# Archival Theory and Digital Historiography: Selection, Search, and Metadata as Archival Processes for Assessing Historical Contextualization

Joshua Sternfeld

**Abstract**

Digital history, a field within the digital humanities, has challenged the disciplinary boundary that in recent decades has come to separate the work of historians and archivists. A new theory and methodology that draws from both disciplines can create a shared vocabulary for the production, use, and evaluation of *digital historical representations*, a broad term that encompasses an array of products such as archives, databases, geospatial visualizations, and mobile applications. This article argues that archival theory, when combined with historiography and technical or computational standards, contributes to a new theory called *digital historiography*. Digital historiography is defined as the interdisciplinary study of the interaction of digital technology with historical practice. Three archival processes—selection, search, and the application of metadata—form the theory's foundation for determining a digital historical representation's contextualization, which may aid in assessing its trustworthiness and effectiveness to communicate historical knowledge.

Digital history spans diverse academic and public activities. Two brief examples can provide a glimpse of the depth and range of digital historical pursuits. The Louisiana Office of Tourism's interactive website,



This article is an expansion of a paper delivered at the 2010 Digital Humanities Conference in London. I would like to thank all those who provided constructive feedback and discussion through its many incarnations, including Joel Wurl, Elizabeth Joffrion, Helen Agüera, Mary Downs, Jennifer Serventi, Brett Bobley, Anthony Grafton, Jessica Sternfeld, and Anne Kasab.

The thoughts and ideas expressed in this paper are entirely my own and do not reflect those of the National Endowment for the Humanities or any other federal agency.


**FIGURES 1A AND B. Screenshots from A Story Like No Other website, http://www.astorylikenoother.com/explore/ and iPhone app, http://itunes.apple.com/us/app/a-story-like-no-other/ id378072152?mt=8, accessed 20 January 2011.**

A Story Like No Other, guides users through a collection of more than 30 museums, heritage sites, institutions, and cultural attractions in all corners of the state.[1] Multimedia content, including narration by actor Louis Gossett, Jr.; geographical position system (GPS) data for each site; editorial commentary supplied through a weblog; and the option to supply user-generated editorial information through social media like Facebook and Twitter together comprise an expanding and evolving database of cultural heritage locations related to Louisiana's African American history. Visitors have the option to "follow in [the] footsteps" of African Americans through one of the site's curated itineraries; however, the site excels at encouraging users to "explore" the sites in any order that they choose. Personal itineraries may traverse the many regions of the state with such designations by the Office of Tourism as "Plantation Country," "Cajun

[1] Louisiana Office of Tourism, A Story Like No Other: Louisiana's African American Heritage Trail, http://www.astorylikenoother.com/, accessed 20 January 2011. The website encourages visitors to download an accompanying iPhone app with virtually the same content as the website as well as functions unique to the mobile platform. Through the app, users customize a personal itinerary that "follow[s] [the] footsteps" of African American ancestors "from street corners and marketplaces to churches, cafés, [and] universities" (see Figure 1b). As a jarring reminder of the fleeting nature of many digital historical representations, the app, which was created in 2010, was no longer accessible in the iTunes App Store as of 5 July 2011. A short, two-minute video is still available on the website and provides a sense of the app's various customizing functions. "iPhone," http://www.astorylikenoother.com/iphone/, accessed 5 July 2011.

Country," or—in a nod to the state's rich blues and jazz heritage—"Crossroads" (see Figure 1a).[2]

On a more academic front, a team of scientists and engineers from Harvard, MIT, and Google released a preliminary study in the magazine *Science* that conducted quantitative textual analysis using a corpus of over five million books digitized by Google, or 4 percent of all works ever published. The study claims to have revealed intriguing historical evidence through mapping word frequency over time. One can infer, the researchers argue, when political states censored artists or when celebrity figures fell into and out of public attention. Based on their findings, they suggest creating a new field of study called *Culturomics*, which they define as "the application of high-throughput data collection and analysis to the study of human culture."[3]

As a concept, digital history has never been stable. Some have argued that it should stand for a methodology, or a set of skills, while others have argued that it should stand as a full-fledged discipline. William G. Thomas III defines digital history as

> an approach to examining and representing the past that works with the new communication technologies of the computer, the Internet network, and software systems. On one level, digital history is an open arena of scholarly production and communication, encompassing the development of new course materials and scholarly data collections. On another, it is a methodological approach framed by the hypertextual power of these technologies to make, define, query, and annotate associations in the human record of the past.[4]

Thomas captures here the tension articulated by advocates and skeptics of digital history. On the one hand, digital history promises an "open arena" in which new material is circulated through more democratic communication channels.

[2] "Explore," A Story Like No Other, http://www.astorylikenoother.com/explore/, accessed 20 January 2011.

[3] Jean-Baptiste Michel et al., "Quantitative Analysis of Culture Using Millions of Digitized Books," *Science* 331, no. 6014 (14 January 2011): 176–82.

[4] Daniel J. Cohen et al., "Interchange: The Promise of Digital History," *Journal of American History* 95, no. 2 (2008), Organization of American Historians, http://www.journalofamericanhistory.org/issues/952/interchange/index.html, accessed 30 August 2010. For an introduction to digital history's definition and development, see Daniel J. Cohen and Roy Rosenzweig, *Digital History: A Guide to Gathering, Preserving, and Presenting the Past on the Web* (Philadelphia: University of Pennsylvania Press, 2006); William G. Thomas III, "Computing and the Historical Imagination," in *A Companion to Digital Humanities*, ed. Susan Schreibman, Ray Siemens, and John Unsworth (Malden, Mass.: Blackwell Publishing, 2004); Jeffrey G. Barlow, "Historical Research and Electronic Evidence: Problems and Promises," in *Writing, Teaching, and Researching History in the Electronic Age*, ed. Dennis A. Trinkle (Armonk, N.Y.: M. E. Sharpe, 1998); David J. Staley, "From Writing to Associative Assemblages: 'History' in an Electronic Culture," in *Writing, Teaching, and Researching History in the Electronic Age*; Orville Vernon Burton, "American Digital History," *Social Science Computer Review* 23, no. 2 (2005); Roy Rosenzweig, "Scarcity or Abundance? Preserving the Past in a Digital Era," *American Historical Review* 108, no. 3 (2003).

On the other hand, Thomas and others recognize that the underlying spirit of openness disrupts traditional practices, thereby raising concern that we should establish a methodology for working with historical materials.

The two examples above, if we accept that they combine digital technology with historical information and given their differences in execution and scholarly intention, reflect how difficult it may seem to sustain an "open arena" while fostering new practices. In spite of their obvious differences, we ought to agree that they both employ a systematic arrangement of historical information, which allows us to extract useful commonalities. For example, both implement criteria to select historical information. The Louisiana website limits its content data to a set of curated cultural heritage sites relevant to African American history. The scientific research team narrows the available fifteen million digitized works in Google Books to a subset of works published between the years 1800 and 2000 that possess "quality" metadata.[5] Additionally, both the site and the scientific study allow users to browse or search the content. With a virtual magnifying lens, users may click on Louisiana sites located on an interactive map to construct their own tour. The research team offers a tool called the "Books Ngram Viewer" that allows users to "run [their] own experiment" by inputting any word or combination of words.[6] Finally, both projects rely upon metadata as the backbone for the content's organization and discoverability. The site, for example, relies upon accurate GPS data to map its cultural sites while the research team depends upon Google Books' chronological metadata to generate time graphs. From this snapshot, we may infer that digital technology—either by inspiring users to take a virtual (and eventually actual) tour of Louisiana's racial history, or by mining five million books for hidden linguistic patterns—has enabled capabilities not possible even just a few years ago, thereby challenging academic and cultural sectors to rethink how to represent and analyze historical information.

Digital archives, collections, databases, websites, pedagogical tools, mobile applications, and geospatial visualizations may possess different, even opposing, formal properties, yet they all share the same objective, that is, they all use digital technology to represent history. It is imperative that we have the ability to discuss digital historical output in a way that allows us to talk across fields and formats. For the remainder of this article, I will refer to the products of digital history collectively as *digital historical representations.* This article will argue that archival theory, when combined with historiography and technical or computational standards, contributes to digital historiography, a new interdisciplinary theory dedicated to the construction, use, and evaluation of digital historical representations.

[5] For a full discussion of how the researchers selected their dataset, see Michel et al., "Quantitative Analysis of Culture Using Millions of Digitized Books," 176–78.

[6] The Books Ngram Viewer can be found at "Books Ngram Viewer," Google Labs, http://ngrams.googlelabs.com/, accessed 20 January 2011.

As the above two examples show, representations with divergent formal properties still share fundamental processes that may serve as points of comparison. Through a series of illustrative examples that analyze elements of digital historical representations, this article will demonstrate that the areas of selection, search, and metadata comprise the framework for digital historiography. All digital historical representations implement a selection process for historical content, driven by selection criteria that guide the content's acquisition, arrangement, and description. Similarly all representations invite users to "search" the content, which may vary from a conventional search query to an unconventional browsing or exploration of an interactive interface. Finally, all representations apply metadata to historical content, which determines the content's trustworthiness, especially with respect to provenance.

These archival processes unlock the many layers of a digital historical representation's *contextualization*. Each relationship or potential relationship between units of historical information—brought together by a selection process, a search inquiry, and archival provenance—reflects an act of historical interpretation by the representation's creator that signifies a convergence of historiographical and archival decisions. In the end, contextualization contributes to a representation's trustworthiness and consequently its effectiveness. The trust bond between archivists and archival users over time has been well established; it is time that we apply a similar set of guidelines to the construction, use, and evaluation of digital historical representations.

