Toward a Notion of the Archive of the Future: Impressions of Practice by Librarians, Archivists, and Digital Humanities Scholars

Tanya Clement, Wendy Hagenmaier, and Jennie Levine Knies

ABSTRACT

With this piece, we seek to interrogate the sites at which library, archival, and scholarly work occurs in order to consider the changing nature of the future of the archive. First, we consider the work of the archive from the perspective of the long-standing tradition of scholarly publication and scholarly editing in archives and libraries. Second, we introduce interviews with five leading humanities scholars and practitioners, who discuss the work that is involved in producing scholarship in archives and libraries. Finally, we explore the topics and themes that surface from the interviews, including centralized digital repositories, open-source methods and applications, and community building. This conclusion gives insight into theories and modes of practice that are developing and shaping notions of the archive as a site of collaborative work among archivists, librarians, and humanists, who are constantly negotiating their shifting roles in the stewardship of the archives of the future.

While the term "archive" has always been slippery, it seems that the most recent debates concerning its meaning hinge on the intersection of archival work, changing digital technologies, and evolving scholarly practices and needs. In a recent blog post entitled "The Problem with the Scholar as 'Archivist,' or Is There a Problem?" Kate Theimer of *ArchivesNext* bemoans what she sees as the blurring of terms such as "library," "archive," and "collection," arguing that "just as you are curating your snack collection when you pull those Doritos off the supermarket shelf, any collection or assemblage of copies of original materials gets called 'an archive'" (2012). On the other hand, Marlene Manoff, who is concerned with how the term "archive" functions across archives, libraries, and museums, notes that "as libraries, museums, and archives increasingly make their materials available online in formats that include sound, images, and multimedia, as well as text, it no longer makes sense to distinguish them on the basis of the objects they collect" (2004, 10). As

well, Kenneth Price, who is codirector of the Walt Whitman Archive, argues that digital scholarly projects such as the Whitman Archive have "come to suggest something that blends features of editing and archiving. To meld features of both—to have the care of treatment and annotation of an edition and the inclusiveness of an archive—is one of the tendencies of recent work in electronic editing" (Price 2009). Finally, in an offhanded remark made while writing about the distinction between archives and databases in scholarly thematic research collections, Jerome McGann begins simply: "Libraries and museums—let's call them archives" (2007, 1590).

Whether or not collecting or archiving an analog original, a digital surrogate, or a born-digital object in a library, museum, or archive (all the while noting that all of these terms are slippery at best) is most or least appropriately the work of any one individual in a given institutional context is beyond the scope of this discussion. As well, we are not asserting that libraries, museums, and archives are the same thing; rather, with a formal qualitative study, we seek to interrogate the sites at which the perceived blurring of library, archival, and scholarly work occurs in order to consider the changing nature of what Jacques Derrida has called an "impression" of the archive. In *Archive Fever*, Derrida writes that the "archive" is "only a notion, an impression with a word and for which, together with Freud, we do not have a concept" (1998, 29). Yet, Derrida considers this "notion" or "series of impressions associated with a word" to be "the possibility and the very future of the concept, to be the very concept of the future" (29). He notes: "It is the future that is at issue here, and the archive as an irreducible experience of the future" (68). As such, with a look toward the future of the archive, we consider in this discussion how emerging digital practices in scholarly publication and scholarly editing in libraries and archives shape notions of the changing roles of the archivist, the librarian, and the humanist in the digital age—and thus the changing work of the archive.

First, we consider the work of the archive from the perspective of the long-standing tradition of scholarly publication and scholarly editing in archives and libraries. Second, we introduce interviews with five leading humanities scholars and practitioners, who give their impressions of the work that is involved in producing scholarship in archives and libraries. Finally, we explore the topics and themes that surface from the interviews, including centralized digital repositories, open-source methods and applications, and community building. This conclusion gives insight into theories and modes of practice that are developing and will continue to shape notions of the archive as a site of collaborative work among archivists, librarians, and humanists, who are constantly negotiating their shifting roles in the stewardship of the archives of the future.

Background

Scholarly work in the archives has a long tradition. When the National Historical Publications and Records Commission began awarding grants to US archives, special collections li-

braries, and historical societies for the publication of annotated editions of archival documents in 1964, the tradition of publishing records and manuscripts in scholarly editions and guides that would attract researchers to institutions had already been thriving for decades in the United States and more than a century in Europe (Jørgensen 1983, 401). European models of historical scholarship privileging primary sources began to influence American academia in the nineteenth century. As a result, the American archival tradition took root as it became crucial to preserve the documentation of cultural heritage and creativity. In order to publicize the richness of their holdings to potential scholars, archives and special collections libraries began encouraging archivists, librarians, and scholars to publish on selections from their stacks.

Beyond providing publicity for various collections, these publications encouraged wider access and helped garner support for the preservation of vulnerable, often fragile, original materials (Cox 1969, 25). In his 1962 report on the "Publication Program of the Philadelphia Archives" in *American Archivist*, Allen Weinberg, the assistant city archivist, emphasizes that "making certain historical studies and publishing some of them were necessary to acquaint the users of the archives with our records" (1962, 193–94). Weinberg, noting that many archivists entered the profession with strong training in history and research, further argues that "an archival installation, even with a very small staff, can produce publications, based on sound research, to aid in the use of and stimulate interest in the collections" (1962, 197). In her history of the origins of the Mississippi Department of Archives and History, Patricia Galloway suggests that the publication program established by archivist Dunbar Rowland, the department's first leader, mirrored other similar endeavors across the nation: "Rowland partook of the trend of his time to edit and publish historical materials to make them available to the educated public, and his annual reports soon became venues for such publications" (2006, 100). Galloway implies that Rowland's position as archivist was particularly advantageous, since as "his aim was to multiply copies of documents as a preservation and access strategy, he obtained copies of Mississippi records where original records could not be secured" (2006, 100).

