Theodore Calvin Pease Award

Archives on the Internet: Representing Contexts and Provenance from Repository to Website

Emily Monks-Leeson

Abstract

This paper examines two repositories of digital material to investigate how representational practices take shape in online archives. Specifically, it questions whether Internet archives work against the archival principle of provenance or reinterpret it to create new and flexible contexts. *Digital, online, and website archives* here refer to websites created by individuals, organizations, or institutions who presumably have little or no grounding in archival theory yet desire to make historical material accessible in digital form. Many of these online archives are defined by their ability *to archive*, rather than any specificity of their meaning as *an archive*. This paper contends, however, that an adherence to the concept of provenance, particularly in its postmodern manifestations, is nonetheless apparent in online archives: that despite their apparently free approach to content, context remains a unifying representational principle for online collections. The paper concludes with a discussion of how the fluidity and malleability of the digital interface both encapsulate and popularize the multifaceted concept of provenance discussed by archival theorists over the past decade, and how the proliferation of online archives will affect the professional practices of archivists.

This paper received the 2010 Theodore Calvin Pease Award at the Annual Meeting of the Society of American Archivists in Washington, D.C., on 13 August 2010. The award recognizes superior writing achievement by a student enrolled in an archival administration program. The paper was written for a master's class at the Faculty of Information, University of Toronto. The Pease Award Committee is chaired by the editor, Mary Jo Pugh. Members of the 2010 committee were Jennifer Johnson of Cargill Incorporated and vice chair of the Education Committee, and Jennifer Meehan, Yale University, appointed by the SAA president.

The last decade has witnessed a burgeoning scholarly and popular interest in archives. For academics and archivists alike, this trend began in earnest with Jacques Derrida's 1996 *Archive Fever*, a deconstructionist reading of the role of archives in cultural memorializing. In the subsequent "aftershock," Terry Cook writes, "Cultural theorists [began] subjecting the archival world to a detailed critique such as it has never before experienced."[1] For many people today, archives are arguably most visible online, where they are popularly understood as a gathering place for old and valuable material. Internet archives, as opposed to the online content of archival repositories, are created by institutions, scholarly communities, historical societies, and individuals for a number of reasons: they bring together large and dispersed collections of material; they offer thematic access to scholarly (and, often, nonscholarly) resources; and they promote certain ideas and viewpoints through the marshaling and visibility of documents, photographs, and other digital surrogates. Archives created online, moreover, represent an expansion of the boundaries of traditional archives by assembling related material in an accessible, easily searchable format. With online archives, however, comes the loss of the fundamental distinction (for archival theory) between a collection and an archives. While a collection gathers disparate material on the basis of interest or subject, archival documents gain their meaning from the preservation of original contexts of creation and use, as represented by the records' provenance. Digital archives, which tend to remove these links and associations in favor of thematic groupings and representative examples, seem to lack the provenancial bonds that archivists take as crucial to a record's meaning and evidential value. An increased attention to the postmodern qualities of malleability and multiple meanings, however, has led some to argue that the online archive in fact generates new contexts for textual, image, and media records.

This paper offers a preliminary exploration of archival representation and the digital or online archive. Building from a discussion on how archivists of the twenty-first century understand the multifaceted concept of provenance, it examines two online repositories of digital archival material: the First World War Poetry Digital Archive, hosted by Oxford, and the Walt Whitman Archive, edited by Ed Folsom and Kenneth M. Price and distributed by the Center for Digital Research in the Humanities at the University of Nebraska-Lincoln. Questions asked of these websites include the type of content being represented and by whom, and how such content elements are represented. In other words, what forms the basis of the content, and how are bonds and links established between related materials, if at all? The purposes of this investigation are to consider how representational practices take shape in online archives and to

[1] Terry Cook, "Fashionable Nonsense or Professional Rebirth?: Postmodernism and the Practice of Archives," *Archivaria* 51 (2001): 21.

question, specifically, whether these websites work against the archival principle of provenance or reinterpret it to create new, flexible contexts for archival material. The growing prevalence of online archives influences how users understand records' contexts and meanings. It is thus crucial that archivists consider the implications of this, both theoretically and practically, as repositories take their own holdings online. It should be noted, finally, that this essay does not address the issue of digitization itself. Representation here is taken to refer to contextual representation; the creation and significance of digital surrogates, items that are themselves representations of original documents, is a topic best discussed more fully elsewhere.[2]

Sites of Knowledge: Reading the Meaning of Archives

The concept of the archive as the storehouse of cultural memory arguably extends back to preliterate societies, where the history of a community or group was transmitted and renewed through oral traditions that communicated collective history. Accordingly, many of those exploring the concept today share the basic idea of the archive as a repository housing the collected artifacts of a culture, and, as Marlene Manoff notes, the term is frequently extended to the contents of museums, libraries, and the entire existing historical record itself.[3] Indeed, the term *archive* has enjoyed a marked surge of popularity in recent years as it has been appropriated by literary and cultural theorists, anthropologists, historians, Web and book designers, and others across academia and popular culture. Spurred by Derrida's deconstructionist examination of history and the *arkhe*, the place where things begin and where power originates, the *archive* as a concept and metaphor is now used to describe sites of knowledge and what Manoff calls the "thorny issues of…interdisciplinary knowledge production."[4] Karen Buckley, moreover, points to the ways that the archive inhabits genres across the spectrum of popular culture, from J. R. R. Tolkien's *The Lord of the Rings* and the historical fiction of Raymond Khoury and Elizabeth Kostova, to

[2] See, for example, Paul Conway, "Modes of Seeing: Digitized Photographic Archives and the Experienced User," *American Archivist* 73 (Fall/Winter 2010): 425–62.

