missing in the online edition. Half of the caption for Figure 19 is missing in the online edition, cut off in midsentence.

For Figure 23, the link provided in the caption is presented as a live link, but it is not; furthermore, following the note link provided in the text

(for note 12) leads to a useless assortment of links on the Links page: the Live Site link is actually to the *Digital History Links* page itself, the Cached link yields a 403 Forbidden error from the CHNM web server, and the PDF link yields a Page Not Found error from the CHNM web server. The URL is provided nowhere in the online or print edition, so it is completely lost to the reader.

Several problems with links can be found with a brief examination. For example, in the Introduction, note 11, the text references 2004 statistics; the note cites a Technorati page from 17 February 2005; but the cached page is from June 2005, giving different stats.

In the text, the online edition includes one hypertext link that has no corresponding note in the print edition: [http://www.st-andrews.ac.uk/%7Ewww\\_sd/jrd4.html](http://www.st-andrews.ac.uk/%7Ewww_sd/jrd4.html) (<http://chnm.gmu.edu/digitalhistory/exploring/4.php>). This appears to be the only live link added to the online edition other than internal cross-references; because it is a hyperlink, it is not in the print edition at all. Curiously, whereas the online edition has several apparent problems with its clever system of crosslinks and external links, in a few places it includes URLs in the text that are not live links and not included on the designated Links page.

In the chapter, “Getting Started: Naming Your Site and Presenting It to the World,” the page contains a URL in the text that is neither hotlinked online nor footnoted—and not included in the list of updated/cached links: a Chinese Culture page by Paul Halsall at <http://academic.brooklyn.cuny.edu/core9/~phalsall/>. There is a new Halsall Chinese culture page at <http://acc6.its.brooklyn.cuny.edu/~phalsall/index.html>, but Cohen and Rosenzweig fail to link to Halsall’s permanent site, as suggested by Halsall in the image they show. Cohen and Rosenzweig also give as a “current” URL for Halsall [www.unf.edu/~phalsall/](http://www.unf.edu/~phalsall/), which is now obsolete, since Halsall had left UNF by the time the print edition went to press.

### **Inadequacy of Both Editions**

The introduction to the Notes section provides one rationale for having the text online, in that it enables the authors to keep an updated list of web addresses there. To facilitate this, they use a system of numbering the links by chapter and then referring to this number within the printed note. The authors state that if they find that a link has disappeared, they will provide the best available reference for the material. However, I found that several of the online links were no longer valid. In response to this

problem, the authors provide cached versions of the original Websites in HTML or PDF format, or both. However, the embedded images in the cached HTML webpages are hotlinked to the original sites, so these images tend to still change.

This method of providing web references only online requires the authors to continuously check and update their reference list, so that theoretically they could update it after the main text is obsolete. On the other hand, at some point they will presumably either stop updating their web references and simply allow them to reflect the most current links available at a particular point in time, or they will revise the main text itself, making the old web reference list unusable.

In either case, the printed book would then become thoroughly obsolete, since there would no longer be an updated web reference list corresponding to the printed text. Moreover, since the printed notes provide none of the URLs themselves, but merely cross-references, without the online key list none of the web references can be reconstructed. Since many of the web references are web-only and give only an author name or document title, the lack of online resources would make these notes superfluous and invalid.

Of course, one could argue that by the time the authors revise the online text or give up on updating the online reference list, the content of the printed book would be irrelevant anyway. However, given the current rate of technological change, it calls into question the whole enterprise of printing a physical book, especially one that is not self-contained with regard to references.

### **Conclusion**

This volume represents an adequate summary of advice for the novice web publisher, or for any author or editor who has felt too intimidated to attempt web publishing. It is particularly good for acquainting a traditional historian with the most recent developments in web culture and technology and in warning about some of the possible problems. The online edition will probably be the most useful for this purpose, assuming it is readily found by a novice using a search engine. The print edition is like an introduction to the online edition or a handy desktop reminder of the basic points, rather than being a standalone volume.

For that reason, a library would be better served by not purchasing the print edition at all but rather providing a link to the Website or a cached version of it; and the University of Pennsylvania Press would be better served by charging significantly less for the print edition. If the online edition were edited properly and regularly updated, it might serve for several years as a good introduction to web publishing.

However, if applications developers succeed in making web publishing more user-friendly for academic authors and editors, the technology could change quickly, making the entire work obsolete except as a historical summary. Interestingly, that describes the history of older digital publishing technologies: as processes are simplified and the technical workings are hidden behind more intuitive user interfaces, the technical knowledge previously considered necessary becomes irrelevant. In their analysis of the history of digital publishing, Cohen and Rosenzweig may have foreshadowed the fate of their own work on digital publishing.