College & Undergraduate Libraries

ISSN: 1069-1316 (Print) 1545-2530 (Online) Journal homepage: https://www.tandfonline.com/loi/wcul20

Digital humanities and the emerging framework for digital curation

Arjun Sabharwal

To cite this article: Arjun Sabharwal (2017) Digital humanities and the emerging framework for digital curation, College & Undergraduate Libraries, 24:2-4, 238-256, DOI: 10.1080/10691316.2017.1336953

To link to this article: https://doi.org/10.1080/10691316.2017.1336953

COLLEGE & UNDERGRADUATE LIBRARIES
2017, VOL. 24, NO. 2–4, 238–256
https://doi.org/10.1080/10691316.2017.1336953

Routledge
Taylor & Francis Group

Digital humanities and the emerging framework for digital curation

Arjun Sabharwal

Carlson Library, University of Toledo, Toledo, Ohio, USA

ABSTRACT

Digital humanities (DH) represents an emerging framework for digital curation. This study focuses on the evolving relationship of the DH (including history) to digital curation and archives and describes the models of collaboration in the digital environment. A conceptual framework is presented, followed by discussions of scarcity and abundance, archival frameworks, models of collaboration, and a review of selected DH projects. Despite the semantic disagreements over the term *archives*, researchers and curators should consider each other's perspectives for continued collaboration.

ARTICLE HISTORY
Received 20 February 2017
Revised 19 May 2017
Accepted 29 May 2017

KEYWORDS
Archives; collaboration; data curation; digital curation; digital humanities; projects

Introduction

Digital humanities (DH) represents an emerging conceptual and practical framework for digital curation in which theory, practice, collaboration, and digital content contribute to a rich foundation for new knowledge. This foundation has roots in the long-standing alliance between archives and the humanities, which has significantly changed over the past two decades. Throughout the history of this alliance, archivists have curated—that is, appraised, selected, arranged, described, and preserved—heritage materials and historic records of individuals and organizations. The work of historians and other researchers in the humanities, in turn, has focused on analyzing and interpreting data and information in the preserved records. Although collaboration between these two communities has continued, the dynamics of this relationship and the perception of what archives are have changed in light of new expectations and emerging practices. While often lumped together in the broader disciplinary context of the DH, DH and digital history have gradually developed distinct characteristics and adjusted to different technological possibilities in their respective domains (Robertson 2016). The models for collaboration in the DH landscape today range between the older models of domain-specific work—that is, archivists doing archival work and humanists

CONTACT Arjun Sabharwal arjun.sabharwal@utoledo.edu Carlson Library, University of Toledo, Mail Stop #509, 2801 W. Bancroft St., Toledo, OH 43606-3390.

engaging in hermeneutic work—to newer models supporting hybrid work—that is, both communities crossing over into each other's domains.

Changes in the humanities, digital technology, and the meaning of archives have affected this alliance; consequently, a paradoxical relationship has evolved over the past two decades. Among the multiple dimensions of this paradox, four are noteworthy for the purposes of the present study. The first one arises from the discourse on the dichotomy of scarcity and abundance familiar in the field of digital history. The second dimension represents the changing perception of archives—more specifically, the meaning of the word *archives* among practicing archivists and digital humanists. The third one refers to the changing dynamic between digital humanists and digital curators, and the fourth one relates to conflict between disintermediation and representation (Rosenzweig 2003) in relation to curatorial agency to mediate between content and user.

This study focuses on the evolving relationship of the DH (including history) to digital curation and archives, and it describes the models of collaboration in the digital environment. There are four parts to this study: It first presents a conceptual framework with reviews of selected literature, bringing DH, DH data characteristics, data curation, digital curation, and archives into a conceptual relationship. The ensuing section addresses the issue of scarcity and abundance faced by digital historians in particular, which nonetheless underscores the need for curators. The study then turns attention to archival frameworks shaping the relationships between archives and researchers in the humanities, followed by a review of various models of collaboration. The study underscores the transdisciplinary aspect of the emerging digital environment. Despite the debates over the meaning of archives, the pervasiveness of "digital"-ness and abundance, the ubiquity of popular albeit short-lived technologies, and the aforementioned paradox involved, there are shared core disciplinary interests that reinforce the continual relationship among DH work, digital curation, and archives focused on extending the life cycle of digital humanistic data and knowledge in this evolving landscape. The review of selected DH projects at the end of the study represents the close alliance holding researcher and archivist together despite disagreements over semantics.