[3] Marlene Manoff, "Theories of the Archive from across the Disciplines," *Portal: Libraries and the Academy* 4, no.1 (2004): 9.

[4] Manoff, "Theories of the Archive," 11. Terry Cook, in "The Archive(s) Is a Foreign Country: Historians, Archivists, and the Changing Archival Landscape," *Canadian Historical Review* 90, no. 3 (September 2009): 497–534, notes a divergence in the terms *archive* or *Archive* (singular) and *archives* (plural). While the latter draws focus to the "history of the documents over time, including the many interventions by archivists" and others, the former points more broadly to issues of power and memory centered on the initial inscription of the document. While often explored in mutually exclusive language, these concepts nonetheless dovetail in their reading of archives (or the Archive) as active sites of agency and power.

mainstream action films such as *Transformers* and *The Bourne Identity*.[5] In her inventory of pop culture representations of archives, Buckley notes four common themes:

1) protection of the record is equated with protection of the truth;
2) the archive is a closed space and the archival experience is an interior one for the characters;
3) records in an archives are "lost" and "buried," and characters must spend much time and effort "digging" to unearth them; and
4) the archival record invariably centers around the search for self or truth.[6]

The proliferation of online archives websites, however, counters certain of these themes by ostensibly making archival material visible to anyone and accessible in virtual form to everyone. The so-called buried records that the average Internet user may never otherwise have been aware of are revealed—and often with very little effort on the part of the user—by thematically organized and subject- or keyword-searchable websites. While the representations of archives that Buckley finds in pop culture may reflect ideas about the characteristics of physical repositories, archives websites increasingly offer the possibility of a new "archival space" for storing and exhibiting heritage material.

For the purpose of this paper, the terms *online archives* or *archive websites* do not refer to the online content of archival institutions; but rather to the websites created by scholarly communities, historical societies, and individuals who presumably have little to no grounding in archival theory but who share a desire to make certain collections of heritage or contemporary materials available. These efforts arguably reflect the popularity of the concept of *the archive* and are evidence of a continued interest in "the residue of history": a residue still widely connected to assumptions about the relationship between evidence and truth.[7] Historically, the authority of the archival record stems from the event it is taken to truthfully and impartially represent. Heather MacNeil traces the association of the record to notions of accountability and reliability back to the seventeenth century and the formulation of a new relationship between probability and evidence.[8] Postmodern destabilizations of widespread (although certainly not exclusive) concepts of truth and objectivity, however, have contributed to shattering the image of archives as bounded and stable sites of knowledge. As a

[5] Karen Buckley, "'The Truth Is in the Red Files': An Overview of Archives in Popular Culture," *Archivaria* 66 (2008): 95–123.

[6] Buckley, "'The Truth Is in the Red Files'," 97–98.

[7] Buckley, "'The Truth Is in the Red Files'," 104.

[8] Heather MacNeil, "Trusting Records in a Postmodern World," *Archivaria* 51 (2001): 36–47.

result, those archival principles once taken to guarantee the authenticity and reliability of the record—principles that include original order, *respect des fonds*, and the unbroken chain of provenancial custody—have come to be viewed by many within the profession as "historically contingent, not universal or absolute."[9] The loss of stability within the records themselves has arguably lead to a renewed interest in ways of representing records' contexts of creation and use to reveal complex and shifting meanings without abandoning a foundation from which such meanings can be gleaned.

The Unsettling of Provenance: Postmodernism and Records' Context

The concept of provenance, in archives, is defined most simply as the origin or source of the records; it is, according to the Society of American Archivists glossary, "the individual, family, or organization that created or received the items in a collection."[10] The principle of provenance thus dictates that records of different origins be kept separate to preserve their context. This concept is so central to modern archival arrangement and description that Tom Nesmith argues that the "intellectual history of the archival profession is the history of thinking about the nature of contextual knowledge about records."[11] Muller, Feith, and Fruin, in the 1898 *Manual for the Arrangement and Description of Archives*, make the rule of provenance "the foundation upon which everything must rest."[12] Archives "must be kept carefully separate" and not mixed with the archives of other creators or placed into artificial arrangements based on chronology, geography, or subject; the arrangement of such archives, furthermore, "must be based on the original organization of the archival collection, which in the main corresponds to the organization of the administrative body that produced it."[13] These rules, Cook notes, comprise the basic concepts of provenance and original order, or *respect des fonds*.

While Peter Horsman points out that the principle of provenance, prior to the writings of the Dutch trio, did not include how the group of records must be internally arranged, "only that it should be kept apart from other groups,"[14]

[9] Cook, "Fashionable Nonsense or Professional Rebirth?," 27.

[10] Richard Pearce-Moses, *A Glossary of Archival and Records Terminology* (Chicago: Society of American Archivists, 2005), http://www2.archivists.org/glossary, accessed 20 October 2010.

[11] Tom Nesmith, "Reopening Archives: Bringing New Contextualities into Archival Theory and Practice," *Archivaria* 60 (Fall 2005): 260.

[12] Quoted in Terry Cook, "What's Past Is Prologue: A History of Archival Ideas Since 1898, and the Future Paradigm Shift," *Archivaria* 43(1993): 21.

[13] Cook, "What's Past Is Prologue," 21.