## A New Theory: Digital Historiography

Digital historical representations generate new interdisciplinary collaborations among scholars and practitioners. Academic and intellectual boundaries separating humanists, information specialists, and computer scientists fade away as members of the various fields rush to make use of the new technologies. A small, dedicated cohort of scholars, working within the emerging field known as the digital humanities, along with practitioners outside conventional academic settings, challenge the limits of intellectual and methodological possibilities. Archivists, librarians, curators, engineers, and scientists, all of whom exhibit an enthusiastic drive to experiment, participate in this rapidly evolving community.

Like all experimental endeavors, the work to construct and use digital historical representations should follow rigorous methods grounded in theory. Unfortunately, up to this point, the onslaught of new research tools, spectacular visualizations, massive datasets, and hastily assembled cultural heritage collections has overtaken our capacity to assign scholarly and pedagogical value. In a major 2010 survey conducted by the American Historical Association (AHA),

Robert Townsend concludes the following: "The scale of interest and array of forms and outlets for publication point to a growing issue in the discipline—that the forms and opportunities for publication seem to be outpacing the tools for assessing them in the monograph-focused reward systems of higher education."[7]

Developers of digital historical representations incorporate complex systems and computational processes that impose barriers for conducting comprehensive analysis at all levels, including a representation's construction, use, and evaluation. Participants in all three areas face the same simple, yet confounding question, Where to begin? Creators may ask: Which content management system, metadata schema, or text encoding schema will optimize my objective? Users of representations may ask: How do I find the content that I am looking for, and what tools do I have at my disposal to assemble information? Faced with having to consider both a representation's structural design and its historiographic strengths or shortcomings, evaluators of digital historical representations may find themselves in a similar quandary as they weigh multiple elements. Until now, each of these activities involving new technologies—creation, use, and evaluation—has operated in a Wild West, with little guidance on how to arrive at critical decisions or awareness of the consequences of those decisions.

Digital history thus demands a theory capable of handling the inherent interdisciplinarity of its products. Regrettably, a noticeable rift fueled by skepticism and distrust has developed between well-established institutions—including academic departments, publishers, funding agencies, and trustee boards—and the digital humanities community. Among the skeptics are historians who question the integrity of inquiries based on digital or quantified evidence and archival specialists who point to the misappropriation of terms such as *finding aid, collection, record,* and *archives.* Advocates of the digital humanities counter that the tools at our disposal, and those that have yet to be invented, promise enhanced capacity to ask new humanistic questions that would otherwise not be possible. Both sides have valid points, and yet, at the moment, they are speaking at odds with one another, unable to find common conceptual or terminological footing.

[7] Robert Townsend, "How Is New Media Reshaping the Work of Historians?," *Perspectives on History* (November 2010), American Historical Association, http://www.historians.org/perspectives/issues/2010/1011/1011pro2.cfm. Townsend's conclusions echo Christine Borgman's strong invocation to overhaul humanities infrastructure. Borgman argues that, as a matter of survival, digital humanities must adopt academic practices commonly associated with the sciences. Christine L. Borgman, *Scholarship in the Digital Age: Information, Infrastructure, and the Internet* (Cambridge, Mass.: MIT Press, 2007) and Christine Borgman, "The Digital Future Is Now: A Call to Action for the Humanities," *Digital Humanities Quarterly* 3, no. 4 (2009), http://www.digitalhumanities.org/dhq/vol/3/4/000077/000077.html. The absence of systematic review for digital humanities work and desire among academic departments to develop one is also summarized in a recent article for the *The Chronicle of Higher Education.* Jennifer Howard, "Hot Type: No Reviews of Digital Scholarship = No Respect," *Chronicle of Higher Education* (23 May 2010), http://chronicle.com/article/Hot-Type-No-Reviews-of/65644/. All accessed 20 January 2011.

A new methodology and theory called *digital historiography* can resolve these differences. But first, an explanation of what digital historiography is not. Digital historiography is not the digitization of analog history or the placement of historiographic essays online in e-journals. It also accepts as a given technological changes to commonplace activities such as word processing or electronic correspondence. The American Historical Association survey mentioned above indicates that professional historians have slowly, yet steadily, adopted digital tools for their daily activities.[8] While these uses have certainly revolutionized historical practice by improving productivity and efficiency, in the broader picture, they only occupy a starting point for digital technology's potential to redefine the discipline.

Digital historiography is something much more expansive than using technology to enhance proficiency. I define it as *the interdisciplinary study of the interaction of digital technology with historical practice.* This definition allows for potentially broad application, as we find technology affecting history at every phase including research, preservation, pedagogy, and presentation. In all of these areas, historiography has traditionally provided the epistemological basis by which we can assess historical knowledge. Historians work toward developing new theories about historical evidence, including its reliability, its authentication, and its use to construct arguments and narratives. They recognize the dynamic forces at play that affect our perception of the past, understanding that the lenses with which we filter evidence may reflect as much about contemporary society as they do about the historical actors under consideration. The "historiography" in digital historiography thus promises to preserve this principle by preserving the "logic of inquiry" that has always accompanied the modern discipline of history.[9]

Similarly, "digital" serves as a convenient, yet admittedly imprecise catch-all term that stands for a wide spectrum of media formats, tools, computational processes, and visualization platforms. More significantly, it encompasses the belief that a common methodological language should be applied across formats. Archival theory provides the appropriate basis for such a language by

[8] Townsend, "How Is New Media Reshaping the Work of Historians?" The AHA survey was conducted in spring 2010. Part 2 of the survey examines the adoption of digital and new media technologies by professional historians for their research and teaching. More than two-thirds of respondents (68.9%) consider themselves "active users." As the report goes on to show, however, use could entail any combination of online search engines, word processing, library databases, online archives, or scanners. Most telling for our purposes is the widespread hesitancy to publish work in a new media format. Only 20 percent of those who have considered publishing online and just over 10 percent who have published online considered using new media to "tell [a] narrative in [a] new or different way."

[9] A comprehensive discussion of historiography is beyond the scope of this article. For a concise overview, see Georg G. Iggers, *Historiography in the Twentieth Century: From Scientific Objectivity to the Postmodern Challenge* (Middletown, Conn: Wesleyan University Press, 2005). In speaking of postmodern historiography of the last two or three decades, Iggers writes: "…[W]hile historians became much more guarded in their belief in the authority of science, they nevertheless worked with the conviction that the historian dealt with a real and not an imagined past and that this real past, although accessible only through the medium of the historian's mind, nevertheless called for methods and approaches that followed a logic of inquiry," 15.

defining and theorizing the processes—including selection, search, and metadata—that transform raw data or information into *evidence*. This transformation begins with distinguishing "records" from "documents" and "archives" from "collections." Citing definitions established by InterPARES 1, Luciana Duranti and Kenneth Thibodeau claim that what distinguishes records from documents is "the nature of [a record's] relationship with the activity of the creator rather than its formal or content characteristics."[10] In other words, a record retains its "recordness" by preserving the context surrounding its creation and transactional history; it embodies more than the information contained within its documentary form. Archival theory has shown that preserving provenancial properties of records, and the relational properties of archives, can produce representations that are more contextually rich than loosely assembled collections.[11]

The distinction between records and documents extends to the broader topic of evidence. Jennifer Meehan asserts that an "archival concept of evidence" leads to "analysis of *relations* between record and event" (emphasis my own). She writes: "An archival concept of evidence helps us to understand that, at a very basic level, the archival treatment of records effectively constitutes records as matters of evidence, that, in identifying, preserving, and communicating the relationships between records and events, archivists select, shape, and situate records such that they can be regarded and used as documentary sources that are capable of serving as evidence of past events."[12] Meehan's definition of evidence, with its emphasis on the relationships among records, carries important implications for defining the "digital" half of digital historiography.

[10] Luciana Duranti and Kenneth Thibodeau, "The Concept of Record in Interactive, Experiential and Dynamic Environments: The View of InterPARES," *Archival Science* 6, no. 1 (2006): 15.

[11] For further discussion on the differences between records and documents, see, for example, David Beard, "From Work to Text to Document," *Archival Science* 8, no. 3 (2009); Duranti and Thibodeau, "The Concept of Record in Interactive, Experiential and Dynamic Environments"; Stephen G. Nichols, "An Artifact by Any Other Name: Digital Surrogates of Medieval Manuscripts," in *Archives, Documentation, and Institutions of Social Memory: Essays from the Sawyer Seminar*, ed. Francis X. Blouin, Jr. and William G. Rosenberg (Ann Arbor: University of Michigan Press, 2007); Geoffrey Yeo, "Concepts of Record (1): Evidence, Information, and Persistent Representations," *American Archivist* 70 (Fall 2007). For a discussion of the differences (semantic and otherwise) between archives, collections, and databases, see Kenneth M. Price, "Edition, Project, Database, Archive, Thematic Research Collection: What's in a Name?," *DH: Digital Humanities Quarterly* 3, no. 3 (2009), http://www.digitalhumanities.org/dhq/vol/3/3/000053.html#, accessed 20 January 2011; and James Currall, Michael Moss, and Susan Stuart, "What Is a Collection?," *Archivaria* 58 (2004).