Certainly, it is now uncontroversial to suggest that, like editors of scholarly editions, archivists shape our interactions with the archive. Derrida reminds us: "The technical structure of the *archiving* archive also determines the structure of the *archivable* content even in its very coming into existence and in its relationship to the future. The archivization produces as much as it records the event" (Derrida 1998, 17). Thus, the archivist's finding aid is a type of expert annotation, and the processes of appraisal, deaccessioning, and arrangement impact the decisions about selection and style that establish the archive. Collecting, arranging, and curating "thematic research collections" and "digital scholarly editions" means that the role of the scholar is likewise becoming increasingly enmeshed with the activities traditionally assigned to the archivist (Palmer 2004, 348). In his book on editorial theory, *The*

Fluid Text, John Bryant opines that because an electronic edition can include so many available resources in the digital age, "the editor-narrator must also become an editor-narrator-librarian of the fluid text 'reading room' wherein all full texts of all versions of a work are stored" (2002, 161). This hybrid librarian must "channel electronic features toward empowering readers to become more familiar with and adept at textual scholarship's techniques . . . to perform the spadework of textual criticism of their own" (161). While the archivist has always occupied a multifaceted role as scholar, editor, publisher, steward, and collaborator, archival work in the digital age has further blurred these roles.

The site of much of this hybrid work, which is at the core of observations made by Bryant, Manoff, McGann, and Theimer, is the scholarly edition. According to the *Oxford English Dictionary*, the term "edition" is "one of the differing forms in which a literary work (or a collection of works) is published, either by the author himself, or by subsequent editors." Historically, however, it has been used more generally to refer to the "action of putting forth, or making public; publication."[1] Through the definitions and examples offered by the Modern Language Association (MLA) and the Association of Documentary Editing (ADE), the term "scholarly edition" originally came to denote a physical, printed text, accompanied by expert annotations. Price puts forth the following argument for what constitutes a first-rate scholarly edition: "Successful scholarly editions yield a text established on explicitly stated principles by a person or a group with specialized knowledge about textual scholarship and the writer or writers involved. What makes the edition scholarly, of course, is the rigor with which the text is reproduced or altered and the expertise deployed in the offering of suitable introductions, notes, and textual apparatus" (Price 2009, 3). The MLA's "Guidelines for Editors of Scholarly Editions" echo these principles: "The scholarly edition's basic task is to present a reliable text: scholarly editions make clear what they promise and keep their promises," ensuring reliability through "accuracy, adequacy, appropriateness, consistency," and "explicitness" (2011). Digital humanities scholar Elena Pierazzo offers a definition of a related form, the "diplomatic edition," which "comprises a transcription that reproduces as many characteristics of the transcribed document (the diploma) as allowed by the characters used in modern print . . . a sort of surrogate of a facsimile edition" (2011, 463–64).

Despite disagreement within the digital humanities field about exactly what constitutes a *digital* scholarly edition, and even whether that is an appropriate term, there is an overwhelming consensus that scholarly editors, while pushing the boundaries of knowledge production in digital publications, should also maintain the painstaking editorial standards and the overriding principle of reliability expected of print scholarly editions.[2] Theorizing on

1. *Oxford English Dictionary*, 2nd ed., s.v. "edition."
2. Pierazzo suggests the term "digital diplomatic edition" (2011, 468). In his article "Digital Editions: Scholarly Tradition in an Avant-Garde Medium," in *Documentary Editing*, Andrew Jewell offers the term digital "thematic research

the future of digital scholarly editions in their 1996 "Prospectus for Electronic Historical Editions," the ADE mandated that electronic editions should "maintain current standards of scholarly editorial excellence," "facilitate changes in scholarly editorial practice," allow for "post-publication enhancements of editions and multiple forms of publication," and "conform to relevant standards for electronic text, images, and other material" (ADE 1996). Around the same time, however, Jerome McGann (1995) argues that "we no longer have to use books to analyze and study other books or texts," and he claims that "electronic tools in literary studies don't simply provide a new point of view on the materials, they lift one's general level of attention to a higher order." As a result of theories like McGann's, enthusiasm for the digital über-edition in literary studies peaked and was codified into recommendations for practice in the 1990s. Indeed, the attempt (in many cases) to implement some of the specifications set forth in guidelines, such as Peter Shillingsburg's essay, "Principles for Electronic Archives, Scholarly Editions, and Tutorials" (originally published in 1993),[3] and Charles Faulhaber's "Guidelines for Electronic Scholarly Editions" (Faulhaber 1997),[4] were far less practical than those espoused by the ADE and encouraged impractical goals and lofty hopes for digital, editorial miracles. At the same time, frustrated editors produced some surprising results, such as the Dickinson Electronic Archives, the William Blake Archive, and the Walt Whitman Archive, which helped to situate the digital scholarly edition as a venue for an expanded notion of what comprises a textual "event," all the while reminding us that print-based and computer-augmented processes of producing knowledge are deeply symbiotic (Smith 2002).

To put it plainly, the digital environment is not the only appropriate venue for displaying the multifarious archives of Blake, Dickinson, or Whitman, but as a medium with different restrictions and different possibilities than the print environment, it does encourage editors to reevaluate their editorial expectations about the method by which they put into practice their editorial theories and how those changing methods may change prevailing impressions of these archives. Further, it is that putting into practice that continues to keep in close synch the work of archivists, librarians, and scholars in the digital age. In the next section, we include a discussion of interviews with five scholar practitioners in libraries and

collection" (2008–9, 31). In his *Digital Humanities Quarterly* article "Edition, Project, Database, Archive, Thematic Research Collection: What's in a Name?" Kenneth Price recommends the term "arsenal" (2009).

3. These guidelines were originally distributed at the Modern Language Association conference in Toronto in December 1993 and became the basis for Charles Faulhaber's MLA "Guidelines for Electronic Scholarly Editions" (1997).