[14] Peter Horsman, "The Last Dance of the Phoenix, or the De-discovery of the Archival Fonds," *Archivaria* 54 (Fall 2002): 1–23.

the concept of provenance itself draws a direct link between the creator of the records as the source of meaning and the records' reliability. As records are assumed to reflect events, the authenticity of the record thus relies on the claim of the recordkeeper to have preserved the original and uncorrupted memory of the event.[15] Once a direct line of custody from the creator to the archives was established, the principle of provenance was seen as a means of preserving evidence of the functional relationships between the records. Records that were created, maintained, and transmitted to the archives as an unbroken whole, separate from those of other creators, preserved their contextual information: the framework in which they were created and used. Wendy Duff and Verne Harris, moreover, suggest that provenance had also been assumed to provide an objective means of organizing archival material.[16] This stemmed from what they term the "Enlightenment origins" of archives, in which the "boundary between text and context is hard and stable," a record's context "is readily knowable," and the archivist's role "is to reveal [the record's] meaning and significance—not to participate in the construction of meanings."[17] Finally, provenance proved to be a practical method for dealing with the backlogs caused by the masses of material transferred to modern archives beginning in the early twentieth century, as it enabled archivists to work with aggregate groupings of records rather than describing individual items.

As Nesmith suggests, the apparent simplicity of these rules and the self-effacement of the archivist in representation is based in a longstanding Western assumption that records, as the means of communicating history, are the neutral mirrors of events as they were.[18] Increasingly, however, postmodern concepts have unsettled many of the ways that archivists understand the contexts in which records are created, maintained, transmitted, and preserved, as well as the ways in which these contexts can be represented in archival theory and practice. Principally, postmodern insights have brought about a wider view of what constitutes relevant context, principally by drawing in context's cultural and societal dimensions.[19] As Brien Brothman asserts, it is not individuals so much as social communities who are the source of the value and order in records, and thus social communities who are ultimately responsible for those records that are made, retained, and destroyed.[20] Cook refers to this as the

[15] MacNeil, "Trusting Records in a Postmodern World," 40.

[16] Wendy Duff and Verne Harris, "Stories and Names: Archival Description as Narrating Records and Constructing Meanings," *Archival Science* 2 (2002): 263–85.

[17] Duff and Harris, "Stories and Names," 264.

[18] Tom Nesmith, "Seeing Archives: Postmodernism and the Changing Intellectual Place of Archives," *American Archivist* 65 (Spring/Summer 2002): 24–41.

[19] Nesmith, "Reopening Archives," 259–74.

[20] Brien Brothman, "Orders of Value: Probing the Theoretical Terms of Archival Practice," *Archivaria* 32 (1991): 78–100.

"deeper contextual realities"[21] of archival records and archival practice in the twentieth and twenty-first centuries: each body of records is, effectively, a shared, constructed product consisting of the amassed "social and technical processes of the records' inscription, transmission, contextualization, and interpretation which account for its existence, characteristics, and continuing history."[22] By extension, archivists and archival institutions contribute to the provenance of the records through their acts of interpretation and inscription. Just as the reality that a creator records and the way that he or she records it are largely brought about by sociocultural factors, archivists, too, realize that they are not simply acquiring and describing records, but are creating value by ascribing a particular order to those materials.[23] Groupings of records, as Brothman writes, are "*microworlds*… demarcated by boundaries of our choosing" and, as such, threaten to "disguise as they conquer a profuse complexity."[24] Thus, while once viewed as natural, the representation and interpretation of records occur within a multiplicity of perspectives.

Postmodernism, while perhaps offering almost limitless possibilities for meaning, is nonetheless compatible with what Cook refers to as the "long-held archival focus on contextuality, on mapping the provenancial interrelationships between the creator and the text."[25] Indeed, as archivists come to understand records as the result of various processes and histories, recent archival literature points to the ways that this plurality of source and meaning, and the awareness of the interrelationships and multiple creators within fonds, do not necessarily unseat the concept of provenance itself so much as expand it. Archivists still depend on provenance as a source of organization, meaning, and elucidating increasingly pluralized and complex records sources. Accordingly, both Cook and Horsman advocate for a method of representation that illuminates multiple-creator relationships within what Horsman calls a "principle of (virtual) provenance"[26]: a provenance that cannot necessarily be captured physically but intellectually through the representation of records' multiple, many-to-many relationships. Laura Millar similarly argues that while provenance as an intellectual reality and the physical reality of records groupings are not equal, given the way that one body of records can stem from many creators, and one creator

[21] Cook, "Fashionable Nonsense or Professional Rebirth?," 26.

[22] Nesmith, "Still Fuzzy, But More Accurate: Some Thoughts on the 'Ghosts' of Archival Theory," *Archivaria 47* (Spring 1999): 145–46.

[23] Eric Ketelaar, "Tacit Narratives: The Meanings of Archives," *Archival Science* 1 (2001): 131–41.

[24] Brothman, "Orders of Value," 84.

[25] Cook, "Fashionable Nonsense or Professional Rebirth?," 28.

[26] Horsman, "The Last Dance of the Phoenix," 23.

can scatter records across many physical locations,[27] provenance can be more properly represented by descriptions that encompass not only biography, but creator history, records history, and custodial history. Each of these reveals different narratives within the history of the records as they came to be grouped within the fonds.