[12] Jennifer Meehan, "Towards an Archival Concept of Evidence," *Archivaria* 61 (2006): 143. For further discussion of the concept of information as evidence, see Terry Cook, "Archival Science and Postmodernism: New Formulations for Old Concepts," *Archival Science* 1(2001); Terry Cook, "What Is Past Is Prologue: A History of Archival Ideas since 1898, and the Future Paradigm Shift," *Archivaria* 43 (1997); Margaret Hedstrom, "Archives, Memory, and Interfaces with the Past," *Archival Science* 2, nos. 1–2 (2002); Joanna Sassoon, "Beyond Chip Monks and Paper Tigers: Towards a New Culture of Archival Format Specialists," *Archival Science* 7, no. 2 (2007); Jennifer Meehan, "Making the Leap from Parts to Whole: Evidence and Inference in Archival Arrangement and Description," *American Archivist* 72 (Spring 2009).

Digital historical representations, as we shall see, invite users to develop relationships among content on an interactive basis. Every search query, computation, or graphical visualization pulls historical data on the fly. An archival understanding of information as evidence reminds us that the process of *how* a representation assembles information carries just as much significance as the information itself. Duranti and Thibodeau contend that contextual knowledge may be embedded within the experience of interacting with digital records: "Given the essential memorial function of a record, the digital components might themselves constitute a record or a set of records, depending on how they are instantiated in the system."[13] We must consider digital historical representations, therefore, as active, dynamic creations rather than static objects. They contain interactive processes both on the development end as well as the user end, including the activities of search, exploration, recombination, and repurposing. Creators and users share responsibility for ensuring that a representation does not distort context through these interactive processes. The analysis that occurs in digital historiography, therefore, must always take into consideration simultaneously a representation's discernible components, as well as its *potential* to generate new, unforeseen historical knowledge.

### Selection

We now have the capacity to pluck sources from multiple repositories, produce sources in-house through reformatting processes, or develop application programming interfaces (APIs) that assemble large datasets. Whereas archivists may be bound by an institutional or administrative acquisition policy, developers of digital historical representations have fewer constraints on aggregating historical information, bounded only by the limits of technology and storage capacity, and often motivated to achieve encyclopedic breadth. The effect, according to Roy Rosenzweig, is that historians must now "[think] simultaneously about how to research, write, and teach in a world of unheard-of historical abundance and how to avoid a future of record scarcity."[14] Enhanced digital capacities have placed more, not less, significance on the selection of historical materials.

Recent archival literature shows that the selection process, despite efforts to maintain objectivity, always possesses an element of subjective appraisal, what humanities scholars define as *interpretation*. Interpretation occurs the moment an archivist must decide which records to keep and which to discard. Margaret Hedstrom argues that "Decisions about which records to describe in greater

[13] Duranti and Thibodeau, "The Concept of Record in Interactive, Experiential and Dynamic Environments," 51.

[14] Rosenzweig, "Scarcity or Abundance?"

detail, and which to digitize for remote access, will influence the characteristics of the documentary past for many users of archives."[15] The selection process (like the processes of search and metadata, as we shall see) carries with it a complex series of decisions—or interpretative moments—that take into account numerous factors ranging from institutional demands to broader social and cultural considerations, all of which affects the historical knowledge inscribed within the content. Terry Cook and Joan Schwartz take this interpretative dimension further by characterizing the process as a performative act. "The archivist is an actor, not a guardian; a performer, not a custodian."[16] Cook and Schwartz encourage archivists to shed the Jenkinsonian ethos that dictates impartiality with respect to archival materials. Performativity incurs a closer involvement in the formation of the documentary record at all points in the record life cycle. If the activity of archival processing is understood as a "performance," then the archivist becomes a "mediator" or "facilitator" tasked with balancing the preservation needs of the record with access to the user.[17]

Digital historical representations undergo a similar interpretative act in the selection of historical content. Problems identifying a selection process often arise because of the difference in the order of magnitude of the data. Instead of working with finite physical holdings, digital historical representations may work with materials that no individual could ever possibly digest in a lifetime. The scale of a representation's data or content, however, must not overshadow the interpretative dimension that accompanies its selection. Developers, like archivists, must accept responsibility for the decisions that they make when establishing limiting criteria, however wide they may seem. Users, meanwhile, must recognize strengths and deficiencies in the comprehensiveness of a representation's content when conducting research or analyzing evidence. Finally, evaluators must consider a representation's engagement (or nonengagement as the case may be) with recent trends in historiography and therefore provide a telling indication of the level of scholarship and rigor undertaken. In all three cases, these decisions depend not just on calculating a quantifiable measure of completeness, but awareness of available historical information that can only come from knowledge of complementary work in the field.

The website The Valley of the Shadow is an example that illustrates the significance of the selection process in framing a digital representation's

[15] Hedstrom, "Archives, Memory, and Interfaces with the Past," 40.

[16] Terry Cook and Joan M. Schwartz, "Archives, Records, and Power: From (Postmodern) Theory to (Archival) Performance," *Archival Science* 2 (2002): 181.

[17] Cook and Schwartz, "Archives, Records, and Power," 183. The term "mediator" has been used elsewhere to describe the archivist, including Francis X. Blouin, Jr. and William G. Rosenberg, "Part 2 Introduction: Archives in Production of Knowledge," in *Archives, Documentation, and Institutions of Social Memory*. The term "facilitator," used in a similar context, appears in Sue Breakell, "For One and All: Participation and Exchange in the Archive," in *Revisualizing Visual Culture: Digital Research in the Arts and Humanities*, ed. Chris Bailey and Hazel Gardiner (London: Ashgate, 2010), 104.

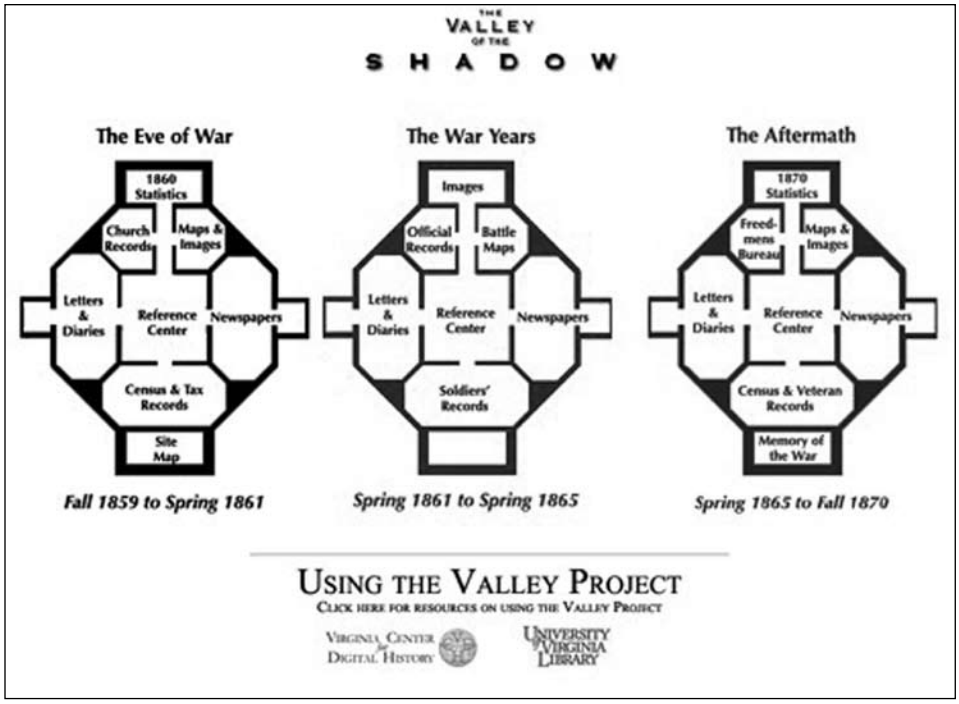


**FIGURE 2. The Valley of the Shadow main interface. From: http://valley.lib.virginia.edu/VoS/choosepart.html, accessed 20 January 2011.**

historical scope and context. The Valley of the Shadow pioneered digital history as a viable medium for research and teaching in the early 1990s during the infancy of the World Wide Web. Although the predominantly "Web 1.0" interface is now beginning to show its age, the site continues to garner awards and attention as a model digital historical representation.[18] For our purposes, the site is useful because it identifies itself as a "digital archive of primary sources… [that] is more like a library than a single book," thereby inviting archival theory into its consideration.

Ed Ayers, the site's creator, developed an experimental repository that could challenge common historical assumptions about the Civil War by constructing a site on a simple yet powerful selection principle, which predated the 1993 Web-based incarnation. According to the site: "[Ayers intended] to examine two places close to the border between the North and the South to see how people in such proximity and similarity went to war[.] [H]e studied maps and guides to military units and indexes of newspapers to find two areas centrally

[18] "The Valley of the Shadow Awards and Press Coverage," http://valley.lib.virginia.edu/VoS/usingvalley/award.html, accessed on 20 January 2011.

involved in the Civil War from start to finish."[19] This statement, which appears in the introductory section of the site, serves as the starting point for an analysis of the site's selection criteria. Ayers selected two counties, Augusta County, Virginia, and Franklin County, Pennsylvania, that through their geographic proximity would form a comparative framework with which to test hypotheses about the Civil War period. The selection of the two counties constitutes a complex series of decisions based on Ayer's command of Civil War historiography and the self-imposed criterion to choose counties according to geography. The project team scoured local archives and libraries to unearth and make digitally available every obtainable resource related to the two counties, including personal diaries and letters, census records, soldiers' records, battle maps, and entire runs of newspapers (see Figure 2). The digitally accessible archive would enable users to conduct a number of possible inquiries, such as tracing political thought across the years immediately before, during, and after the Civil War or studying the effects of slavery on the counties' social and economic institutions.[20] The digital archive, in other words, aims to be a microcosm that can both support and challenge a broader Civil War narrative.