4. The academic community used these 1997 guidelines until the latest revision in 2005, which now incorporates guidelines for both print and electronic editions in the combined "Guidelines for Editors of Scholarly Editions" (MLA 2011). For a full account of how the MLA Committee for Scholarly Editions produced the current guidelines, see "Report from an Editors' Review," at http://www.iath.virginia.edu/~jmu2m/cse/Editors.rpt.htm, and the presentation to the Society for Textual Scholarship, in 2001, at http://www3.isrl.uiuc.edu/~unsworth/sts2001.html. More information on this study and its relationship to the creation and implementation of the Dickinson Electronic Archives can be found in Tanya Clement's essay, "A Digital Regiving: Editing the Sweetest Messages in the Dickinson Electronic Archive" (2008).

archives—each of whom incorporates humanities scholarship, archival practice, librarianship, and editorial expertise into his or her work. Considering their impressions of how scholarship is an essential part of their work with digital archives and libraries helps shape our expectations concerning areas of significant development in the future of the archives.

Interviews

In this section, we discuss interviews with five archivists, librarians, and humanities scholars who offered insight on their own work with digital collections. The participants were chosen to reflect a range of perspectives, from those who were trained as literary scholars to those who were trained as archivists and librarians. Because the work on which we are focused in this study is scholarly, literary archives and editions, literary scholarship is the prevailing humanities background of our interviewees, most of whom hold positions within academic libraries and collections. All five of the participants are involved in what Doug Reside has called the work of the "hybrid scholar" (2011).[5] During the hour-long telephone interviews, we asked the interviewees to discuss their definitions of digital editions, the advantages and disadvantages of creating projects in the digital realm, the challenge of defining manageable scopes for digital projects and integrating the resulting data into institutions' larger data sets, best practices for the role of the host institution in the creation and maintenance of digital projects, and predictions for the future of digital editions. The interviewees reflected on the ways emerging digital practices are shaping how archivists and humanists consider the work of archiving, the nature of archives, and the role of the archivist and the scholar in the digital age. As we have seen, scholarly articles and scholarly editions are a long-standing tradition in libraries and archives. Examining this tradition, both then and now, gives insight into theories and modes of practice that may best facilitate how the librarian and archivist of the future might approach the stewardship of digital humanities data.

Andrew Jewell, associate professor of digital projects at the University of Nebraska–Lincoln Libraries and editor of the Willa Cather Archive, identifies several major advantages of the digital sphere for scholarly editions (which he often prefers to call "digital thematic collections"): the opportunity for smaller, niche projects to secure publication (editions that might not otherwise make it into print), the possibility of creating more expansive projects that contain unprecedented quantities of material (e.g., creating a digital scholarly edition incorporating every single one of Cather's letters), and the power to reach new audiences by providing free online access. He cautions against creating "new for new's sake," and he emphasizes that digital scholarly edition creators should clearly delineate costs and goals in their decision-making processes.[6]

---

5. Reside specifically discusses the scholar-programmer, but an analogy could easily be drawn to the scholar-librarian or scholar-archivist.

6. Andrew Jewell, interview with the authors, January 30, 2012.

Trained as a literary scholar, yet working within a library, Jewell considers fellow scholars to be the primary audience for his digital editions. He believes that his interdisciplinary role within the Center for Digital Research in the Humanities offers him the chance to utilize his scholarly intuition and his understanding of metadata concerns at the same time that he can apply his practical experience as a "builder" of digital humanities projects. Jewell thus unites the skills of the researcher, the librarian, and the digital humanist within one position: in the domain of the digital humanities, his renaissance identity is not at all unusual. According to Jewell, the responsibility of the library or archive in the creation of digital scholarly editions should not only be to provide long-term preservation and stewardship of data; the library/archive should also act as a content creator, collaborating with scholars to build digital publications and encouraging peer review by external groups, such as Nineteenth-Century Scholarship Online (NINES). He envisions the model of scholarly editions progressing until digital publications are viewed as clusters of data that can be remixed and reinterpreted through multiple interfaces and presentation formats. Experimentation with interfaces, he argues, should be a top priority.

Like Jewell, Amy Earhart, assistant professor of English at Texas A&M University, espouses the scholar's view of the creation of digital editions.[7] Unlike Jewell, however, Earhart works under the academic umbrella of a traditional English department. Her involvement with the Nineteenth-Century Concord Digital Archive (for which she serves as director) takes place largely outside the boundaries of the print publication work she is expected to accomplish within the tenure system. Earhart prefers the term "cultural collection" over the phrase "digital scholarly edition" for her own digital publication projects, and she sees data mining, visualization, and other innovative tool-based projects as the greatest advantages in doing scholarship with digital collections.

Emphasizing the fact that most modern literary scholars lack several skills that are crucial in digital humanities work—a background in textual theory and analysis, an understanding of interdisciplinary collaboration and communication strategies, and a facility for managing and educating teams—Earheart laments the dearth of large, robust data sets in the digital humanities, the raw materials that would enable tool-based projects to thrive. Scholar-creators of digital projects, Earhart argues, are required to do "double duty" to meet traditional academic expectations and to master "the new and the different." Like Jewell, she cautions against novelty for novelty's sake, and she voices concern over grant funders' emphases on innovation at the expense of perhaps less original but higher-quality detailed work: the "new and exciting and hot stuff isn't making a new digital edition. We need to experiment like crazy, for sure," she argues, "but not to overemphasize that experimentation."

7. Amy Earhart, interview with the authors, February 10, 2012.

With regard to the role of the library and archive in the creation of digital projects, Earhart sees librarians taking a greater role in traditionally scholarly activities such as annotation. "Scholars forget that librarians are scholars too," she emphasizes, highlighting the Scholars' Lab at the University of Virginia as the ideal model of an institution that fuses research, preservation, and training under one roof. Earhart hopes to see increased support for scholars at smaller institutions, for connections and collaborations between related digital publications, and for preservation plans for orphaned projects that are no longer being updated. She also hopes to witness a continued blurring of the lines separating the scholar, the librarian/archivist, the editor, and the information technology (IT) expert. Earhart and Jewell published their coedited anthology, *The American Literature Scholar in the Digital Age*, with the University of Michigan Press, which recently became a division of University Libraries. Earhart draws inspiration from this new model (Earhart and Jewell 2011). She applauds the publisher's open-access framework and its hybrid library-editorial identity.