Thus, even within a pluralist, postmodern perspective, context is taken to be a source of truths—multiple, conflicting, and interwoven as they may be—about the records. While postmodern discourse, with its emphasis on "locality, ambivalence, contingency, multiplicity, and difference," often seems to preclude the assertion of a stable reality, archivists need to "adhere to some conception of truth in order to anchor the integrity of archival practice."[28] Similarly, Bernadine Dodge posits that while most are likely to agree that a reconstruction of an "authentic" past is unachievable,

> most of us still adhere to the notion that our professional practices lend themselves to the preservation of authentic fragments and textual artefacts which indicate something of real events, real deeds, real decisions, real administrative structures, real lives.[29]

Provenance, as the context(s) of records creation, arguably continues to provide the frame within which interpretations of the record and its focal event can be interpreted, discussed, and debated. As Geoffrey Yeo suggests, "[s]uch interrelationships are intrinsic to the *fonds* but will almost certainly be difficult to comprehend when its components are dispersed and undocumented."[30] The primary question, now, is how those interrelationships—how provenance itself—are represented in the host of online websites that purport to be the digital archives of heritage materials. If records are multiprovenancial in nature, do online archives offer a postmodern reinterpretation of archival representation based in the ever-changing contexts of society? Or does such technology, in the words of Lilly Koltun, "threaten to change 'how archives mean,' and how they relate to that concept of truth as definable, in whole and in parts"?[31] Margaret Hedstrom suggests that acts of contextualization, representation, and use of digital archives receive scant attention in archives or humanities literature in

[27] Laura Millar, "The Death of the Fonds and the Resurrection of Provenance: Archival Context in Space and Time," *Archivaria* 53 (Spring 2002): 5.

[28] MacNeil, "Trusting Records in a Postmodern World," 45–46.

[29] Bernadine Dodge, "Across the Great Divide: Archival Discourse and (Re)presentations of the Past in Late-Modern Society," *Archivaria* 53 (2002): 17.

[30] Geoffrey Yeo, "Where Lies the Fonds? Custodial History and the Description of Personal Records," *Third International Conference on the History of Records and Archives (I-CHORA) Conference Program and Participants' Papers* (Boston: Massachusetts Historical Society, 2007), 193.

[31] Lilly Koltun, "The Promise and Threat of Digital Options in an Archival Age," *Archivaria* 47 (Spring 1999): 115.

comparison to the creation, capture, or transfer of digital information.[32] In response, this paper turns to an examination of two Internet archives to begin to consider how archival and archived materials are represented online, and how these websites either work against or reinterpret the archival principle of provenance in the contexts they build for archival material.

A Review of Two Online Archives

The First World War Poetry Digital Archive

The First World War Poetry Digital Archive (FWWPDA) was launched on 11 November 2008, with the intention of making available a wide array of archival resources relating to the literature of the Great War.[33] The archive was one of twenty-two projects funded by the Joint Information Systems Committee (JISC) digitization program from April 2007 to March 2009. The JISC is an independent advisory board that supports the delivery of online content for higher education communities. The archive is hosted by Oxford University under the direction of Stuart Lee, a member of the English faculty at Oxford, the director of Oxford University Computing Services, and a reader in E-learning and Digital Libraries.

Self-described as an online repository with over 7,000 digital images of primary source material (including poems, letters, manuscripts, and diaries) and "contextual information" in the form of images, audio recordings, and film from the British Imperial War Museum, the archive's main focus is the work of the major poets of the period. In 2008–2009, the available collections were limited to those of Robert Graves, Vera Brittain, Wilfred Owen, Isaac Rosenberg, and Edward Thomas. Since that time, the FWWPDA has grown rapidly to include Siegfried Sassoon, Edmund Blunden, Ivor Gurney, David Jones, and Roland Leighton. The original materials were drawn from university and private collections in Britain and the British Imperial War Museum. Notably, the site also features a separate archive of over 6,500 items contributed remotely by the general public over a four-month period in 2008. This separate body of material was intended to facilitate the release of records stored by individuals that had been inaccessible to scholars and the general public. As stated on the website, these contributions "proved a powerful means of building the archive and making use of the Internet's ability to tap into amateur digitization and bring together unknown collections."

[32] Margaret Hedstrom, "Archives, Memory, and Interfaces with the Past," *Archival Science* 2 (2002): 21–43.

[33] See the website at http://www.oucs.ox.ac.uk/ww1lit/, accessed April 2010.

As a historical resource, the FWWPDA has selectively brought together materials from archival institutions and individual collections within a thematic context: that of the Great War generally, and its literature more specifically. From the homepage, a welcome note provides this contextual framework and offers several ways of accessing the collections. Each of these access points arguably provides a different representation of the digital materials, as each begins from a separate focus. The navigation bar down the left side of the page offers three different means of accessing the materials ("The Collections," "Education," and "The Great War Archive"), as well as links to outreach functions and self-guided reference aid ("News and Events," "About," "Permitted Use," "Book Store," and "Help"). The right side of the homepage mirrors the access points to the materials, but with larger, image-rich tabs and a one-line description of the materials content for each ("Browse the Collections," "Education," and "The Great War Archive"). The welcome note on the homepage also features hyperlinks to the collections, educational resources, items contributed by the general public ("The Great War Archive"), multimedia artifacts from the Imperial War Museum, and the archive's exhibition in the virtual world of SecondLife. There is also a search bar on the page, from which users can perform keyword searches in the FWWPDA or the donated collections of the "Great War Archive." This function seems to return mainly item-level hits, with no immediate links to related records or larger aggregations in which the materials have been grouped. A search for the term *gas mask*, for example, returned audio, visual, video, and textual material displayed with a thumbnail image, title, author/subject, item date, and content (material type).

In exploring the representational elements of this online archive, the poetry collections are the primary method of organizing the materials.[34] Clicking on a link to the collections, one arrives at a page that divides the materials into different contexts. First are the ten poets, represented by a hyperlinked image and name, and, below them, a section titled "War in Context" in which digital records have been arranged by format into a photographic collection, an audio collection, a film collection, and publications of war. Within each of these last four collections, the shared context of the materials is, again, mainly thematic, and each aims to supplement the archive as a whole by providing access to materials representative of the war experience. The photograph, audio, and film collections can be searched by keyword, with the returning hits organized at the item level according to item date, author/subject, or content, with a thumbnail image and content note (material type). The publications collection is organized either by the publication title (*The Hydra*, for example), subject (such as

[34] It should be noted that the "Education" page, while interesting and highly developed, is of less relevance to this study in that it has been constructed to provide educational resources and lesson plans to teachers and students. As something more akin to an archival outreach program, it is thus not expected to represent context in the ways this paper aims to explore.