This conceit obliges us to analyze the site's selection process in terms of the representativeness, or comprehensiveness, of its Civil War content. The Valley of the Shadow promises it has a complete archive of available materials with which to draw sound historical conclusions, in essence re-creating the experience of sifting through archival records and drawing connections that span multiple repositories and media. Nonetheless, the ease with which the user can access these records should not conceal the initial careful selection of the counties. Besides adhering to the self-imposed geographic constraint, Ayers undoubtedly considered whether the counties shared equally representative collections of common document types such as local newspapers, diaries, and soldiers' records, which required significant interpretative analysis. He needed to determine in advance whether users would be able to interrogate these collections and derive plausible conclusions. Had the archival records for Augusta County, say, possessed more diaries or a more complete newspaper series than Franklin

[19] "The Story behind the Valley Project," The Valley of the Shadow, http://valley.lib.virginia.edu/VoS/usingvalley/valleystory.html, accessed 20 January 2011.

[20] Evidence of Ayers's command of Civil War historiography can be found in an electronic article that he and William G. Thomas III produced using the Valley of the Shadow archive. In "The Differences Slavery Made: A Close Analysis of Two American Communities," the authors use the database to complicate the assertion that slavery was the primary factor precipitating the war. In the introduction, they argue "that slavery did not bear a simple relationship to emergent forms of modernity in the economic, cultural, or political realm. The very pervasiveness of slavery throughout the South meant that it exerted complex effects on every aspect of society…." Edward L. Ayers and William G. Thomas III, "The Differences Slavery Made: A Close Analysis of Two American Communities," http://www2.vcdh.virginia.edu/AHR/, accessed 20 January 2011. For a critique of the electronic article that raises important questions about the use of new media formats to construct a historical argument, see Allan Megill, *Historical Knowledge, Historical Error: A Contemporary Guide to Practice* (Chicago: University of Chicago Press, 2007), 5–13.

County, the site would have run the risk of skewing potential conclusions toward one county or the other.

If we take a step back, therefore, we can see that even the site's overall comparative framework determined documentary selection. Although one may conduct a study using sources from either of the two counties, the more worthwhile studies compare records from both. Other selection decisions found throughout the site reinforce this element. The project team concedes that in the indexing of the counties' newspaper serials, "only those national events that touched directly on Augusta or Franklin County" were included, while international events were not indexed at all.[21] Such decisions affect how a user interacts with the site, by steering the user toward predominantly localized conclusions. As evaluators, we may legitimately raise the question of whether, despite the site's claims of providing a representative set of records of two central regional participants in the Civil War, users have the capacity to derive conclusions that may shed light on a broader national context. By interrogating the site's archival selection, we introduce a question central to the effectiveness of the site, specifically, whether the representativeness of the records translates to the representativeness of the counties within our collective understanding of the Civil War.

With The Valley of the Shadow, we see selection criteria enforced at every level of the site, from the selection of the two counties, to the selection of individual record series, to the meta-selection of a comparative framework for the site. We can also begin to see how an assessment of a digital historical representation's selection can lead to larger questions about the representativeness of historical content, which ultimately points to the archival matter of trust. A representation is only as effective as the extent to which the user trusts that the creators have exhaustively considered available historical content. Such questions even become more evident when we consider the other two areas: search and metadata.

### Search

Digital historical representations promote recombination of information and with it the prospect for enhanced independent discovery. Steve Anderson remarks that the rise of "database histories" leads to "histories that are comprised not of narratives that describe an experience of the past, but collections of infinitely retrievable fragments, situated within categories and organized

[21] "About the Valley Newspapers," The Valley of the Shadow, http://valley.lib.virginia.edu/VoS/newspapers/about.html, accessed 20 January 2011.

according to predetermined associations."[22] A database of "retrievable fragments" invites users to generate associations previously unconsidered, but, at the same time, it can establish the dangerous precedent of generating erroneous narrative constructions. A search query can lead to the false expectation that all hits within the query possess an inherent link to one another. Users may be tempted to construct false relationships among disparate pieces of information, swayed by the notion that they share some commonality under the umbrella of a search term.

Historical recombination therefore requires sufficient contextual means to search for relevant data. Constructing sound historical questions is just as critical as accessing content. Representations, however, often do not guide users to conduct searches that maintain contextual integrity. How often do we come across the phrase, "browse the archive" with an empty search bar awaiting input from the user? Frequently this seeming innocuous activity can have an intimidating—even panic-inducing!—effect. What can I search for? Which terms are acceptable and which terms will produce zero results? As Douglas R. White and William Kules point out, "search engines, bibliographic databases and digital libraries provide adequate support for users whose information needs are well-defined. However, they do not work well in situations where users lack the knowledge or contextual awareness to formulate queries or navigate complex information spaces."[23] Digital historical representations have a basic responsibility to reveal the scope of their content.

Scholars have shown how the archival interface, the gateway to search, plays an important part in fashioning contextual knowledge. Building upon the interpretative qualities found in content selection, as discussed in the previous section, Margaret Hedstrom argues that archivists also build interpretation into computer interfaces. The interface functions as a "boundary where archivists exercise power and negotiate over what constitutes legitimate evidence of the past, and less directly, shape social memories."[24] The interface contributes to the archive's representational form by determining the limitations and constraints upon information access. From traditional finding aids and indexing schema to sophisticated digital design features, an archival interface governs use. In our triad of selection, search, and metadata, the search interface serves as the bridge linking selected content with its descriptive and provenancial metadata, all the while regulating the interpretative properties of both.

[22] Steve Anderson, "Past Indiscretions: Digital Archives and Recombinant History," forthcoming in *Interactive Frictions*, ed. Marsha Kinder and Tara McPherson, University of California Press, http://www.rimusnoc.com/download/pdf/digital%20cinema/anderson_article1.pdf, accessed 20 January 2011.

[23] The quote was reprinted in Mike Pringle, "Do a Thousand Words Paint a Picture?," in *Revisualizing Visual Culture*, 22.

[24] Hedstrom, "Archives, Memory, and Interfaces with the Past," 26.

Archival theory provides the conceptual framework with which to assess how an interface prioritizes data or records, to suggest topics of inquiry, or to guide the user toward historical conclusions. A well-structured system for search and retrieval can tear down the barriers separating what Thomas Kirchhoff et al. call "digitization islands." They argue that "the major advantage of…a joint portal [that unites cultural heritage databases] is that it connects [the databases] in one user interface. Objects can be searched and retrieved using uniform search procedures, and presented in uniform result lists."[25] Evaluation of a digital historical representation should include an assessment of how the search interface facilitates the construction of contextual bonds among its content.

For the purposes of limiting our discussion of search, we will focus on search interfaces that employ keyword schema. Developers have a range of possible keyword schema, from the traditional authoritative subject keyword headings, to in-house thesauri, to folksonomies, as well as numerous interfaces with which to display the schema. To illustrate the range of possible schemas and interfaces, which in turn demonstrate a range of control over the search process, three sites have been selected: The History of the Accademia di San Luca, c. 1590–1635: Documents from the Archivio di Stato di Roma; Gulag: Many Days, Many Lives; and the Hurricane Digital Memory Bank. Each site employs a set of search functions including controlled vocabulary menus, tag clouds, and open search mechanisms. Comparison of these features captures the tension between constructing a system with a controlled structure and one that relies upon user input. Analysis of these systems, as we shall see, invariably raises the question of a representation's visual organization of information. How a representation suggestively guides search, categorizes keywords, and prioritizes query results reflects a convergence of decisions in design, aesthetics, and information management.[26] An assessment of a representation's search must therefore approach the issue not as a simple algorithmic optimization of keywords, but as a mixture of interface features and contextualizing information that contribute to the overall search experience.

The National Gallery of Art's History of the Accademia di San Luca is an example of a controlled, contained search system. A brief introduction explains that the project team has assembled a "searchable database [that] provides access to a complete transcription of every extant notarial record of the period from the Archivio di Stato di Roma."[27] The academy flourished during the Renaissance as a gathering place for some of Italy's most influential artists and thinkers, and political and ecclesiastical figures. Applying analysis of the site's

[25] Thomas Kirchhoff, Werner Schweibenz, and Jörn Sieglerschmidt, "Archives, Libraries, Museums and the Spell of Ubiquitous Knowledge," *Archival Science* 8 (2008): 255.

[26] See, for example, Chris Bailey, "Introduction: Making Knowledge Visible," in *Revisualizing Visual Culture*.

[27] "A Brief History of the Accademia di San Luca," The History of the Accademia di San Luca, http://www.nga.gov/casva/accademia/intro.shtm, accessed 20 January 2011.