Conceptual framework

Digital humanities

The field of DH has evolved from humanities computing, defined however inadequately as the "application of the computer to the disciplines of the humanities" (McCarty 1998) because, according to McCarty, "[it deletes] the agent-scholar from the scene and … [overlooks] the mediation of thought that his or her use of the computer implies" (1998, 2). McCarty notes that humanities computing is not merely an interdisciplinary scholarly activity but rather a discipline concerned with praxis instead. The *Companion to Digital Humanities* (Schreibman, Siemens, and Unsworth 2004) marks a new milestone for DH with theorists and practitioners, interdisciplinary focus, and new directions paved by the advent of the World Wide

Web (Schreibman, Siemens, and Unsworth 2004). As of 2008, however, the field of DH had not been recognized as "a unified field but an array of convergent practices" (Schnapp and Presner 2008). The Digital Humanities Manifesto 2.0 by Schnapp and Presner further presents DH as a discipline "that [explores] a universe in which: (a) print is no longer the exclusive or the normative medium in which knowledge is produced and/or disseminated; instead, print finds itself absorbed into new, multimedia configurations; and (b) digital tools, techniques, and media have altered the production and dissemination of knowledge in the arts, human and social sciences" (Schnapp and Presner 2009). The definitions from the first decade, therefore, describe the field as an emerging discipline with high potential to contribute new knowledge.

In more recent definitions, DH is "the use of digital media and technology to advance the full range of thought and practice in the humanities, from the creation of scholarly resources, to research on those resources, to the communication of results to colleagues and students" (Cohen, Frabetti, Buzzetti, and Rodríguez-Velasco 2011). Elsewhere, the field "refers to new modes of scholarship and institutional units for collaborative, transdisciplinary, and computationally engaged research, teaching, and publication" (Burdick, Drucker, Lunenfeld, Presner, and Schnapp 2012, 122). There is a new emphasis on the relationships of digital media, technology, and new modes of scholarship to archives. "Curation, analysis, editing, and modeling comprise fundamental activities at the core of Digital Humanities. Involving archives, collections, repositories, and other aggregations of materials, curation is the selection and organization of materials in an interpretive framework, argument, or exhibit" (17). According to Kathie Gossett, DH is an "umbrella term that covers a wide variety of digital work in the humanities: development of multimedia pedagogies and scholarship, designing and building tools, human computer interaction, designing and building archives, and so on" (as quoted in Gold and Klein 2012). Mark Tebeau views "the digital humanities as a collaborative, open, and emerging field of inquiry. A state of mind, a methodology, and theoretical approach to knowledge, it forces us to reconceive our practice" (as quoted in Gold and Klein 2012). Finally, in an earlier and widely cited definition from Wikipedia (n.d.), the DH are an area of research, teaching, and creation concerned with the intersection of computing and the disciplines of the humanities. Developing from the field of humanities computing, DH embrace a variety of topics, from curating online collections to data mining large cultural data sets. DH currently incorporate both digitized and born-digital materials and combine the methodologies from traditional humanities disciplines (such as history, philosophy, linguistics, literature, art, archaeology, music, and cultural studies) and social sciences with tools provided by computing (such as data visualisation [*sic*], information retrieval, data mining, statistics, text mining) and digital publishing (Wikipedia n.d.).

These definitions represent a continually expanding scope of interests and practices within DH, the most recent interest being data visualization. With increasing

references to digital media, it is important to address digital curation to ensure long-term preservation, access, and scholarly engagement—a topic that has dominated scholarly and professional discourse since 2001.

Characterization of DH data

Flanders and Muñoz (2011) have identified several types of data available to DH scholarship: (1) Scholarly editions; (2) text corpora; (3) marked-up digital texts (using TEI-XML markup); (4) thematic research collections (including aggregation of XML data, image files, style sheets, and configuration files); (5) research data (enhanced with analysis and annotations); and (6) archival finding aids (some using EAD-XML mark-up). DH data comes from narratives as well as raw, structured, and interpreted data for hermeneutic study. Archival finding aids are particularly important because their structure features a combination of narratives and a variety of data that helps researchers map up entire organizations based on information in series- and folder-level descriptions. The ArchiveGrid finding aid discovery system is particularly suitable for DH use because of its data harvesting and mapping applications, which can aid researchers using archival records to extract provenance information for any given geographical location.