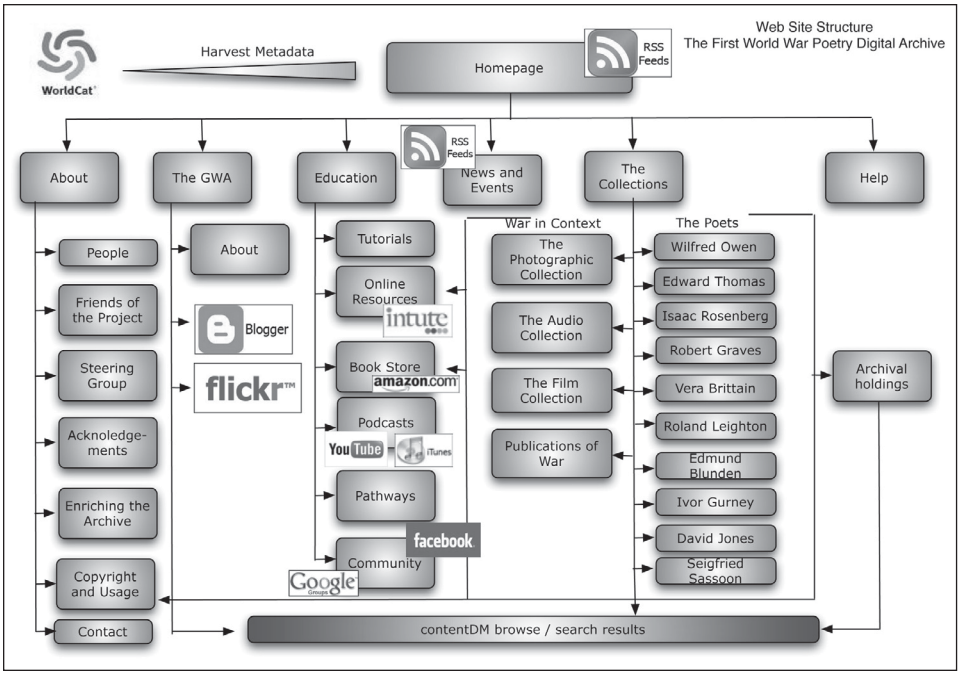

FIGURE 1. I FWWPDA Website Structure.

"U.K. Propaganda Posters"), and ephemera or material types. These items are, again, returned at the item level. It is interesting to note that while the FWWPDA itself does not immediately highlight the contextual linkages among items in its collection, the detailed information on each individual item returned includes the repository housing the original and the collection ID number at that repository. Clicking on this number will, in fact, result in a new results page that groups items from the same repository together. Thus, while building and foregrounding its own contexts, the online archive also preserves access to the archival/repository contexts from which the materials were drawn for those researchers who might be interested. To illustrate the intended structure of the website and links between the collections, the documentation section offers links to the workflows and technical specifications of the FWWPDA site, including a PDF website structure diagram that shows the collections' contents feeding into each other and back to the original repository holdings.[35]

The poetry collections in particular do the most to preserve their provenancial links and are grouped according to the poet that generated them. Using the Vera Brittain collection as an example, a small pane on the left side of the page maps the available commands for searching the collection. A biographical

[35] FWWPDA, "Documentation," http://www.oucs.ox.ac.uk/ww1lit/about/documentation, accessed April 2010.

sketch provides the context of the materials as pertaining to the wartime experiences of Brittain. Following this, centered on the page, is a keyword search bar and, below, a drop-down menu to specific poems by title or first line. These can be limited to images, texts, or images and texts together. Further down the page are links to the materials organized by format: correspondence from Vera Brittain; correspondence to Vera Brittain; photographs of Vera Brittain; and extracts from Vera Brittain's war diary. As in the other areas of the digital archive, the search results provide slender information, but clicking on the item itself offers a number of different contexts from which a record was taken and in which it can be placed. The digital copy of the poem "To My Brother," for example, offers the following representational elements: author, title, notes, item date, creation place, file type, item source, item medium, writing medium, content, collection ID, repository name, repository address, repository URL, cataloger, filename, copyright, digital repository, and reference URL.[36] From the "Vera Brittain Collection" page, finally, there is a link to a summary of the items digitized from physical holdings at McMaster University, as well as to other online resources for Vera Brittain and an extended bibliography.

While links do occur, moreover, between the different collections, these are not highlighted through the written descriptive elements so much as present in the arrangement of the materials. For example, letters between Vera Brittain and her fiancé, Roland Leighton, can be found in both collections, organized under "correspondence to" and "correspondence from." In this sense, the digital archive demonstrates its advantage in revealing the flexibility of provenance and the multiple contexts to which an item can be ascribed. A letter written by Leighton and received and preserved by Brittain certainly has a shared provenance, although in the traditional repository the letter is likely to be housed physically and ascribed provenancially to the Vera Brittain collection or fonds. Letters such as these, moreover, are created within the context of the Great War, which arguably becomes part of their social provenance—one of the factors, as Nesmith would argue, "of the records' inscription, transmission, contextualization, and continuing history."[37] The war, moreover, is a factor in what Brothman terms the socially generated value of the record. The FWWPDA acknowledges the widespread popularity of First World War poetry as a result of the considerable public attention the war continues to attract. The created social value of the war generates the value of the records themselves and, as such, creates yet another part of the context in which such letters are preserved and made available. The First World War Poetry Digital Archive thus shows the possibilities of acknowledging multiple contexts for archival materials, while

[36] See Vera Brittain, "To My Brother," FWWPDA, http://www.oucs.ox.ac.uk/ww1lit/collections/item/2840?CISOBOX=1&REC=2, accessed April 2010.