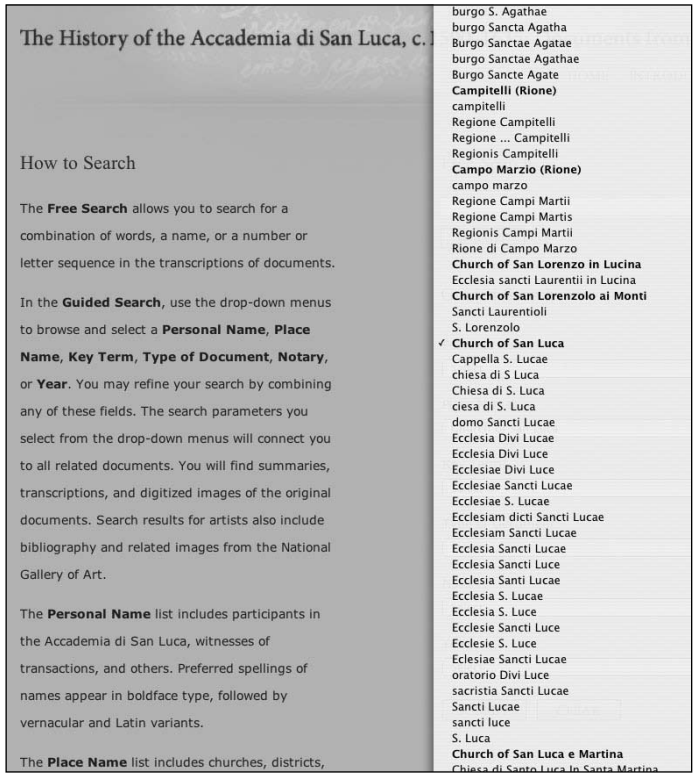


**FIGURE 3. Drop-down menu from History of the Accademia di San Luca, http://www.nga.gov/casva/accademia/intro.shtm, accessed 20 January 2011.**

content selection, we find that the digital archive was constructed around a finite, comprehensive set of records related to the administrative history of the academy. The user is assured that the archive contains "every extant notarial record" available to the project team.

Identifying the site's selection process can help explain the layout of the search interface. The user has two choices in how to conduct a search: a "free search" marked by a single search bar, or a guided search with predetermined fields including personal name, place name, type of document, keyword term, and year. Each field contains a drop-down menu with a complete list of indexed terms. Even more, the menus provide all possible variant spellings and appellations for each term. Personal names include "preferred names [appearing in boldface type], followed by vernacular and Latin variants," while the place name category "includes churches, districts, and neighborhoods within Rome; outside Rome it provides the names of cities, countries, and regions."[28] The term "Church of San Luca," for instance, possesses an impressive twenty-five

[28] "How to Search," The History of the Accademia di San Luca, http://www.nga.gov/casva/accademia/search_eng.shtm, accessed 20 January 2011.

variations (see Figure 3). The breadth of available terms and appellations instills confidence that a search query will result in the desired historical figure or place. The menus' expansiveness has the added benefit of also circumscribing the historical content by chronicling the actors, locations, and themes that appear in the documents. Both knowledgeable and casual users may grasp the historical scope of the digital archive simply by perusing the drop-down menus.[29]

In contrast to the hierarchical structure of the Accademia menu system described above, Gulag: Many Days, Many Lives employs a tag cloud as its primary search mechanism. A team of Russian and American scholars and curators created the site to document the history of the Soviet-era gulag, in particular Perm-36, which was converted into the Gulag Museum. Gulag has a different overall purpose from History of the Accademia di San Luca in that its contents


**FIGURE 4. Tag cloud from Gulag: Many Days, Many Lives, http://gulaghistory.org/items/tags, accessed 20 January 2011.**

[29] "How to Search," The History of the Accademia di San Luca.

have been assembled and curated to produce virtual exhibits that guide users through a detailed tour of Perm-36 or a discussion of life in the gulags. In addition to the virtual exhibits, users have the option to browse all contents of the site's "archive" or to "browse by tag," which directs the user to the tag cloud.

A tag cloud visually organizes keywords by accenting the most frequently cited terms in a larger font size. The tag cloud can reveal in a single snapshot the project team's organizational and contextual vision for the digital objects. The emboldened, enlarged words "prisoners," "labor," "arrest," "deprivation," and "guards" draw attention to the prisoner experience, which is reinforced elsewhere by other contextual information found in the exhibits (see Figure 4). The introduction to the exhibit "Days and Lives" promises to "[take the user] inside the brutal system of forced labor concentration camps and the internal exile institution called the Gulag."[30] The exhibit also pairs the user with a prisoner, whose story frames the other themes of the virtual narrative. The tag cloud, in highlighting particular keywords that point to the prisoner experience, invites the user to access artifacts reaffirming this perspective. The user is encouraged to build a unique experiential understanding of the gulag with the assumption that the terms in larger font represent the most important elements of the experience.

Both History of the Accademia di San Luca and Gulag employ variations on a controlled vocabulary system, separated more by degree in structural integrity than by overall intent. The former embeds a rigid hierarchical structure within the drop-down menu system, while the latter employs a looser construction with the tag cloud by assembling all keywords into a single visual "box" prioritized by font size. Despite differences in design and presentation, the creators have selected keywords and metadata fields in advance, implementing a top-down approach that contrasts with the bottom-up approach found in the third and final example of a search interface.

Hurricane Digital Memory Bank was created by George Mason's Roy Rosenzweig Center for History and New Media and the University of New Orleans to document the aftermath of Hurricane Katrina. Unlike the previous two sites, Hurricane is a user-driven site where anyone can freely contribute content in the form of digital images, video clips, written anecdotes, blog postings, or podcasts. The site envisions itself as participating in a "growing practice of using the Internet to preserve the past through 'digital memory banks.'"[31] Just as the site does not place restrictions on the content available for upload, users are free to tag their content however they choose. Hurricane in turn creates a two-tiered system for searching the content. Users may elect to "browse"

[30] "Introduction: Days and Lives," Gulag: Many Days, Many Lives, http://gulaghistory.org/exhibits/days-and-lives, accessed 20 January 2011.

[31] "About the Hurricane Digital Memory Bank," Hurricane Digital Memory Bank, http://hurricanearchive.org/about/, accessed 20 January 2011.

the entire site, or they may select from a list of curated collections. The browse function permits search either by content type ("Stories," "Image," "Other Files," "Outside Links," "Map") or by a tag cloud. Whereas Gulag limits the tag cloud to a preselected set of keywords, the memory bank's tag cloud contains hundreds of terms assembled alphabetically and restricted only by the contributors' own descriptive imagination. With the exception of the terms "Rita, Katrina, New Orleans, and st bernard parish [*sic*]," there is little visual distinction among the terms, which defeats the purpose of the tag cloud as a system for guided search (see Figure 5).[32]

A cursory inspection of the site's contents reveals a wide discrepancy in uploading practices. Content accessed through the tag cloud system generally contains many more tags added by the individual source creators, but leaves little provenancial information about the source itself, while curated collections, such as those by the Smithsonian National Museum of American History and the United States Coast Guard, document the provenance of the items much better, but fail to tag them with useful keywords.[33] The overall effect is an archive that engenders little confidence in discoverability. Tags are misspelled (two terms for "chocolate city" and "chocalate city" [*sic*]) or follow no rules for consistency in subject headings, names, or places ("francesco di santis," "francisco," and "francisco di santis").

The differences in the three search systems discussed here, especially with regard to discoverability of content, reflect telling information about their methods of representing historical knowledge and expectations for how that knowledge will be used. The drop-down menus for History of the Accademia di San Luca, organized by fields such as name, place, and year, encourage users to reconstruct the *network* of people and locations associated with the academy. By combining terms from each field, users may build contextual relationships, whose strength derives from the controlled vocabulary of the menus. The menu system reinforces the archive's underlying historical narrative. Users are encouraged, through exploratory searching of the records, to uncover the "story and transformations [of the academy, which are] not to be found in a single written source; rather it has to be reconstructed from the fragmented documentation that has been recovered and rediscovered in the collections of the Accademia . . . ."[34] With its visual emphasis on certain key terms, the tag cloud reflects the Gulag project team's objective to reconstruct the emotional,

[32] "Browse the Memory Bank," Hurricane Digital Memory Bank, http://hurricanearchive.org/browse/, accessed 20 January 2011.

[33] The two mentioned collections are listed as Smithsonian National Museum of American History Hurricane Katrina Photos and United States Coast Guard Released Photographs of Hurricanes Katrina and Rita. Hurricane Digital Memory Bank, http://hurricanearchive.org/collection/, accessed 20 January 2011.

[34] "Introduction," The History of the Accademia di San Luca, http://www.nga.gov/casva/accademia/intro.shtm, accessed 20 January 2011.


**FIGURE 5. Tag cloud from Hurricane Digital Memory Bank, http://hurricanearchive.org/tags/, accessed 20 January 2011.**

physical, and psychological prisoner experience. Most telling in this example are the categories of terms that are *not* present in the tag cloud, which in effect inhibits users from asking certain types of questions. For instance, terms (at least those in English) relating to Russian state institutions that may shed light on the bureaucratic infrastructure of the gulags are absent from the tag cloud, thereby limiting the site's political contextualization. Finally, Hurricane Digital Memory Bank's unregulated tagging system renders materials virtually undiscoverable through conventional search mechanisms. Instead, the sea of tagged terms—

the heart of the site's search capability—functions much better when considered as a historical record unto itself. The site entices users to consider how anonymous and venerated institutions alike chose to document the tragic event. Each contribution to the memory bank reflects a unique conception of search that transcends in historical significance the results of any individual search query.

These three examples of keyword mechanisms suggest how the subtle interplay between visual and textual features within a representation's interface can guide the user toward contextual information, which in turn can influence how a representation is searched or browsed. In short, the search interface is one element of the overall user experience, an area that requires much more study within digital historiography.