Owens (2011) has treated humanistic data essentially as a textual artifact for textual analysis and interpretation. From an archival perspective, data possess evidentiary value

> as a species of human artifact, as a cultural object, as a kind of text, and as processable information [and it is] open to a range of hermeneutic tactics for interpretation. In much the same way that encoding a text is an interpretive act, so are creating, manipulating, transferring, exploring, and otherwise making use of data sets. (Owens 2011)

In archival practice and theory, collections with evidentiary value (while they are still being created) pass into having informational (or research) value after their transfer from the record creator's office to the archive, but, in the digital environment and in DH context, this transition may either be very short, immaterial, or possibly evaded.

There are three methods for processing humanities data: (1) interpretive layering, which involves curating metadata and annotations; (2) documenting data capture and preparation, which involves decisions on data curation; and (3) capturing scholarly agency vital to hermeneutic work. These methods correspond to different levels of curation: Data capture and preparation correspond to levels where metadata and other forms of documentation exist for long-term preservation and curation, while interpretive layering and capturing scholarly agency clearly belong to those levels of curation that support hermeneutic work and scholarly discourse. Data visualization is related to interpretive layering with the various temporal, geospatial, and social data gathered and interpreted through visualization technologies.

Finally, as part of the accession process, archivists prepare finding aids and metadata with potentially useful data for DH aggregation and visualization (maps, charts, and timelines), which includes provenance, contributor, biographical and

historical sketches, temporal coverage, and geospatial data needed for hermeneutic and historiographical work. Archival curation practices include exhibitions, exhibition catalogs, lecture series, presentations, and other educational events with rich textual data. Visualization can lead researchers to discovering relationships of events, people, and organizations in temporal and spatial proximity, which may have been overlooked by historians in the predigital environment. While provenance research has been a common practice among museum curators, provenance data can potentially elucidate historically misunderstood relationships and facts. In fact, The Mapping the Republic of Letters (Stanford University 2013) visualization project has analyzed and utilized data in case studies focused on notable authors such as Voltaire, D'Alembert, Galileo, Locke, and Franklin to visualize their correspondence, publications, and letters preserved by archivists. With these data, researchers can revisit, refute, or clarify historical facts and relationships.

Digital humanities data curation

The focus on data curation in the DH community differs from preservation- and conservation-centered curation practices in the archival community. The interests of digital humanists lie in building scholarly editions, data sets (and data visualizations), digital thematic research collections, websites, and digital archives. Much of the data going into projects like the Perseus Digital Library (2017), The Valley of the Shadows (Ayers 2007), Voyages Database (Emory University 2013), Walt Whitman Archive (Folsom and Price 2017), and the Digital Literary Atlas of Ireland, 1922–1949 (Travis 2017) has come from archives, libraries, museums, and historical and literary societies. However, any new knowledge generated through DH research would require archival curation skills to ensure long-term access. The cyclical process necessarily involves description, arrangement, finding aids, and preservation, yielding new data to support future DH data curation.

Data curation has been defined as "the active and ongoing management of data throughout its entire lifecycle of interest and usefulness to scholarship" (Cragin, Heidorn, Palmer, and Smith 2007), and it has been closely associated with DH practice in addition to the digital preservation and curation of cultural heritage material. It includes the following activities and processes: description (documenting the context and relationship of various forms and of research data); annotation (that is, enhanced information on the data with more granularity and context); collection and aggregation (connecting data and teams); storage (maintaining a platform for stable and accessible data); and migration (to ensure continued access via emulation or preservation).

Humanities data differ from digital objects in that they are

> presented in specialized aggregations that themselves have significance for understanding, using, and curating the data. Some of these aggregations are digital extensions of long-standing traditional forms such as finding aids, concordances, and scholarly editions. Thematic research collections or digital text corpus, in contrast, are products of new digital research methods. (Flanders and Muñoz 2011)

There is also a difference between digital curation and data curation in that the former focuses on preserving, contextualizing, and providing access to digital objects supplied with metadata, while the latter " … carries with it the burden of capturing and preserving not only the data itself, but information about the methods by which it was produced" (Flanders and Muñoz 2011).