[37] Nesmith, "Reopening Archives," 262.

nonetheless preserving access to the provenance information, bonds, and links present in the physical repository.

The Walt Whitman Archive

The Walt Whitman Archive shares with the First World War Poetry Digital Archive a thematic context, although in this case one centered on the works of a sole creator. The Whitman Archive, which receives funding from the National Endowment for the Humanities and the U.S. Department of Education, among others, is managed and edited by Ed Folsom (the Carver Professor of English at the University of Iowa) and Kenneth M. Price (Hillegass University Professor of American Literature at the University of Nebraska-Lincoln). It is distributed by the Center for Digital Research in the Humanities at the University of Nebraska-Lincoln. As an electronic teaching and research tool, it brings together Whitman's vast body of work to make it "easily and conveniently accessible to scholars, students, and general readers."[38] In arranging their digital archive, Folsom and Price do not directly address issues of provenance or the bonds between records (assuming, perhaps, that the provenance of the material is obvious). Instead, the website emphasizes the textual difficulties of arranging Whitman's work—the enormous amount of manuscript material, six different edited editions of *Leaves of Grass*, and many notebooks, manuscript fragments, prose essays, letters, and articles that offer, as Folsom and Price suggest, the cultural and biographical contexts for Whitman's poetry.[39] While the main context of the materials is obviously Whitman himself, the Whitman Archive draws its digital collection from libraries and archives around the United States and the world.

The site's homepage offers a number of ways into the records, based mainly on format. "Published Works" includes links to *Leaves*, periodical printing, and foreign editions; "Manuscripts" leads to transcriptions and images, finding aids, and more; "Biography and Correspondence" includes life, letters, and chronology; "Criticism" contains reviews, commentary, disciples, and bibliography; and "Pictures and Sound" contains portraits of Whitman and audio material. In addition to these are links to "Resources" (including teaching material) and information "About the Archive." This menu demonstrates that the Whitman Archive represents itself more as a collection than an archive in any traditional sense. The link to "Criticism," for example, clearly does not contain materials originating with Whitman, but rather contemporary reviews, selected criticism,

[38] As stated on the website's "About the Archive" page, http://www.whitmanarchive.org/about/index.html, accessed April 2010.

[39] It is interesting to note the way this statement fundamentally assumes context to both include and extend beyond the creator of the records and his or her intentions for them.

a bibliography of published works about Whitman and others by and about those devotees who gathered around Whitman during his lifetime. Such works are strictly beyond the frame of provenance as it has traditionally been imagined and, arguably, as it is currently envisioned. It does, however, fit with Price's carefully elaborated re-imagining of the digital archive as both database and electronic scholarly edition. The concept of the database, Price asserts, is suggestive of the "re-configurable quality of our material" as well as the "simultaneously 'finished' and 'unfinished' qualities of the Whitman Archive itself." The archive, he continues,

> resembles a database in that its content is discrete computer files that function atomistically: as functional units within a computing system each item is just as important as every other item.[40]

The suggestion is thus that the Whitman Archive, as a digital archives, is both more open and less value-laden than the traditional repository, a belief demonstrated by the fact that the Whitman manuscripts themselves appear no more privileged on the website than the secondary source material.

The two links most likely to lead to digital copies of original materials are "Manuscripts" and "Biography and Correspondence." "Manuscripts" contains transcriptions and page images; a link to the Library of Congress's website on Whitman's notebooks; an "Integrated Guide" to Whitman's poetry manuscripts that describes each individually, notes their relationships to his published works, offers repository location information, and provides links to finding aids for manuscripts at the individual repositories from which the surrogates have been drawn. The website notes that finding aids "provide information about these particular manuscripts and briefly describe associated materials, such as correspondence, photographs, and other primary source materials." Within the link to the poetry manuscripts, the user can view transcriptions or digital images of the individual manuscripts. Clicking on "After certain disastrous campaigns," for example, opens a page with a transcription of the original manuscript that represents Whitman's revisions, as well as a digital image of the page.[41] The contextual information given is limited to the date of the original, an editorial note on the poem's publication, the original's location at the Morgan Library in New York, and an ID number for the image on the Whitman Archive.

Information on the manuscripts can also be found, however, through the "Integrated Guide to Walt Whitman's Poetry Manuscripts." This guide is much closer in format and content to what one would traditionally expect from an

[40] Kenneth M. Price, "Edition, Project, Database, Archive, Thematic Research Collection: What's in a Name?," Walt Whitman Archive, http://www.whitmanarchive.org/about/articles/anc.00346.html, accessed April 2010.

[41] See Walt Whitman, "After certain disastrous campaigns," Walt Whitman Archive, http://www.whitmanarchive.org/manuscripts/transcriptions/pml.00006.html, accessed April 2010.

archival or special collections finding aid. The guide includes only those original records created by Whitman and digitally housed in the Whitman Archive, and it was created through the work of the EAD Project Team at the University of Nebraska-Lincoln. A scope and content note states that the guide is arranged according to uniform title, and that it lists, item-by-item, all identified poetry manuscripts located in archival repositories throughout the United States and the United Kingdom. Item-level entries each contain the following representational elements: title, date, a link to images, content note, a link to information on the repository holding the original, and the identification number for the Whitman Archive. Multiple manuscripts of the same work are noted separately under each uniform title. The guide also provides subject indexing terms: "Whitman, Walt, 1819–1892"; "Whitman, Walt, 1819–1892—Manuscripts"; and "Poets, American—19th century". In addition, perhaps the most helpful feature of the website in terms of preserving provenance information is a page with links to the finding aids of individual archival institutions that have contributed to the Whitman Archive. These finding aids are arranged alphabetically according to the repository and provide an overview of the collection to which the individual manuscripts belong as well as the hierarchical levels of control. While these digitized finding aids are extremely useful to those wishing to pursue serious scholarly work, it would perhaps be more helpful if they were linked directly from the individual digital copies in the Whitman Archive so that the new context provided by the online archive could be aligned more readily with that of the repository, the repository's description of the item, and the links among original items.