### Metadata

The previous two sections discuss how selection and search contribute to a user's trust in a digital historical representation. With selection, trust is established when the creator demonstrates the representativeness of historical content, while a search interface provides clues that reveal the content's intellectual organization. Both areas depend upon an even more fundamental level of trust derived from the historical content's reliability. Heather MacNeil reminds us that archival records earn their trustworthiness according to how they adhere to accepted archival principles that ensure they are "uncontaminated by the distorting influence of time, bias, interpretation, or unwarranted opinion…." She writes:

> The methods for assessing a record's truth-value are underpinned specifically by an assumption that a unitary and stable relationship exists between a representation (that is, a record) and its referent (i.e., a pre-existent reality); and an attendant belief in the capacity of language to reflect and give presence to a world of fixed and generalizable knowledge about the nature of a record and the conditions necessary to ensure its trustworthiness.[35]

A record's trustworthiness, therefore, depends on preserving the integrity of its form and content. What are the content's referents, and in the case of digital historical representations, how do elements like the interface and computational tools establish those referents? Advanced electronic records systems enable associations with multiple record creators and consequently multiple contexts, which clearly raises difficult questions about how to assess the authority of those contexts.

Digital historical representations embrace the potential to generate multiple contexts by providing the means to search and locate historical data in a

[35] Heather MacNeil, "Trusting Records in a Postmodern World," *Archivaria* 51 (2001): 42.

seemingly infinite number of ways. This unbridled freedom to sort data consequently threatens the salience of historical context. In the rush to digitize texts, images, and data, digital humanists prioritize informational content over the integrity of the original record, potentially breaking a record's critical chain of custody and creating what Joanna Sassoon describes as "a databank of orphans which have been removed from their transactional origins and evidence of authorial intent."[36] Sassoon's remark suggests that what is at stake with digital historical representations is their reliability as sources of historical information.

We must consider methods for evaluating the reliability of historical content that must come not only from the information contained within the documentary content, but from the contextual information surrounding the content as well. According to archival theory, a sound metadata infrastructure establishes the reliability of digital records. Anne Gilliland et al. define metadata in the broadest sense as

> all types of structured information, including archival description, that is created manually or automatically by recordkeeping systems including metadata that documents the juridical-administrative, business and technical contexts within which records are created; identifies records and delineates how the records behave, their function and use; identifies and describes the relationships within and between records and other information objects; and expresses and supports how records should be managed, and what happens to them over time.[37]

Gilliland et al., in their work with InterPARES 1 and 2, explore how metadata contributes to the construction of context. Context is built within a metadata structure by dictating the relational behaviors of records. Levels of contextual information in metadata are quite varied and can reveal diverse information about the use value of digital records.[38]

Archival theory has shown that trust in metadata is synonymous foremost with trust in provenance. In their seminal article, "The Power of the Principle of Provenance," David Bearman and Richard Lytle argue that archivists bring a unique perspective to the "organization and management of information by prioritizing records' provenance, i.e. organizational activity, especially how organizations create, use, and discard information."[39] Although they do not

[36] Sassoon, "Beyond Chip Monks and Paper Tigers," 139.

[37] Anne Gilliland et al., "Towards a 21st Century Metadata Infrastructure Supporting the Creation, Preservation and Use of Trustworthy Records: Developing the InterPARES 2 Metadata Schema Registry," *Archival Science* 5 (2005): 44, fn 1.

[38] Gilliland et al., "Towards a 21st Century Metadata Infrastructure." See also Joanne Evans, Sue McKemmish, and Karuna Bhoday, "Create Once, Use Many Times: The Clever Use of Recordkeeping Metadata for Multiple Archival Purposes," *Archival Science* 5 (2006).

[39] David A. Bearman and Richard H. Lytle, "The Power of the Principle of Provenance," *Archivaria* 21 (1985–1986): 14.

explicitly frame their discussion around the concept of context, we may infer from their understanding of provenance that contextual information is derived from the relationship between "creating activity and information created by organizations." Archivists, they suggest, "can thus know from provenance rather than from subject indexing certain elements of the intellectual contents of records…. If the archivist's use of provenance in arrangement and description…is reversed, a potential exists for a practical and powerful means of gaining access to and managing information."[40] Archival scholars have since expanded the principle of provenance beyond predominantly organizational, or institutional, activity to include numerous other layers of contextual relationships. Building on the work of Terry Cook, Tom Nesmith, and Eric Ketelaar, Jeanette Bastian reconceptualizes the principle of provenance within the framework of a "community of records," which encompasses a wider understanding of societal activity and the "multi-representational" power of records to contextualize such activity. Bastian challenges us to consider how provenance may contribute to a "wider context" when she asks, "How far should archivists go in establishing a context that will enable the full interpretation of the record?"[41]

If we accept the principle of provenance—updated according to the democratizing, postcolonial, and postmodern perspectives by Bastian, Cook, Nesmith, and others—we may conclude that provenancial metadata helps digital historical representations preserve the integrity and reliability of their information. An evaluation of a representation, therefore, must address whether and how provenancial metadata is sustained across its functions, a task that is more difficult than may at first seem, considering the ease with which a representation may draw upon sources across multiple repositories or fashion new sources through digitization processes. According to Luciana Duranti, electronic records possess fundamentally different compositional features from physical records. She writes, "[W]hat distinguishes electronic records from their traditional counterparts is not their components, which remain the same, but the fact that the record's elements through which they are manifested are not inextricably joined to one another, as in traditional records. They exist separately as metadata, and can be managed separately, unless they are consciously tied together for the purpose of creating and maintaining reliable and authentic records."[42] For digital representations, therefore, it is critical that historical data retains its provenancial metadata given the high probability of losing such context in the course of data recombination.

[40] Bearman and Lytle, "The Power of the Principle of Provenance," 22–25.

[41] Jeannette Allis Bastian, "Reading Colonial Records Through an Archival Lens: The Provenance of Place, Space and Creation," *Archival Science* 6 (2006): 281–83.

[42] Duranti's conclusions are based on the findings of a research project conducted at the University of British Columbia in collaboration with the U.S. Department of Defense. Luciana Duranti, "The Impact of Digital Technology on Archival Science," *Archival Science* 1 (2001): 48.

To illustrate the importance of preserving archival provenance through metadata, let us consider History Engine, a site founded by professors at the University of Richmond that has since attracted over a dozen collaborating universities. History Engine, according to the homepage, "is an educational tool that gives students the opportunity to learn history by doing the work—researching, writing, and publishing—of a historian."[43] Students select a single document from a local archive, investigate its broader historical significance using limited secondary literature (typically no more than one or two sources), and publish a short "episode" that describes their findings. The site grew out of the course, The Rise and Fall of the Slave South, taught by Ed Ayers at the University of Virginia. Most episodes highlight source materials from local archives, which inevitably skew subject and geographic coverage because all of the participating schools are located along the East Coast and in the South. Entries must follow a strict formula founded by the site's creators, which includes limits on an episode's word count. A "Teacher's Guide" describes how students should compose their episode as a narrative rather than as an interpretative argument: "The goal of the episode is not to recount textbook-like facts, but to try to understand the past from the perspective of the Americans who lived it.... [The students'] job will be to tell the story of this source and explain its significance to American life.... [They] will make an important contribution to understanding the American past by uploading [their] finished episodes to the History Engine database."[44]

The bottom-up construction of a database of brief "episodes" in "American life" presents a tantalizing opportunity to discover and aggregate historical information that remains hidden in small archives. History Engine's creators hope to bypass rote learning from textbooks, proposing that students instead have the "opportunity to become historians." The site offers a number of ways to search and explore the content from "Basic" and "Advanced" text searches to visualization tools that plot episodes and their accompanying keyword tags along time plots and Google maps.

Were History Engine only a vehicle for motivating students to visit archives, work with original documents, interact with their local archivists and librarians, and experience the simulated pressures of publishing, it would represent a remarkable success. The site's interest in archiving all student work in a central database, however, reflects the creators' loftier ambitions. They hope that the wikilike database of student-composed episodes will "provide a previously untapped resource for researchers and academics in general," steering History Engine away from its original intent as an innovative collaborative teaching tool

[43] History Engine: Tools for Collaborative Education and Research, http://historyengine.richmond.edu/, accessed 20 January 2011.

[44] "Your Goal," History Engine: Tools for Collaborative Education and Research, http://historyengine.richmond.edu/pages/students/goal, accessed 20 January 2011.

toward developing a research portal.[45] In a 2009 article, the creators describe History Engine as an "online" or "history archive that would be both interesting and useful to students, the general public, and historians."[46] Their intent reveals a flawed archival understanding that reinforces a lackluster way of conducting historical research and working with cultural heritage materials.