Digital curation

Curation had originally meant "guardianship" under the purview of curators selecting, organizing, overseeing, and presenting collections and artifacts in exhibitions (Oxford English Dictionary 1971). However, in the DH context, a more recent definition is:

> the activity of managing and promoting the use of data, from its point of creation, to ensure it is fit for contemporary purpose, and available for discovery and re-use. For dynamic datasets, this may mean continuous enrichment or updating to keep it fit for purpose. Higher levels of curation (as in the Digital Humanities) will also involve maintaining links with annotation and with other published materials. (Lord and Macdonald 2003, 12)

Digital curation involves "the actions needed to maintain digital research data and other digital materials over their entire lifecycle and over time for current and future generations of users" (Bearie 2004, 7). This definition is important because it refers to preservation as articulated in Yakel's definition of curation as an "umbrella concept that includes digital preservation, data curation, electronic records management, and digital asset management" (2007, 335). Lee and Tibbo (2007) have defined digital curation as "stewardship that provides for the reproducibility and re-use of authentic digital data and other digital assets." They have also emphasized the need for trustworthy repositories utilizing open standards for long-term access and rigorous application for metadata standards for interoperability. Digital curation is an archiving activity to ensure the long-term preservation of research data and media content—a critical part of collaboration between DH and archives.

Debate on archives

In the evolving digital landscape for DH and curation, the definition of "archives" has taken center stage and has led to debates about its emerging purpose, meaning, and usage. Pearce-Moses (2012) has defined "archives" as

> materials created or received by a person, family, or organization, public or private, in the conduct of their affairs and preserved because of the enduring value contained in the information they contain or as evidence of the functions and responsibilities of their creator, especially those materials maintained using the principles of provenance, original order, and collective control.

Cox has highlighted the significance of archives and the trustworthiness of records amidst growing criticism of archives as "bureaucratic obstacles" (2005, 209) in the information age. Cox alludes to the vision of paperless offices, which begs the

question about the accountability of institutions and corporations not committed to preserving historic records. The meaning of "archives" had become the source of disagreement, however, as it began to change in DH discourse.

Kenneth Price, a coeditor of the Walt Whitman Archives, has questioned the meanings of such terms as archives, databases, editions, projects, and thematic research collections, asking whether these had any different and significant implications for DH scholarship or merely are legacies of predigital institutions of humanities, archives, and scholarly communication in general. He wrote,

> In the past, an archive has referred to a collection of material objects rather than digital surrogates. This type of archive may be described in finding aids but its materials are rarely edited and annotated as a whole. In a digital environment, archive has gradually come to mean a purposeful collection of surrogates. As we know, meanings change over time, and archive in a digital context has come to suggest something that blends features of editing and archiving. (Price 2009)

However, Price appears to have discounted the fact that archives continually edit their finding aids when collections are reappraised or the donors add new records. Price has indicated that "digital thematic research collections" may well represent the union of edition, project, and database—concepts that are endemic in DH practice.

In her response, Theimer (2012) pointed out that while archivists and digital humanists might use the term *archives* correctly in their respective communities of practice, "there is the potential for a loss of understanding and appreciation of the historical context that archives preserve in their collections, and the unique role that archives play as custodians of materials in this context." Theimer has pointed to four principles in archival practice: arrangement by provenance (or *respect des fonds*) telling the history of the record; collective control involving description and arrangement (often according to provenance); the concept of original order to preserve the original context of records; and intrinsic value of truly unique and surviving artifacts.

The divergence in the usage of the term *archive* is nevertheless important because many DH projects—including so-called "digital historical representations" (Sternfeld 2011)—have continued to apply this term in reference to materials from traditional, digital, and hybrid archives. Digital humanists can collaborate with archivists and digital curators to develop highly specialized digital archives while taking advantage of high-quality collections of digitized and born-digital primary sources and finding aids to navigate knowledge in that field. In order for such collaboration to be possible, it is important for archivists to consider the disciplinary perspectives of researchers and the prospects for future use of digital archives.

Scarcity, abundance, and the effect on DH data curation

While the digital environment has ushered in new opportunities for collaboration between digital humanists and archivists, challenges caused by information overload have also affected both communities of practice. Scarcity and abundance are

opposite conditions that have affected humanities scholarship and the work of archivists. According to Rosenzweig (2003), there are threats and benefits in this transition. He warned that historians may be facing a fundamental paradigm shift from a culture of scarcity to a culture of abundance. Not so long ago, we worried about the small numbers of people we could reach, pages of scholarship we could publish, primary sources we could introduce to our students, and documents that had survived from the past. At least potentially, digital technology has removed many of these limits: Over the Internet, it costs no more to deliver the AHR [*American Historical Review*] to 15,000,000 people than 15,000 people; it costs less for our students to have access to literally millions of primary sources than a handful in a published anthology. And we may be able to both save and quickly search through all of the products of our culture. Rosenzweig also sees a growing potential for existing and emerging media not being accessible to the equipment of the future, and the degradation of the media has become a serious concern to historians using digital formats for scholarly editions and collections. This issue affects various analog and digital media formats as well as computer-generated materials created with equipment now considered obsolete.