Discussion: Digital Archives and the Re-vision of Archival Provenance

Overall, both online archives more closely match the definition of a collection—a group of materials determined by custodianship and drawn from a variety of sources—as opposed to an archival fonds or "grouping determined by [a common] context of creation."[42] Both the FWWPDA and the Whitman Archive, moreover, embrace the more popular concept of the archive as a warehouse of information and materials, rather than a repository with its own traditions of meaning and ordering. Only the Whitman Archive explicitly discusses the use of terminology in a linked article written by Price. Price asserts that, while in the past *an archive* has referred to a collection of unedited, unannotated material objects, in a digital environment *archive* "has gradually come to mean a purposeful collection of surrogates…something that blends

[42] Yeo, "Where Lies the Fonds?," 192.

features of editing and archiving."[43] It is worth noting a certain naïveté in Price's supposition that while traditional archives are "described in finding aids," this does not constitute an act of editing or annotation. As Cook notes, archives are not unproblematic storehouses of records awaiting the interpretation of historians,[44] but subjective spaces containing meaning at every level of inscription. However, the founders of the Whitman Archive, one might surmise, assume an implicit adherence to notions of archival neutrality that allows them to shape their materials into what they feel is a more meaningful form. Thus Price advocates the use of the term "digital thematic research collection" to more precisely delineate the nature of the online archive. Whereas, he writes, people ordinarily "assume that materials in a traditional print-based archive are unedited," thematic collections "are not static. Scholars add to and improve the content, and work on any given collection could continue over generations."[45] It is both interesting and troubling that, in discussing the use and nature of the term *archive*, Price makes no reference to its meaning within archival theory nor to any of the professional connotations and principles it might encompass. The archive is instead a blank slate, ready for a process of inscription that the Internet archives can provide.

The two online archives discussed do at least contain digital representations or surrogates of materials that would be commonly considered archival. In contrast, a great number of other Internet archives are strictly text based: for example, the H. P. Lovecraft Archive and the Marxist Internet Archive. Indeed, these archives function more appropriately as libraries or databases (the Marxist Internet Archive refers to itself, in the introductory text, as a library) by bringing together various transcriptions of the written work of their subjects and related secondary research articles. Other online archives also refer to themselves as archives in their title but as libraries in their content notes, blending the format of their materials without concern. The Internet Archive, for example, describes itself as a "non-profit digital library" with "150 billion archived web pages." What defines an archive online thus seems to depend on its ability *to archive*, rather than any specificity to its meaning as *an archives*.

It is thus important for archivists to ask what is meant, in these websites or online contexts, by the use of the term *archive*. As suggested in the writings of Manoff and Buckley, as well as by Derridian and humanities formulations, the term is ill defined, widely used, and—to many archivists—widely misused. While there is good reason to posit the need for a more universal understanding, we, as archivists, cannot ignore that *archive* and *archives* have acquired particular connotations within other disciplines and theoretical contexts, and the use of

[43] Price, "What's in a Name?"

[44] Cook, "The Archive(s) Is a Foreign Country."

[45] Price, "What's in a Name?"

the term on the Internet both points toward and builds new sets of meanings. Accordingly, it is vitally important that as archivists we not ignore or dismiss these meanings as simply incorrect or imprecise. We must instead seek to understand how different groups or individuals use the term in different contexts. As Cook suggests, the influence of postmodern and poststructural thought has provided the impetus for archival theory to emerge from the "cloisters"[46] of our own profession; indeed, we must engage not only with postmodern ideas but also postmodern realities. The pluralism and fluidity that we discuss as present in records are evident in the online archive and with this comes the potential unsettling of professional concepts that archives as a discipline has only recently established. And yet, as this paper asserts, a certain adherence to the concept of provenance, particularly in its "postmodern" manifestations, is found within Internet and online archives—that despite their apparently free approach to content, context can be seen as a unifying representational principle for online collections.

As both of these online archives suggest, context is primarily thematic and based not so much in the origins of the records as in the ways in which they can be linked to a central idea or person. As surrogate copies for archived material, the website, rather than the creator or the archival repository, provides the records' present context by bringing them together and making them available for a certain purpose. In no small way does the Internet itself form part of the records' context, as the primary benefit to such online archives is their ability to establish active connections between dispersed records and collections. Hedstrom speaks of the "interface" of the archive as "a tangible set of structures and tools that place archival documents in a context and provide an interpretive framework."[47] In computer-mediated archives, the interface—the website, the Internet—becomes a critical element in the interaction between records and researchers and the meanings and contexts that are created. Digital archives allow for juxtaposing and constant migration of records and texts that, for Koltun, mirror the postmodern condition: "eternal migration, the forever changing, never at rest, never at a final destination."[48] These changing contexts, from repository to thematic collection to multiple collections within a thematic framework, generate new meanings and add to those that have already been established.