Representing the librarianship perspective on the creation of digital editions, Dorothy Porter, associate director for digital library content and services at Indiana University,[8] echoes Jewell and Earhart's enthusiasm for the digital—particularly the opportunity to pair TEI-encoded text (Text Encoding Initiative; http://tei-c.org) with extremely high-resolution images of archival materials and the open-access nature of so many digital publications (e.g., Dekhtyar et al. 2006). More than anything else, she celebrates the editorial transparency the digital realm enables. Porter has significant experience with the scholar's temptation to "encode it all!" as she phrases it—to think only of what is imaginable and ignore what is feasible. She encourages several tactics to combat overambitious projects, including drafting of TEI guidelines before the project begins, participating in centralized repository efforts (particularly for scholars at smaller institutions that lack digital infrastructure), specifying the level of TEI-encoding the project will achieve, and attempting to accomplish the project's primary goal with the least amount of encoding. Porter stresses that libraries should support both "big data" and "small data" (or close-reading) projects. They should work with scholars to encode project data according to recognized standards and ensure that the base files are available over the long term. Accordingly, project teams should realize that presentation mechanisms will change over time. Furthermore, Porter would like to see librarians and archivists encouraging scholars to make use of peer review infrastructure for digital editions.

Like Earhart, Porter applauds the model of library publishing established at the University of Michigan, despite the fact that such merging of worlds can sometimes incite conflict over terminology: for example, when the textbook publishing team at Indiana University began referring to their electronic editions as "e-texts," they failed to realize that Porter's digital libraries team had already claimed the term for their own digital texts. Porter's col-

8. Dorothy Porter, interview with the authors, January 30, 2012.

league asked the TEI listserv to help her brainstorm a new term, and one respondent suggested, simply, "book." "As the vast majority of books, even though they may end up being printed on paper, are created using digital means," the respondent argued: "can't we now call all self-contained or even hyperlinked text objects 'texts' or 'books' and simply step over the cute prefixes? Or, if we must have a word that describes the form the text takes, like scroll or codex, can we just create something entirely new that doesn't contain an 'i' or 'e' or 'digi' or other reference to the technology currently in vogue. . . . Let's . . . make paper-based practitioners add their own modifiers to the name of their paper objects."[9] This very blurring or redefinition of publication terminology seems to echo the increasing overlap between the roles of the various creators of digital editions. It seems the rise of the digital humanities has inspired practitioners to redefine their formats as well as their professional identities.

Gretchen Gueguen is the digital archivist for digital curation services at the University of Virginia Libraries, and like Porter, she believes in the power of infrastructure—from metadata to project management.[10] As the digital archivist for the Mellon-funded initiative Born Digital Collections: An Inter-Institutional Model for Stewardship, Gueguen is at the cutting edge of digital content stewardship and preservation.[11] She has the insight that only diverse experience in the field can bring. She has been involved in the creation of digital projects from the digital humanities, library, and archives perspectives at several universities. Despite the fact that she has been able to cross perceived lines between these perspectives, Gueguen sees a lack of communication between digital humanities centers and libraries and archives. This is sometimes due to the interests of scholars not being well aligned with the collections in the library. But it can also be based on a lack of awareness between the two groups.

Gueguen understands the prime importance of storing digital data in trusted archival repositories, and she laments the fact that the digital components of many digital humanities projects never make it into the repository. She offers several suggestions based on her experience. First, digital project managers should take a lesson from IT and divide projects into achievable phases with manageable scopes. And, second, because university budgets are inevitably tight, digital stewards should spread their resources across projects and avoid investing too heavily in specific endeavors. For example, curators of a large collection of an author's e-mails would do well to describe the e-mails at the aggregate, series level by performing "more product, less process" methods and thereby making the collection acces-

9. Hope Greenberg, e-mail to the TEI listserv, January 19, 2012. http://listserv.brown.edu/archives/cgi-bin/wa?A2=ind1201&L=TEI-L&P=R9291.
10. Gretchen Gueguen, interview with the authors, February 27, 2012.
11. University of Virginia Library, "AIMS Born-Digital Collections: An Inter-Institutional Model for Stewardship," http://www2.lib.virginia.edu/aims/.

sible to researchers as soon as possible, rather than to encode each individual letter in TEI and further delay researchers' discoveries (Greene and Meissner 2005). Such an activity would provide a baseline level of access that would not impede future scholars or archivists from doing further work such as encoding or other types of analysis.

Most of all, the individual components and data behind digital projects should be stored in a shared centralized repository. This kind of synergy could facilitate the development of links between different projects using the same resource. So, in the future, if a visitor to the university digital repository encounters an image of a letter that is accompanied by a link to the digital project featuring that image, she can then also discover the project. If possible, Gueguen suggests, digital material could even be hosted and shared between the digital repository and the project from the same servers through a single work flow. Looking ahead to born-digital projects, Gueguen recommends the idea (inspired by the British Library's Digital Lives Research Project [http://www.bl.uk/digital-lives/] and the PARADIGM Project [http://www.paradigm.ac.uk/]) of "enhanced curation," in which the archivist begins to collect material from the creator while the creator is still living, enriching accessions of digital material with accessions of information acquired through interviews—information about the creator's digital habits, computer usage, software preferences, and so forth. In other words, these details about the creator as a curator of his or her own digital material becomes classified as archival. This information can be crucial for being able to recreate an environment in which the original digital material can be viewed again.

As the first ever digital curator of the performing arts at the New York Public Library for the Performing Arts, Doug Reside has a PhD in English, a programmer's vision, and a groundbreaking perspective on digital curation.[12] He asserts that the most significant power afforded to edition creators by the digital is the power to offer a sustained, rigorous scholarly argument on a text while also enabling the reader to make choices and to customize the content based on his or her own needs. According to Reside, creators of digital editions need to prepare projects according to widely accepted standards, to divorce raw data from changeable interface trends, and to centralize data in collaborative, preservation-friendly repositories. As a result of working with performing arts materials, he knows all too well that edition makers must grapple with the limitations of copyright and provide as much access as possible. He believes creative tools for digital exhibition and sharing (blogging, social media, mini-editions) can help with this. The increasing predominance of born-digital materials in libraries and archives will intensify the challenges we already face with digitized content, while also inviting further innovation.