Digital preservation is a critical strategy to ensure what Hedstrom (2001) has called "temporal interoperability [or] the ability of current systems or legacy systems to interoperate with future systems that may use new formats, data models, languages, communication protocols, and hardware." In fact, digital preservation and format migration are two curatorial strategies with vital importance to the future of the DH. Digital preservation is a "series of managed activities necessary to ensure continued access to digital materials for as long as necessary" (Jones and Beagrie 2000, 10), while format and content migration are routines in migration and data transformation to preserve access to the information on obsolete and unstable storage media.

Rosenzweig (2003) has also pointed out that "authenticity" and "trust"—usually associated with historic primary sources in print formats (except for forged and fabricated materials)—present serious concerns in validating digital content. While prints have come with unique features such as watermarks and uniqueness in the paper's fiber, a systematic differentiation has not been the case with digital content. Even where a digital master (e.g., a TIFF or PDF/A file) may be the original, it is difficult to establish its authenticity in the absence of password-protected internal documentation. Metadata for historical texts must contain documentation such as provenance information as well as a digital signature, along with other possible encrypted data to be saved into a PDF/A file, to help researchers establish the authenticity and credibility of digital masters and publicly accessible digital surrogates. The advantage of this file format is that it can prevent users (other than the curator) from altering the documentation of information embedded in the file, hence underscoring the reputation and credit of the archives.

In addressing the previous problems, Rosenzweig clearly pointed to two critical issues: technical and cultural. The first one speaks to the onslaught of digital content that many archives may not be able to keep up with for reasons related to staffing,

equipment, and funding. The cultural issue points to researchers' distrust toward digital resources regarding veracity and integrity, but this is where digital historians and humanists can work with archivists to temper these concerns. A curator's certification, as discussed earlier, may put some concerns to rest.

Archival frameworks in DH context

The French Revolution and the Age of Enlightenment represent milestones and the starting point of a new era in a much longer history of archives going back to ancient Mesopotamian civilizations.

Cook's chronology presents the modern history of archives in four stages he called "paradigms … as frameworks for thinking about archives" (Cook 2013, 97), which are (in chronological order) evidence, memory, identity, and community; each framework also represents an important milestone in the professionalization and integration of archives in scientific and humanistic research. These frameworks not only signify the place of archives in society but also the changing dynamics of collaboration between humanities scholars and archivists. Placed within the context of DH scholarship, these frameworks present a succession of evolving models of collaboration between these two communities of practice. The four frameworks in Cook's (2013) chronology are: (1) "Evidence: pre-modern archiving [where] the custodian-archivist guards the juridical legacy" (106); (2) "Memory: modern archiving [where] the historian-archivist selects the archive" (107); (3) "Identity: postmodern archiving [where] the mediator-archivist shapes the societal archive" (109); and (4) "Community: participatory archiving [where] the activist-archivist mentors collaborative evidence- and memory-making" (113). Throughout these stages, as archivists have gradually come out of their closeted existence in government archives, they have begun to ally themselves with historians and researchers in other disciplines, engaged in professional training for certification, and begun proactive participation in their communities. From evidence to community, the focus of archival curation has steadily shifted from mere description and arrangement to working with memory, identity, and community; to DH researchers, this means the transformation of the archival record as well as the increase in the volume of humanistic data over the past two centuries. The availability of social data in the "community" stage does not mean only abundance but also the self-reflectivity of that data in comparison to what researchers may find in the "evidence" stage.

The approaches taken to curating collections have been closely allied with those in other disciplines, as historians and social scientists—such as anthropologists and ethnographers—have worked with very different methodologies in representation. The most dramatic turn in this timeline, however, was the flipping of roles: As archivists initially tasked in curating records of governmental bureaucracies had become researchers and mentors in curating community identity and memory, humanists, in turn, have turned their attention to curation. In fact, this transformation is very similar to armchair anthropologists turning into ethnographers doing fieldwork. In the DH context, therefore, researchers would have to consider two

factors when analyzing and interpreting humanistic data: the role of the curator and the provenance of DH data in the various paradigms. Both are necessary to evaluate their sources in the proper context. While Cook's chronology underscores Price's view on changing archives, it also communicates the importance of incorporating archival expertise in future DH projects.