Like so many digital humanities projects, History Engine blurs the boundaries between scholarship and pedagogy, a topic that requires much more extensive discussion in the humanities and information fields. For the purpose of this article, let us focus on the site's flawed archival understanding, which will unavoidably overlap with some of its pedagogical objectives. The developers assume that, with enough entries, users may begin to visualize historical patterns, spurring them to conduct historical research that cuts across time, geography, and archival boundaries. Unfortunately, the method of archiving episodes robs them of historical context. Despite claims that students must "contextualize" their episodes, the History Engine episode format accomplishes precisely the opposite by inhibiting sound historical contextualization at every level, from the individual episode all the way up to the database-wide search functionality. We may trace the problem back to a decision to deliberately neglect applying archival provenance at the episode unit. The episode entitled "Runaway Slaves," which appears as an entry for the common tag "African-Americans," serves as a typical example. "Runaway Slaves" is supposed to describe an advertisement for three slaves who escaped a Louisiana plantation in 1827, although it is extremely difficult to determine this from the episode interface and narrative. Although the episode lists a date ("27 January 1827"), a location ("ST CHARLES, Louisiana"), and two keyword tags ("African Americans," "Slavery"), it does not list a title for the document used to write the episode, or even a documentary type. We do not learn that the episode is documenting an advertisement until the fifth sentence of the last paragraph: "One such advertisement ran in the newspaper, *Le Courrier de la Louisiane*, advertising the loss of three slaves to a subscriber" (see Figure 6).[47]

This episode represents the site's muddled understanding of historical context as demonstrated through the site's disregard for archival provenance. The "Runaway Slaves" episode does contain a basic bibliographic citation that, with a little online searching, would pull up which libraries have the Louisianan

[45] "What Is the History Engine?," History Engine: Tools for Collaborative Education and Research, http://historyengine.richmond.edu/pages/about/what_is_the_history_engine, accessed 20 January 2011.

[46] Robert K. Nelson, Scott Nesbit, and Andrew Torget, "The History Engine: Doing History with Digital Tools," Academic Commons (2009), http://www.academiccommons.org/commons/essay/history-engine, accessed 20 January 2011.

[47] "Runaway Slaves," History Engine: Tools for Collaborative Education and Research, http://historyengine.richmond.edu/episodes/view/3233, accessed on 20 January 2011.

ABOUT | EXPLORE THE ENGINE | FOR TEACHERS | FOR STUDENTS

HISTORY ENGINE *Tools for Collaborative Education & Research*

**Runaway Slaves**

Date: January 27, 1827
Location: ST CHARLES, Louisiana
Tags: African-Americans, Slavery
Course: "Rise And Fall of the Slave South," University of Virginia

Rating: No votes.

The harshness of the institution of slavery is something that historians can agree on. The lifestyle that enslaved people had to endure is one that many cannot imagine today. It is not surprising therefore when slaves wished to be free of their masters and escape from plantations. The reasons for slaves running away varied but usually included fear of punishment or resentment because of punishment already administered, avoiding the harsh labor forced upon them, as well as sheer cruelty imposed by their masters. Although the great mass of Louisiana slaves remained at home, the number of runaway slaves in the state compelled the legislature to act. In an attempt to centralize the problem of runaway slaves, the Louisiana State legislature created acts in the 1820s and 1830s to set up central depots to which runaways, unclaimed for two months or longer, were to be transported. These depots kept the slaves together and used them on public works. Jailors were required to advertise the runaway slaves they kept in their prisons to successfully return them to owners. These jailors would be paid upong the death of a slave for expenses incurred with regard to the deceased.

Although Louisiana attempted to produce an efficient system to recover runaway slaves, it did not always function as intended. The legislature created several more parish depots in the late 1820s and this decentralized the effort to keep the runaway slaves in central locations such as Baton Rouge and New Orleans. Slaveholders took it upon themselves to advertise to the community the loss of a slave and incurred the expenses themselves. Plantation owners paid large sums of money to advertise slaves' looks and ages in the hope of their property being returned. One such advertisement ran in the newspaper, *Le Courrier de la Louisiane*, advertising the loss of three slaves to a subscriber. The subscriber offered a handsome reward to anyone that helped return Lewis, Isaac, and Snaiel to his fields. Judging from their descriptions, these slaves were very important slaves to their owner. Three strong young men missing from the plantation deprived the owner of not only labor bu capital and time. The January advertisement volunteered the information that one of the slaves had been missing since September, demonstrating the long time period a plantation owner could spend searching for runaway slaves. It it not certain if Lewis, Isaac, and Daniel ever returned to their owner. Runaway slaves cost their masters' time, money and a loss of control over a system that the economy of the antebellum Louisiana was based upon. The state believed that the situation of runaway slaves was grave enough to address it in their laws to gain a grasp on the problem. When this failed, the slaveholders turned to the resources of newspaper advertisements. The advertisements from newspapers in the South serve as reminders of the lengths that plantation owenrs would go to in order to insure the safe return of their chattel.

**Citations**

*New Orleans (LA) Le Courrier de la Louisiana*, January 6, 1827.

Loren Schweinger, "Counting Costs: Southern Planters and the Problem of Runaway Slaves, 1790-1860," *Business and Economic History* Vol. 28, No. 2 (1999): 267-276.

Joe Gray Taylor, "Runaways," in *The Louisiana Purchase Bicentennial Series in Louisiana History, Volume IV: Antebellum Louisiana, 1830-1860 &34;Part A Life and Labor&34;*, ed. Carolyn E. De Latte (Lafayette: Center for Louisiana Studies, 2004), 249-263.

HOME | SEARCH | DIGITAL SCHOLARSHIP LAB | HELP | RSS: NEW EPISODES



**FIGURE 6. "Runaway Slaves" History Engine episode, http://historyengine.richmond.edu/episodes/view/3233, accessed 20 January 2011.**

newspaper in their holdings. Archival provenance instead is absent in the isolation of the document from other documents. Instead of committing to either a close reading of a single document, as the site purports to encourage, or a more general history of a topic, the creators have hedged their bets and encouraged episodes to try to do both simultaneously. In the case of "Runaway Slaves," the author could have been encouraged to place the advertisement in a broader context by examining similar advertisements about runaway slaves that may have appeared in the same newspaper or other regional newspapers. This process would have mirrored more closely the activities of a historian, who selects documents by comparing them with other documents within the archive or across multiple archives. Furthermore, the restrictive episode format, which

emphasizes narrative over interpretation, prevents the author from conveying why the document was selected. Was the advertisement indicative of runaway slave advertisements from the period, or did it contain features that made it stand out from other examples?

The problems with the episode format are compounded when we move to the keyword tagging schema and the visualization tools associated with the tags. According to the creators, the database "[leverages] the metadata associated with each episode to produce historical visualizations." Metadata consists of three pieces of information including an episode's geotagging, a "historical date," and keywords. Leaving aside the first two fields, both of which are problematically handled by the site (Should the historical date signify the date of the historical event in question or the date the source was created? What if the creation date of the source is unknown or in doubt?), we find in the site's documentation regarding the creation of keywords no indication that students should follow standards for building a controlled vocabulary. According to the "Quick Guide," "Each episode should have at least two tags. There are a number of available tags in the database already but you may create your own tags if you think those are insufficient." History Engine then feeds an episode's keywords into its visualization tools, including a time plot. Plugging in the term "African-Americans," for example, produces a series of peaks and valleys along a timeline that indicates its frequency as a keyword (see Figure 7). At face value, the graph provides some contextual information regarding student contributions to the site, enabling us to conclude that more episodes were written about African Americans in the years 1865–1866 and 1895 than other years. The creators' intent, however, is not to document student activity but to document or quantify *historical* activity. We are expected to surmise that 1865 and 1895 may reflect some significant uptick worthy of further investigation, the protean data that may compel scholars to formulate an inquiry. The notion that there was an increase in African American discourse for certain years and not others, considering the centrality of race throughout the nineteenth century, produces only historical distortion. The time plot, regardless of how many episodes are generated by students, will never reflect anything more than the idiosyncratic interests of participating instructors and students and their access to local archival records. The "high hopes" of creating a "fine-grained account of U.S. history" in the form of "a large interpretive finding aid for historical sources located in archives and libraries across the country" reflects a misguided faith in quantitative results, one that belies any rigorous social science and, more to the point, archival methodology.[48] How many entries are required, or how many

[48] Nelson and Nesbit, "The History Engine: Doing History with Digital Tools."

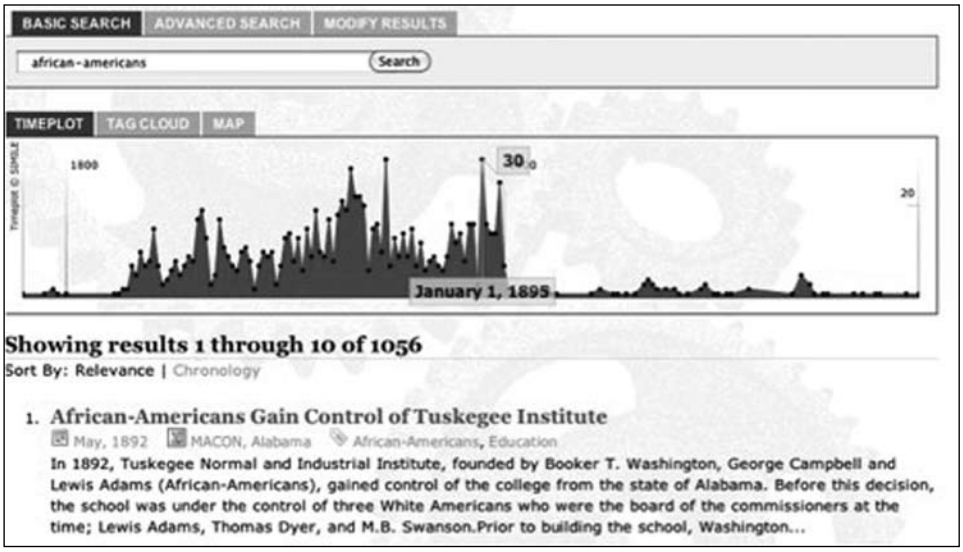


**FIGURE 7. History Engine time plot for search term "African-Americans," http://historyengine.richmond.edu/search/basic, accessed 20 January 2011.**

institutions must participate, before the database captures a reliable cross-section of historical activity? No quantity of episodes—certainly not the "tens of thousands of episodes" the creators hope to obtain—will ever capture a precise portrait of American history unless each episode adheres to rigorous metadata standards.