Reside is a TEI skeptic, arguing that the initiative represents a remnant of a preweb world. He suggests that edition creators should instead transfer their efforts to advances in recog-

12. Doug Reside, interview with the authors, February 24, 2012.

nition algorithms and automated processing. When humans encode content, they provide context for and insight into content that machines cannot yet mimic. According to Reside, however, the semantic web and linked data offer the possibility of constructing that context artificially. If we can assemble vast enough data sets, he posits, we can mine them and train technological systems to act as our efficient partners in data curation. In conjunction with this movement toward automated techniques, Reside recommends that edition creators should "make the introduction of new data into a data ecosystem easier by adopting standards and practices that are bigger than the TEI world," bigger than the library or digital humanities worlds—standards like HTML5 and HTML+RDF that facilitate semantic markup. A project can reach a broad audience and maintain relevance if it speaks the language of that audience: the language of the web. Scholars from scientific and technical fields have long been grappling with the challenge of immense data sets. Now that humanities data also exist in bits and bytes, digital project creators are coming to understand the vocabulary of data curation. Increased frequency of interdisciplinary conversation will ensure the cross-pollination of emerging data stewardship techniques.

In order to establish legitimacy in the eyes of the scholarly establishment, Reside suggests that curators and creators must build tools for gathering metrics into their projects and cooperate with external peer review groups—perhaps interdisciplinary peer review consortia modeled on NINES. In addition, in order to facilitate open-access publishing processes by taking a stand against traditional proprietary subscription structures, Reside promotes the idea that a portion of academic and professional society membership fees should be used to fund open-access journals. As well, project publishers can facilitate the preservation and reuse of data by working together to establish "data stores." At the 2010 HASTAC conference, Reside proposed "a collaboratory of web-based archives, with participants agreeing to publish all content at open and stable URIs for 10–15 years (in-copyright materials exempted)" (Trettien 2010). He suggested that "while software goes out-of-date rapidly, text and image files have 'remained relatively stable' for the last twenty years, and libraries still tend to adopt a 'protectionist approach,' guarding content that 'can only be accessed easily through the interface that the library provides.' Not much is gained by tying content to the interface . . . and 'potentially everything is lost.'"

In this section, we have discussed five interviews in which impressions of the scholarly work of archives and libraries have been conceived and perceived and actuated by archivists, librarians, and scholars as they negotiate the changing terrain of new technologies at different institutions. The impression is that archival, librarian, and scholarly work continues to be blurred within the context of advancing technologies, but these five practitioners and scholars highlight three main areas in which they see the combined efforts of future work in the archive. First, there is an implicit and explicit demand for the resources to build central, digital repositories that are open, interdisciplinary, multimedia, and built to support cross-

institutional projects, Second, each of these practitioners identifies the continued need for resources at archives and libraries to produce and publish small and large curated data sets or scholarly editions and projects. Third, there is a sense that these products must circulate for peer review and remixing within a wider community that is educated in the scholarly work these kinds of projects entail (Rockwell 2011; Schreibman, Mandell, and Olsen 2011).[13] In order to look more closely at these areas of future development, the final section of this discussion will examine progressive trends in community building in digital humanities.

Discussion: Imagining What We Don't Know

As we have seen, the separation of "preservation," "access," and "scholarship" has always been untenable, and it is with the implication that libraries should expand their scope to include more dynamic and flexible solutions to promote scholarly collaborations with faculty that Dorothea Salo asserts that library-run institutional repositories must "adapt or die" (2008). The work of an academic library in supporting, storing, and providing a framework for scholarly projects is still very much undefined, however. The final and concluding section of this article will discuss the third area of future development put forth by the digital humanities practitioners and scholars we interviewed as a means for achieving deepened collaborations— namely, that scholarly projects created in collaboration with libraries and archives must circulate within a wider community that is educated in the scholarly and archival work these kinds of projects entail (Zorich 2008). This section will discuss how some libraries, archives, and scholarly organizations achieve this circulation through community building, tool building, and peer review.

First, the emergent work of the archives of the future show an increased reliance on a deeper sense of community building among archivists, librarians, and scholars who work with new technologies. Institutional repositories that archive published faculty research, technical reports, and other scholarly research already play a role in allowing libraries to provide access to a wide range of scholarly materials, but a major goal of an academic library is to make its archival and other resources available as widely as possible. In particular, adopting open-source projects invested in sustainability and innovation gives libraries and archives access to active user communities. Communities such as Fedora Commons, the TEI Consortium, and the World Wide Web Consortium (W3C), for example, gather in online forums not only to create tools and standards but also to discuss issues concerning performance, interoperability, and sustainability. Other user-builder communities, sometimes called "digital humanities centers" (Zorich 2008), are often embedded within academic li-

13. In particular, the cluster of articles titled "Evaluating Digital Scholarship" in *Profession* (2011), published by the Modern Language Association (MLA) and edited by Susan Schreibman, Laura Mandell, and Stephen Olsen, is very helpful in identifying significant aspects of evaluating digital scholarship. In particular, see Rockwell (2012) and Schreibman et al. (2012).

braries as brick-and-mortar hubs that are built around using and building open-source projects for research and pedagogy, such as the Center for Digital Research in the Humanities (CDRH), the Maryland Institute for Technology in the Humanities (MITH), the Roy Rosenzweig Center for History and New Media (CHNM), and the Scholars's Lab at the University of Virginia Libraries, among others. These scholarly communities not only create tools and work with local scholars, they spawn international virtual communities, such as CenterNet, an international network of digital humanities centers first conceived at MITH;[14] the Praxis Program at the Scholars' Lab, which was created to train humanities scholars in alternative academic ("alt-ac") careers and which has already generated cross-institutional collaborations toward this goal;[15] THATCamp (The Humanities and Technology Camp), a roving "unconference" that local organizers develop and CHNM supports from afar (http://thatcamp.org/); and the new online journal *Scholarly Editing: The Annual of the Association for Documentary Editing*, an updated version of the *Documentary Editing* journal edited and maintained by CDRH that now includes peer-reviewed scholarly editions. All of these projects are built on and with open-access tools and standards and are open to and have gathered communities outside the walls of each of the centers' home institutions.