Evolving models of collaboration in DH practice

The timeline of successive archival paradigms poses two other important questions: How did archivists and curators interact with humanities researchers? On what types of DH projects have they collaborated in the past two decades? For present purposes, this study proposes three models of collaboration in the order reflecting the succession of the previously discussed archival paradigms: disjunct, tangential, and conjunct. These models indicate the degree of collaboration and the level (or lack) of interdisciplinary activity (Klein 1994). The earlier projects (1970s and earlier) fall into the disjunct category in which archivists did not interact professionally with researchers to produce scholarly work. In this model, each community of practice remains within its own boundary, limiting professional discourse to disciplinary boundaries. The tangential model, however, brings the two communities into closer contact with more professional interactions, leading to collaboration on projects and scholarly work. Archivists from the second paradigm forward were professionalized but also degreed in academic disciplines like history, anthropology, arts, and humanities, which has enriched professional discourse with more interdisciplinary work. Finally, the conjunct model breaks down the traditional barriers between the two communities of practice and transcends traditional disciplinary boundaries. Collaboration in this model can lead to hybrid (or crossover) work engaging scholars in curatorial projects and archivists and curators outside their traditional zones. It also introduces transdisciplinary dimensions in any professional discourse resulting in this model with humanities scholars in roles once associated with archivists by engaging in the "collecting, arranging, and curating 'thematic research collections and 'digital scholarly editions'" (Clement, Hagenmaier, and Knies 2013, 114). For archivists, it also means significant changes because "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" (115). Researchers do not completely leave their own disciplines, but the range of overlap can vary from minimal to considerable. The last two paradigms in Cook's analysis present the landscape for collaboration of this kind.

DH is ideally suited for collaboration among librarians, researchers, and technologists, as libraries provide space for digital collections and DH projects (where the level of IT support is amenable to such extents). There has been significant progress in collaborative relationships since "humanities disciplines have been slow to embrace these digital opportunities in comparison to their colleagues in the natural and social sciences. However, the field of DH contrasts with the traditional model of humanities scholarship as it shares its purpose of

amassing digital collections with libraries" (Cunningham 2010, 1). Strong partnerships are necessary for faculty–library collaboration in DH projects such as scholarly editions and digital archives using digital tools, media digitization, text encoding (TEI) projects, and data visualization.

In the area of digital history, Sternfeld has described a new landscape for collaboration between historians and archivists through the merger of historiography and archival theory: In this environment, historians no longer work with physical collections in the archive but, instead, manage a combination of digital content, new media, and data called "digital historical representations" (2010, 1). Hybrid work for digital historians may not involve building the interface per se but rather collaboration with technologists on the design and effectiveness of the search interface and with archivists on matters of metadata. While some historians have been deeply involved in the design of the search interface implemented by technologists at the institution or the academic software firm, the development of descriptions and metadata schema remains within the curators' purview.

The digital environment offers yet another issue for researchers: disintermediation. It briefly means the absence of an intermediary—curators, librarians, and archivist—as a mediator of history and memory. However, without maintaining intellectual control over masses of content, historians and curators may lose leverage over historical content, knowledge, and their handle on digital historiography. Moreover, lecture series, outreach programs, and exhibitions are programs allowing curators and historians to interpret—and thus represent—history to the public. Rosenzweig (2003) wrote, "If historians are to set themselves 'against forgetting' … then they may need to figure out new ways to sort their way through the potentially overwhelming digital record of the past." After all, Roszenzweig believed, the method of treating sources as traditional historians had done in the past may not be as effective in the digital environment. Therefore, the issue of abundance may be welcome by the public with the time, the passion, and the logistical advantage (i.e., more volunteers to plow through vast volumes of digital records) to frame its own history, memory, and identity. From the public's perspective, disintermediation may pave the way for a greater degree of freedom in historical self-representation. From the perspective of curators, however, representation and disintermediation may represent competing—if not altogether antithetical—issues affecting their curatorial roles as central or peripheral respectively. There may be another benefit to disintermediation, which involves granting the public unmediated access to history without ideologically influenced metanarratives. Disintermediation puts society in charge of interpreting its own history, and the same thing is possible with digital technology as long as curators retain the integrity of collections without overcurating such material. Although the preservation of digital content remains a high priority, this is where collaboration between historians and archives may run into another contradiction, which underscores the paradox. This is because archivists approach collection development through the lens of appraisal and selection while disintermediation assumes that the public asserts control over its own history as seen in the fourth archival paradigm in Cook's (2013) chronology. The holistic look at digital content

does not resonate in the archivist community; therefore, the gap between preservation and representation may at times be unhelpful in solving concerns about the future of historical work involving digital historical representations. The potential conflict between those favoring disintermediation on the one hand and representation on the other may also underscore the paradox of the relationship between digital humanists (especially historians) and digital curators.