Most worth noting about this example of digital history is the brazen misappropriation of archival terminology such as "history archive" and "finding aid." When buzzwords such as these appear in digital historical representations, historians and information specialists should probe deeper for evidence of archival best practices. History Engine does not promote the construction of an archive composed of interrelated records, but rather a mass database of isolated documents, divorced from their archival provenance both physically and intellectually, which renders it incapable of serving as a portal to conduct historical inquiry.

A few simple changes may mitigate some of these deficiencies. Rather than limit episodes to a single document, students should be required to build episodes based upon sets of documents. This would guide students to consider historical contextualization built around selection, more stringent provenancial metadata, and standardized keywords, all the while learning to apply higher-level interpretative analysis. Finally, students should be permitted to publish a digital surrogate of each document, which would open further comparative possibilities across the database by allowing users to conduct their own analysis of the documents.

## Conclusion

Returning to the two examples of digital historical representations at the opening of this article, we find that digital historiography can illuminate more than the fact that they share archival processes in common. Identifying archival processes through digital historiography can raise pointed analytical questions that users and evaluators should consider. For the website A Story Like No Other, we should inquire about the representativeness of the thirty-three African American cultural heritage sites. If a user elects to explore certain sites and ignore others, how does that affect the overall contextual experience? Ultimately, we will want to question exactly what "story" about the state's complicated racial history is conveyed by learning about and visiting the sites. The Google Books Ngram project likewise highlights difficult questions about the evidential value of word trajectories. At the very least, archival theory can assist in weighing how to handle the multitude of metadata errors that are known to exist in the Google Books corpus. In his preliminary evaluation of the study, Geoffrey Nunberg writes: "…[While] the Harvard researchers have purged the research corpus of a large proportion of the metadata errors that have plagued Google Books, there are still a fair number of misdated works, and there's no way to restrict a query by genre or topic."[49] Digital historiography can help us theorize how to move from the data stream to guided close reading of selected texts.[50]

The three archival processes discussed in this article—selection, search, and metadata—build a framework for digital historiography that reconstructs a theoretical and methodological bridge between the historical and archival disciplines. An archival perspective to digital historical representations leads to the identification of interdependent information structures, including selection criteria, search interfaces, and metadata schema, that contribute to the contextualization of historical content. The integrity of that structure is measured according to the representation's potential to generate historical knowledge, and, more importantly, evoke humanistic inquiry. Admittedly, this framework for digital historiography also needs to incorporate disciplines within the wider cultural heritage sector besides archival studies such as library science, museum studies, and informatics, which are equally dedicated to understanding matters of representation, user interactivity, information retrieval, and cataloging.

[49] Geoffrey Nunberg, "Counting on Google Books," *The Chronicle of Higher Education* (16 December 2010), "Chronic Review," http://chronicle.com/article/Counting-on-Google-Books/125735, accessed 20 January 2011.

[50] The *Science* study sparked a healthy dialogue among scholars in the blogosphere about the utility of the Ngram Books Viewer. Dan Cohen, one of its strongest advocates, suggests that currently one of the biggest problems is the inability to move from the data's viewer to source texts. Dan Cohen, "Initial Thoughts on the Google Books Ngram Viewer and Datasets," *Dan Cohen's Digital Humanities Blog* (2010), http://www.dancohen.org/2010/12/19/initial-thoughts-on-the-google-books-ngram-viewer-and-datasets/, accessed 20 January 2011.

As digital historical representations appropriate archival characteristics and vice versa, we must adapt our methods of development, use, and evaluation by merging salient theoretical constructs from both the humanistic and information fields. Fortunately, archival theory is undergoing a "paradigm shift" of its own that may accommodate such an alignment, as we have seen. Archival theory has incorporated a renewed interest in the historicity of archival frameworks. At the center of this shift is a reassessment of the archivist's relationship to the formation and preservation of the documentary past. From selection and appraisal, to description, to building a metadata infrastructure, we have come to appreciate the archivist's involvement throughout the decision-making process, and the influence of social, professional, political, and personal values. New experimental archival models are currently being explored that enhance user and archivist participation, decentralize curation, challenge power structures, and form multiple layers of interpretation.[51] As a result, archivists face greater transparency in their methods, theory, and background, all of which becomes fertile ground for analysis and interpretation in digital historiography.[52]

Transparency is intimately linked with the archival notion of trust, and it should be applied with equal rigor to digital historical practices. Until now, determining trust in digital historical representations lacked a systematic approach that could deconstruct a representation's components in terms of use, organization, and presentation. Digital historiography brings into focus the transparency—or lack thereof—of each formative decision, from choices in the organization of content, to aesthetic design, to the construction of an interface. Open, variant contextualization—one of the lynchpins of digital history—depends upon transparency in a representation's production. The elements that are absent from a representation, such as a robust metadata schema, may be just as revealing as those elements that are accounted for by the creators. Similarly, for users to rely upon evidence derived from digital historical representations, they must demonstrate clearly the process by which information was searched, selected, and combined.

[51] The field of experimental archival forms is much too broad and diverse to cover here. For examples of recent investigative work, see the contributions in Carolyn Hamilton et al., eds., *Refiguring the Archive* (Dordrecht, The Netherlands: Kluwer Academic Publishers, 2002), as well as Isto Huvila, "Participatory Archive: Towards Decentralised Curation, Radical User Orientation, and Broader Contextualisation of Records Management," *Archival Science* 8, no. 1 (2008); and Magia Ghetu Krause and Elizabeth Yakel, "Interaction in Virtual Archives: The Polar Bear Expedition Digital Collections Next Generation Finding Aid," *American Archivist* 70, no. 2 (2007).

[52] The 2010 Society of American Archivists conference devoted an entire panel to the issue of archivists' transparency: Session 407—"Trust Me, I'm an Archivist: Transparency, Accountability, and Archival Documentation," chair, Laura Millar; held on 13 August 2010. Heather MacNeil's paper, "Trusting Description: Authenticity, Accountability, and Archival Description Standards," particularly captured the advantages and consequences, along with the justifiable reluctance shared by many archivists, to increasing transparency of their methods and personal history.

Finally, archival theory can serve as a cornerstone for evaluation of digital historical representations. We may consider the application of archival processes in representations analogous to the application of reference citations in a scholarly text. Working through a text's footnotes, bibliography, table of contents, and index can reveal the layers of the author's intellectual framework. Piecing together these elements reconstructs and illuminates the author's historical argument. The same is true for a digital historical representation: identifying the archival structures can help reconstruct context. In both cases, the reconstructive process indicates the quality and nature of the author-creator's historiographic and interpretative relationship with historical materials, and the expected relationship of the user-reader to those materials.

Terry Cook reminds us that the notion of archival relationships has not changed with the advent of sophisticated digital means of organizing content or data:

> Archivists are now perceiving that a world of relational databases, of complex software linkages, of electronic office systems, of hypermedia documents, of multi-layered geographical information systems, is, when all the high-technology rhetoric is put aside, still a world of information relationships, of interconnections, of context, of evidence, of provenance. Re-creating such relationships for complex electronic records should be no different for the archivist, at a conceptual and theoretical level, than unraveling the interconnections of the many series of records that were typical of the nineteenth-century office, and linking them to their animating functions and creators.[53]

Cook makes a valuable point that we should not discard fundamental archival principles such as evidence, context, and provenance with technological advancement. Instead, we should understand how archival principles are transformed through technology. This motive to understand technology's transformative effects drives digital historiography. While the means to build contextual relationships among data or content have expanded in complexity, their underlying importance for securing trust has not diminished. Cook's mentioning of geographic information systems and hypermedia documents invites us to consider participant groups beyond archivists for assuring that representations preserve strong contextual relationships. The responsibility to apply archival standards to historical representational practices should not rest with archivists alone, but should include historians, information specialists, engineers, curators, librarians, and many others.

As we have seen in the digital history examples provided here, a comprehensive approach to the development, use, and evaluation of representations requires a combination of subject, archival, and technical knowledge. Given the current divisions within secondary education, such efforts

[53] Cook, "What Is Past Is Prologue," 41.

may, in the short term, continue to require a collaborative model. In the long term, we should consider how to (re-)integrate humanities and information academic programming.[54] By revising accepted theories and methodologies from both archival studies and history, digital historiography can provide a more seamless transition between analog and digital work and pave the way for new curriculum models. Digital historiography also reminds us that in spite of the advent of advanced computational processes, digital tools, and new representational formats, we must continue to preserve the fundamental humanistic activities of close reading, interpretation, and historiographic engagement.



[54] NYU's Archives and Public History program has taken up the call to teach archival theory alongside digital history theory and practice. It recently unveiled a new website showcasing a revamped academic program, Archives and Public History Digital, http://aphdigital.org/, accessed 20 January 2011. Encouraging discussions "across the aisle" occurred at panels held at the 2010 Digital Humanities Conference in London and at a day-long workshop conducted by Joshua Sternfeld, Johanna Drucker, and Stephen Davison at the 2009 Archival Education and Research Institute held at UCLA. The latter event was reported in Sarah Buchanan, "Accessioning the Digital Humanities: Report from the 1st Archival Education and Research Institute," *Digital Humanities Quarterly* (2010), http://www.digitalhumanities.org/dhq/vol/4/1/000084/000084.html, accessed 20 January 2011.