Open-source technologies encourage building. The flexibility inherent to open-source technologies, such as Fedora, for example, means that multiple communities can use the base system for a variety of solutions in different institutional settings. As Carl Lagoze and others explain, Fedora is "implemented as a set of web services . . . well-suited to exist in a broader web service framework and act as the foundation layer for a variety of multi-tiered systems, service-oriented architectures, and end-user applications" (Lagoze 2006, 2). Communities are developing a means for building a content management system like Drupal (http://drupal .org/) on top of the Fedora repository, which would allow for audience comment, participation, and collaboration. For example, the University of Prince Edward Island's Islandora (http://islandora.ca), an open-source digital asset management system that works with Fedora Commons and Drupal to support a scholarly editing environment that includes scanning, encoding, and editing, is currently working with the Editing Modernism in Canada project (http://editingmodernism.ca/2011/09/commonwealth-of-modernist-studies/) to produce a Digital Humanities Sprout. Other projects are also investigating ways of creating open-source Fedora overlays that are more robust. For example, the Hydra Project, a multi-institutional, distributed development project, allows for the grafting of "heads" or applications, such as image and rich media libraries, work-flow management, and exhibits on top of a Fedora-based repository. Hydra's strength lies in its flexibility and vibrant community of adopters who seek

14. For further information on CenterNet, see http://digitalhumanities.org/centernet/.
15. In particular, please see the TEIDisplay Omeka plug-in project described at http://www.scholarslab.org/announcements/collaborative-mentoring-at-ut-and-uva-co-developing-an-updated-teidisplay-for-omeka/.

to build strong products that meet a variety of different uses. Project Hydra's "ultimate objective is to produce a community-sourced, sustainable application framework that provides rich and robust repository-powered solutions as an integrated part of an overall digital content management architecture" (University of Virginia Libraries 2013). Likewise, the key innovation in many of these projects is the approach of purposefully developing a system with outside collaborations in mind.

There are other examples of this type of layered interaction in which the development of open-source web publishing platforms depends on collaborative, community building. Libraries most commonly use an open-source content-management system, such as Omeka (http://omeka.org/), for the purposes of creating online exhibits for their own collections, but Omeka's open-source structure and support for community involvement encourages other communities to create plug-ins for all to use that may be integrated in these systems. The Scholars' Lab, for example, has created plug-ins for Omeka that facilitate fuller integration between the contents of a digital repository and the Omeka interface, such as Neatline (Nowviskie 2009), which "allows scholars to combine timelines, maps, and scholarly narratives to explore literary and historical materials in new and exciting ways," and TEIDisplay (http://omeka.org/codex/Plugins/TeiDisplay), a plug-in for rendering TEI documents in Omeka. In another example, the Interedition Project (http://www.interedition.eu), which originated at the Huygens Institute for the History of the Netherlands at the Netherlands Royal Academy for Arts and Sciences in 2006 and has since become a COST (European Cooperation in the Field of Scientific and Technical Research) Action, attempts to produce a road map for the technical infrastructure involved in collaboratively preparing, editing, publishing, and analyzing digital literary research materials. The Interedition team makes a point to emphasize the importance of sustainability—the preservation of both digital project infrastructure and the raw textual data contained in digital editions. The developers of Interedition's CollateX (http://gregor.middell.net/interedition/) have developed a modular infrastructure that runs underneath the new web service for the Juxta Commons web service (https://github.com /performant-software/juxta-service/wiki/API-Documentation). Following the Gothenburg model of Collation,[16] Juxta Commons has an interface layer (the Commons) to support collation and comparing multiple witnesses to a single textual work, privately storing collations, sharing visualizations, and exporting formats such as TEI Parallel Segmentation in one environment. These developments pose even more opportunities for opening up Fedora repositories to scholarly work and collaborations.

16. At a 2009 conference sponsored by the European group, COST Action 32 (Open Scholarly Communities on the Web), in Gothenburg, Sweden, the makers of Juxta and CollateX combined forces to theorize best practices in collation software. The result of these conversations was the "Gothenburg Model of Collation," in which the previously interconnected processes of collation were separated into distinct modules to allow other users and applications to better interface with the tools.

The available, aggregated community surrounding these technologies also guides peer review and recommendations for developments and products. Essentially, the "public" element inherent to these open-source technologies means that scholars and libraries who use them and have questions or find bugs or use them in some innovative sense leave documentation that steers how these open-source technologies are implemented and by whom. For instance, TEI's strong community and well-documented guidelines have encouraged more traditional organizations like the MLA, which grants the preeminent prize in scholarly editions, to recommend that electronic files are "encoded in an open, nonproprietary format (e.g., TEI XML rather than Microsoft Word or WordPerfect"; MLA 2011). Smaller, cutting-edge communities like NINES, which is intent on creating alternate forms of peer review, state in their criteria for inclusion that for text-based projects TEI is recommended unless "overriding intellectual concerns justify an alternative schema" (MLA 2011). Funding agencies, which perform an essential process of peer review for methods and tools, consistently recommend the use of open-source software. For Digital Humanities Start-up Grants, the National Endowment for the Humanities (NEH) insists that "the use of open-source software is a key component in the broad distribution of exemplary digital scholarship in the humanities," while the Mellon Foundation also "prefers the development of open-source, modular applications, and open standards and specifications that are freely available and usable cost-effectively" since these are commonly "easily shared, extensible, and reliable" (Andrew Mellon Foundation 2010; NEH 2012). Lagoze sees these kinds of recommendations and "the motivation for integrating content management and the semantic web" as "originat[ing] from requirements defined by the broader Fedora user community" (Lagoze et al. 2006, 1). The underlying sense here with words like "available" and "reliable" emphasizes the fact that these emergent open-source technologies (software and standards) are free but also that their robust user communities make them common (and thus more communities will work to sustain their use); they are vetted by users who are knowledgeable and experienced (and publish and discuss their findings); and they are constantly in development (since no one must reinvent the wheel, the wheel becomes more and more refined).