Gorzalski has pointed out a crucial component of collaboration between archivists and researchers—that is, archivists should understand researchers' perspectives. He wrote, that

> while digital humanities offers a rich area of collaboration between archivists and digital humanists, the partnership has not been without friction between the professional theories and the practices of both disciplines. During a recent digitization project undertaken jointly by archivists and theater faculty at Southern Illinois University Carbondale (SIUC), different assumptions and perspectives revealed competing ideas about context and the authenticity of primary sources. (Gorzalski 2016, 162)

Respect for cross-disciplinary perspectives is essential to collaborative relationships of all types whether the communities of practice are engaged in collaboration models of the tangential or conjunct model. The perspective of the "historian and a culturally inclined historical geographer [with] a shared interest in the use of Geographical Information Systems (GIS)" (Foley and Murphy 2015) also underscores Gorzalski's concerns. Spatial visualization, for instance, is an important component of spatialized knowledge contained in archives and museums (Dallas 2007), but it is not a part of archival description at the collection or item levels. While archivists have adhered to standards in archival description combining DACS with Dublin Core or other schemata, the addition of temporal and geospatial data to the metadata record may help historical geographers further represent knowledge in their own disciplinary contexts. "A key benefit of a GIS approach is that it enables the researcher to identify spatial patterns within data which may otherwise be unapparent" (Foley and Murphy 2015).

This brief review of collaborative models in the digital environment has aimed to demonstrate the potential for productive collaboration across the communities of curators, researchers, and practices (despite the blurring distinctions) in the digital environment. However, all models of collaboration—especially in the tangential and conjunct models focusing on interdisciplinary approaches—require recognition and respect of perspectives of engaged practitioners. Therefore, if it means that DH practitioners get an opportunity to recognize the merits and advantages of true archives as vehicles for preserving humanistic data and knowledge, archival curators will also be in the position to extend the scope of their practice to accommodate digital humanists by enriching or enhancing some metadata records. Despite the evolving rhetoric on what true "archives" are, at the core of the aforementioned paradox is—as recent publications pointed out—the relationship between the two communities that is still conducive of long-term collaboration on scholarly production. The following review of selected projects in chronological order demonstrates the extent of collaboration between DH researchers and archivists.

Selected digital humanities projects

The following descriptions of early to more recent DH projects illustrate the evolving models of collaboration between curators and humanists as these projects were envisioned and implemented. They illustrate the work of historians, literary scholars, archivists, and librarians who have contributed to these projects to various degrees. The projects discussed present four distinct models of collaboration that have defined the DH and continue to do so in the present.

The Perseus Project

The Perseus Project 1.0 (presently at 4.0) was launched in 1987 (the planning stages in 1987) at Tufts University as an experiment to explore the effects of the electronic environment on libraries (Preece and Zepeda 2009). It covers Greco-Roman history, literature, and culture but has gradually expanded to modern English, American history, history and topography of London, history of mechanics, and Arabic language (Perseus Digital Library 2017). The project has been a collaboration of digital librarians, technologists, editors, student assistants, and researchers (from institutions across the United States). It features a metadata catalog linked to WorldCat and a mission statement focused on three areas: (1) human-readable information, (2) machine-actionable knowledge, and (3) machine-generated knowledge. While the first mission focuses on digital representations needed for human processing, the other two combine metadata and scripts to enable computing equipment to make other types of data available in some structured format (Perseus Digital Library 2017). Perseus 1.0 was published in print form in 1992, while 2.0 was also available in CD-ROM from 1996 to 2000. Perseus 3.0 became available as a website in 2000, in 2005 as a full web version, and in 2006, it featured a TEI-XML-coded collection of English, Greek, and Latin works that were released with a Creative Commons license.

Valley of the Shadow

The Valley of the Shadow project site is a digital archive in the Virginia Center for Digital History at the University of Virginia. It contains primary sources—letters, photographs, maps, and diaries—from the eve of the Civil War in 1860 to 1870 with a focus on Augusta County, Virginia, and Franklin County, Pennsylvania. Although Edward L. Ayers, an historian, had originally planned it as a print book, he instead created it as a hypertext (SGML-coded) project to create "a research library in a box, enabling students at places without a large archive to do the same kind of research as a professional historian" (Ayers 2007). In fact, all of the materials were selected form archives in these two counties, the National Archives, and the United States Army Military Institute library in Carlisle, Pennsylvania. IBM computers were used in converting the archival documents into hypertext, and in 1993, the site was viewed using the hypertext browser Mosaic (predecessor of Netscape, which is a predecessor of