Digital archives, with their malleability and thematic focus, are certainly more akin to collections. In referring to themselves as archives, however, they must arguably also reflect certain elements of provenance by maintaining and demonstrating some link to those social factors, functions, institutions, or

[46] Cook, "What's Past Is Prologue," 35.

[47] Hedstrom, "Archives, Memory, and Interfaces with the Past," 22.

[48] Koltun, "The Promise and Threat of Digital Options in an Archival Age," 124.

individuals that constitute their origins. Crucial to this inquiry, then, is whether Internet archives work against the guiding principles of traditional archives. While those digital archives created by individuals and organizations unaffiliated with physical archival repositories cannot necessarily be expected to consciously engage with the concerns of our profession, they nonetheless demonstrate many of the facets of malleable context and meaning that archival theorists have discussed over the past decade. Increasingly, archivists argue that while records' meanings may be context based, context itself is boundless. Nesmith, for one, questions how archivists can formulate certain knowledge of the provenance or origin of records: "We must act on some view of origins many times a day. We must begin somewhere, but where do we even start? What *is* the beginning?"[49] Far from discovering or representing one authoritative origin for a group of records, archivists must make choices about what elements of provenance are the most meaningful. Naturally, such choices change how records are contextualized in the present. Thus, Nesmith continues, while provenance itself "does not change," concepts of provenance "evolve[s] towards greater complexity and variety rather than consisting of simple shifts of custodial responsibility."[50]

The fluidity and changeability of the digital interface seem in many ways suited to demonstrate the archival profession's changing views on representing provenance. Elizabeth Yakel contends that archival representation is "a fluid, evolving, and socially constructed practice."[51] As collections are used, their understood meanings and significance change over time, and archivists must thus think less of a single means of representation—a single definition of provenance for records—and more of "continuous [and] relative…on-going representational practices."[52] George Bornstein suggests that the presence of any one text serves as a reminder of any number of other texts that could also be present;[53] in the same way, an expanded conception of provenance reminds us that archival collections are never complete in themselves, but always point to other, related records that form part of a larger context. The cross-linking that occurs in online archives, therefore, can bring together that which had otherwise been scattered across different libraries and archives, and thus both restore and establish contextual bonds that would have remained hidden. MacNeil argues that our understanding of how records functioned in their original context demonstrates precisely how they no longer fulfill those same functions.[54]

[49] Nesmith, "Still Fuzzy, But More Accurate," 140.

[50] Nesmith, "Still Fuzzy, But More Accurate," 147.

[51] Elizabeth Yakel, "Archival Representation," *Archival Science* 3, no. 1 (2003): 2.

[52] Yakel, "Archival Representation," 4.

[53] George Bornstein, "How to Read a Page: Modernism and Material Textuality," *Studies in the Literary Imagination* 32, no. 1 (1999): 29–58.

[54] Heather MacNeil, "Archivalterity: Rethinking Original Order," *Archivaria* 66 (Fall 2008): 1–24.

Accordingly, as records take on new meanings and new contexts, understandings of provenance can shift to encompass not only the original contexts of creation (which should be preserved), but also those new contexts to which records come to belong.

The greater challenge to re-envisioning provenance in online archives may well have less to do with accepting and representing multiple contexts of creation and use than with negotiating what these mean for related concepts dependent on provenancial stability. As previously suggested, archives are often characterized within the profession in terms of their relationship to evidence and memory. Given that this relationship has been more than appropriately problematized in recent archival literature, it seems most useful to question the implications for this relationship when looking at nontraditional archives such as those found online. Issues that need to be further considered and addressed include, in particular, how online or Internet archives can assert their reliability as sources of history and memory. Is this to be achieved through the size and variety of the collection given, and the proportion of records that are scanned originals rather than transcriptions? Is the reliability of an online archive as an information resource based on the perceived reliability of the hosting body or source community? Or perhaps it is based on the transparency with which online records (or representations) are linked back to their original, physical repositories and the degree to which different possibilities in arrangement and description are made evident? Such issues will certainly come to the fore as increasing numbers of researchers—academic and amateur—go online for their material and institutions take advantage of the Internet as a platform for their collections and scholarly projects.

Conclusion

In sum, the growth of online archives will almost certainly have an impact on the ways that researchers understand and use archival repositories, as well as on the ways repositories manage their own records as they take their holdings online. Cook suggests that popular understandings of archives—including their functions and arrangements—are likely to influence "general public opinion, shape the outlook of new archivists coming to us, and transform researchers' and sponsors' expectations."[55] For Koltun, however, this does not mean that digital archives and their representations of context will supplant those of traditional archives; rather, "each medium adds to the collective body of expressive communication in unique ways, rarely entirely replaced in power or scope by a

[55] Cook, "Fashionable Nonsense or Professional Rebirth?," 21.

subsequent medium."[56] Given that the Web facilitates thematic and subject searching while building collections and context around historic persons and events, archivists may do well to seize the opportunities presented by the Internet to provide richer, more detailed, and multinarrative contextual information. To do so, archivists must deconstruct those contexts they are trying to describe and, as Nesmith has suggested, think about familiar concepts in different and complex ways.[57] Many websites are not unlike the thematic guides that archives produce as another means of accessing records; as such, online archives are less an alien means of representation than a familiar adaptation of ongoing practices and concerns. Thus, while provenance provides an interpretive grounding in the creator or creators, the multiple contexts that influenced the records' creation, transmission, and reception will continue to manifest in the "fluidity, flexibility, and ultimately uncontrollable nature"[58] of records' contexts.

[56] Koltun, "The Promise and Threat of Digital Options in an Archival Age," 117.

[57] Nesmith, "Still Fuzzy, But More Accurate," 142.

[58] Cook, "Fashionable Nonsense or Professional Rebirth?," 32.