At the same time, Gueguen reminds us that developing creative, future-looking implementations in the archive, even with a supportive community and access to open-source, cutting-edge tools and platforms, can still rely on the ingenuity and stamina of individuals with foresight (and a little spit and gum). For instance, when Teresa A. Sullivan was forced by the Board of Visitors of the University of Virginia to resign as university's president on June 10, 2012, the University of Virginia Library staff decided to begin archiving materials surrounding this riveting, well-publicized incident (2012). It was the first time the libraries had tried to preserve materials from a real-time, continuously unfolding event. To further complicate the collecting process, "the staff did not begin collecting materials on the subject until a rally on June 18," at which point, "as of June 22, the team ha(***d) archived nearly

20,000 tweets, 61 blog posts, over 200 media posts, and about 100 physical objects, such as signs from protests" (Chen 2012). Gueguen comments on the process of archiving such a large number of constantly updated new media resources by prefacing the new frontier this kind of collecting represents. First, they did not know what the scale of the response would be, and they began manually saving web pages in the browser as well as creating screen shots in order to capture both the code of sites and their look and feel. Beyond capturing websites related to the event, Gueguen collected news sources (e.g., local papers, the *Washington Post*, the *New York Times*, and the *Chronicle of Higher Education*), second versions of the sites she collected early on, video, and some other "'objects' like pictures, or pdf [*sic*] documents, etc."[17] For capturing tweets (or mini, 140-character postings on a platform called Twitter), Gueguen used a tool called "The Archivist" to create an XML output of the tweets, which she then post-processed into a tabular spreadsheet (http://archivist.visitmix.com/). She notes with interest: "The Archivist tool requires that I manually open the thing and search and save tweets, so it's not an ideal solution for a really chatty Twitter stream," especially since Twitter imposes a limit of 1,500 tweets for search results. As a consequence, Gueguen found that waiting past midnight to download and back up new tweets was necessary in order to capture as much activity as possible. Even though the Twitter data, the website, and the online objects will eventually be described as an archival collection in a finding aid (EAD) and be managed in the library's repository infrastructure, Gretchen notes that the tools she had at hand did not save other kinds of information upon which Twitter users regularly rely, such as images, profiles, or "twitpics" (pictures attached to tweets), making her options for collecting and archiving few. One of the most innovative aspects to the Sullivan project was the choice to create a website using Omeka for people who wished to upload their own contributions. Selecting and archiving continuously updated blogs and tweets in the public realm requires innovative collecting practices, but soliciting and then collecting and archiving public commentary on an event requires a whole host of other considerations that impact how this sort of exhibit or site or "collection" will evolve.

Conclusion

In conclusion, a sense of "textual performance," of live phenomena "situated in space and time," and of interactive communities gives a certain sense of fluidity to archives, scholarship, and modes of access in the digital realm (Clement 2011). On the one hand, this fluidity embodies the spirit of discovery made possible by emergent digital technologies in the archives of the future; on the other hand, it also comes with a sense of uncertainty and suspense and change that reflects our impressions of the blurred roles of archivists, librarians,

17. Gretchen Gueguen, e-mail to authors, July 25, 2012.

and scholars. At the same time, this fluidity also reflects our continued impressions of the future archive: an archive situated in the walls of a specific institutional setting in which there are a disparate range of resources and unrelenting demands for highly consistent standards that will be sustainable well into the future; an archive designed to produce and sustain and vet data that has both significant "contextual mass" (Palmer, Zavalina, and Fenlon 2010, 3) or thematic strength and great "analytic potential" (Palmer, Weber, and Cragin 2011, 2); a future archive built on a reliance on shared, interoperable infrastructure endeavors, such as the models discussed above and on community members giving back to the community; and a future archive enabled by an investment in mindfulness throughout the processes of creation, curation, and preservation that are the future archive's fundamental "purpose" (Pierazzo 2011, 475).

References

Andrew W. Mellon Foundation. 2010. "Research in Information Technology." http://web.archive.org/web/20100129073027/http://www.mellon.org/grant_programs/programs/rit.

ADE (Association for Documentary Editing). 1996. "A Prospectus for Electronic Historical Editions." http://xml.coverpages.org/mep-prosp.html.

Bryant, John. 2002. *The Fluid Text: A Theory of Revision and Editing for Book and Screen.* Ann Arbor: University of Michigan Press.

Chen, Angela. 2012. "UVa Creates Archive Documenting Sullivan's Resignation." *Chronicle of Higher Education*, June 22. http://chronicle.com/blogs/wiredcampus/uva-creates-archive-documenting-sullivans-resignation/37295.

Clement, Tanya. 2008. "A Digital Regiving: Editing the Sweetest Messages in the Dickinson Electronic Archive." In *A Companion to Emily Dickinson*, edited by Martha Nell Smith and Mary Loeffelholz, 415–36. Malden, MA: Blackwell, 2008.

Clement, Tanya. 2011. "Knowledge Representation and Digital Scholarly Editions in Theory and Practice." *Journal of the Text Encoding Initiative* 1 (June). http://jtei.revues.org/203.

Cox, Henry Bartholomew. 1969. "Publication of Manuscripts: Devaluation or Enhancement?" *American Archivist* 32 (1): 25–32.

Dekhtyar, Alex, Iacob E. Ionut, Jerzy W. Jaromczyk, Kevin Kiernan, Neil Moore, and Dorothy Carr Porter. 2006. "Support for XML Markup of Image-Based Electronic Editions." *International Journal on Digital Libraries* 6 (1): 55–69.

Derrida, Jacques. 1998. *Archive Fever: A Freudian Impression.* Chicago: University of Chicago Press.

Earhart, Amy E., and Andrew Jewell. 2011. *The American Literature Scholar in the Digital Age.* Ann Arbor: University of Michigan Press.

Faulhaber, Charles B. 1997. "Guidelines for Electronic Scholarly Editions." Modern Language Association of America, Committee on Scholarly Editing." Updated January 30, 2002. http://sunsite.berkeley.edu/MLA/guidelines.html.

Galloway, Patricia. 2006. "Archives, Power, and History: Dunbar Rowland and the Beginning of the State Archives of Mississippi, 1902–1936." *American Archivist* 69 (1): 79–116.

Greene, Mark A., and Dennis Meissner. 2005. "More Product, Less Process: Revamping Traditional Archival Processing." *American Archivist* 68 (2): 208–63.

Jewell, Andrew. 2008–9. "Digital Editions: Scholarly Tradition in an Avant-Garde Medium." *Documentary Editing* 30 (3/4): 28–35.

Jørgensen, Harald. 1983. "The Publication Policies and Practices of the Nordic Archives." *American Archivist* 46 (4): 401–13.

Lagoze, Carl, Sandy Payette, Edwin Shin, and Chris Wilper. 2006. "Fedora: An Architecture for Complex Objects and Their Relationships." *International Journal of Digital Libraries* 6 (2): 124–38.

Manoff, Marlene. 2004. "Theories of the Archive from across the Disciplines." *Libraries and the Academy* 4 (1): 9–25.

McGann, Jerome. 1995. "The Rationale of Hypertext." *IATH*. http://www2.iath.virginia.edu/public/jjm2f/rationale.html.

McGann, Jerome. 2007. "Database, Interface, and Archive Fever." *PMLA* 122 (5): 1588–92.

MLA (Modern Language Association). 2011. "Guidelines for Editors of Scholarly Editions." http://www.mla.org/cse_guidelines.

NEH (National Endowment for the Humanities). 2012. "Digital Humanities Start-Up Grants." http://www.neh.gov/files/grants/digital-humanities-start-sept-25-2012.pdf.

Nowviskie, Bethany. 2009. University of Virginia Library, Scholars' Lab. *Neatline.* http://www.scholarslab.org/uncategorized/neatline/.

Palmer, Carole L. 2004. "Thematic Research Collections." In *A Companion to Digital Humanities*, edited by Susan Schreibman, Ray Siemens, and John Unsworth, 348–65. Oxford: Blackwell.

Palmer, Carole L., Nicholas M. Weber, and Melissa H. Cragin. 2011. "The Analytic Potential of Scientific Data: Understanding Re-use Value." *Proceedings of the American Society for Information Science and Technology* 48 (1): 1–10.

Palmer, Carole L., Oksana L. Zavalina, and Katrina Fenlon. 2010. "Beyond Size and Search: Building Contextual Mass in Digital Aggregations for Scholarly Use." *Proceedings of the American Society for Information Science* 47 (1): 1–10.

Pierazzo, Elena. 2011. "A Rationale of Digital Documentary Editions." *Literary and Linguistic Computing* 26 (4): 463–77.

Price, Kenneth M. 2009. "Edition, Project, Database, Archive, Thematic Research Collection: What's in a Name?" *Digital Humanities Quarterly* 3 (3). http://www.digitalhumanities.org/dhq/vol/3/3/000053/000053.html.

Reside, Doug. 2011. "On Ant-Lions and Scholar-Programmers." *#alt-academy.* http://mediacommons.futureofthebook.org/alt-ac/pieces/ant-lions-and-scholar-programmers.

Rockwell, Geoffrey. 2011. "On the Evaluation of Digital Media as Scholarship." *Profession* 2011:152–68.

Salo, Dorothea. 2008. "Innkeeper at the Roach Motel." *Library Trends* 57 (2): 98–123.

Schreibman, Susan, Laura Mandell, and Stephen Olsen. 2011. "Introduction." *Profession* 2011:123–35.

Smith, Martha Nell. 2002. "Computing: What Has American Literary Study to Do with IT?" *American Literature* 74 (4): 833–57.

Theimer, Kate. 2012. "The Problem with the Scholar as "Archivist," or Is There a Problem?" *ArchivesNext.* February 28. http://www.archivesnext.com/?p=2522.

Trettien, Whitney. 2010. "Rethinking How We Publish Electronic Edition: Doug Reside @ HASTAC2010." http://hastac.org/blogs/whitneyt/rethinking-how-we-publish-electronic-edition-doug-reside-hastac2010.

University of Virginia Libraries. 2012. "Materials Relating to the Resignation and Reinstatement of President Teresa Sullivan." Archive in development.

University of Virginia Libraries. 2013. "Project Hydra." http://www.library.virginia.edu/blog/projects/hydra/.

Weinberg, Allen. 1962. "Publication Program of the Philadelphia Archives." *American Archivist* 25 (2): 193–97.

Zorich, Diane M. 2008. *A Survey of Digital Humanities Centers in the United States.* Washington, DC: Council on Library and Information Resources. http://www.clir.org/pubs/reports/pub143/pub143.pdf.

Tanya Clement: assistant professor, School of Information at the University of Texas at Austin. Clement has a PhD in English literature and language and an MFA in fiction. Clement has published on scholarly information infrastructures, digital humanities, digital literacy, digital scholarly editing, text mining and visualization, and modernist literature. E-mail (corresponding author): tclement@ischool.utexas.edu.

Wendy Hagenmaier: digital collections archivist at the Georgia Institute of Technology Library. Hagenmaier received her master of science in information studies degree, with a focus on digital archives and preservation, from the School of Information at the University of Texas at Austin. Her areas of scholarly and professional interest include the role of the digital in archives and humanities scholarship, new models of publishing, and the expansion of access to cultural heritage materials through technology. E-mail: wendy.hagenmaier@library.gatech.edu.

Jennie Levine Knies: manager, digital stewardship, University of Maryland Libraries. Knies has been managing digital projects at the University of Maryland Libraries since 2008 and is currently the manager of digital stewardship. Before this position, she worked for eight years as the curator for historical manuscripts at the University of Maryland Libraries. She received a bachelor's degree and a master of library science from the University of Maryland. She served as the chair of the EAD Roundtable of the Society of American Archivists from 2007 to 2008 and as vice chair of the Mid-Atlantic Regional Archives Conference from 2007 to 2009. E-mail: levjen@umd.edu.