Mozilla Firefox). The Institute for Advanced Technology in the Humanities (IATH) supported the project by donating the project site and providing key personnel to develop the coding for the individual parts of the project. The CD version was created under the leadership of William G. Thomas III (after taking over from Anne Rubin, who was leaving the project) in 1996. In 2002, the project moved to the web with much of the text converted to the XML schema, while maps were made accessible using GIS technology. The site architecture features a very organized hypertext environment with navigational consistency across the temporal zones (fall 1859 to spring 1861, spring 1861 to spring 1865, and spring 1865 to fall 1870), each of which has linkable regions to a Reference Center, Letters and Diaries, Census Tract Records, and Newspapers. This project involved historians, technologists, subject librarians, archivists, and digital curators.

Voyages databases

The project's aim was to create a "single multisource dataset of trans-Atlantic slave voyages" (Emory University 2013) based on the research of Herbert S. Klein and other scholars between the 1960s and the late 1980s. They developed a data set of about 11,000 voyages based on archival records from various European (British, Dutch, French, and Portuguese) governmental archives. By the end of the 1990s, the CD-ROM version of the data set included over 27,000 records; currently, there are almost 36,000. The website was launched in 2006 with Spanish and American data on slave trade and shipments from archival data. In addition to searchable data, the site presents data visualized in tables, maps, timelines, bar charts, and animations. The site provides period images from the beginning of the slave trade, including manuscripts and visual representations of African slaves shipped to the Americas and the vessels used for those shipments. A registry of African names (transliterated into Latin script) describes individuals, their origins, and destinations. The database is searchable and suitable for classrooms using this as an educational resource.

Thematic research collections

The William Blake Archive is a digital thematic research collection—alternatively referred to as a "digital archive"—with a focus on the original poems and paintings of William Blake. Launched in 1996, the website was a collaborative project between the editors and the IATH. The project involved an international collaboration of several universities, museums, libraries, private collectors, and societies contributing content to the digital archive. The site is searchable and organized by genres (illuminated books, commercial book illustrations, prints, drawings and paintings, and manuscripts); each page image comes with so-called diplomatic translations showing the original editing by Blake, and TEI (Text Encoding Initiative) encoding has enabled the site to include changes in his manuscripts. The archive features a biography, glossary, biographical timeline, and a hypertext narrative on illuminated printing. Pop-up images open when users click or tap the footnotes in the text. Although

many libraries have contributed content, there is no catalog or finding aid in this digital archive.

The Walt Whitman Archive (Folsom and Price 2017) is another digital archive created under the lead of Folsom and Price in 1995. The full text-searchable site makes extensive use of TEI editing to be viewed side by side with the corresponding page images, especially in the case of his poetry manuscripts, which is where most of the editing happened. Many pages contain metadata pointing to the source (an archive or library) of the manuscripts, which enables skeptical scholars to examine the editions. There are also criticisms and critical reviews of Whitman's publications by various literary scholars, supplied with bibliographies and catalogs of metadata showing on the individual pages as well. This is not a standard library catalog, but it is compliant with the currently used EAD (Encrypted Archival Description) standards used in many archives. The site presents a comprehensive biography with links to related biographies and a link to the encyclopedia index.

Perhaps one of the most recent innovative developments in DH is the use of geospatial data to emphasize the spatial—in addition to the temporal and social—dimensions of DH since narratives migrate when people do, which directly contributes to the diffusion of artistic, musical, literary, and other creative genre over extensive geographical spaces. The Digital Literary Atlas of Ireland (Travis 2017) project at the Trinity College of Dublin combines (prior to the site being decommissioned) biographical data with interactive time line and geospatial technology, which present Irish literary history and biographical knowledge in a temporal-spatial context. The project has focused on the literary, historical, and cartographic perspectives on Ireland between 1922 and 1949 through the eyes of fourteen selected Irish writers with great significance in the context of Irish national history and cultural identity. The user interface features three access points to biographical information: (1) by featured authors' life paths and a biographical narrative with references to sources; (2) time line, using Google Earth time line outlining the life of these authors as they moved around; and (3) a Vimeo presentation of maps featuring an overview of the terrain, street-level zooming with a Google Earth plug-in that allows a significant level of interactivity for viewers interested in the geographical aspects of humanities. The source of the map images used in this project was Trinity College's Glucksman Map Library Collection. Other than the bibliography and footnotes in the biographies, the site does not provide references to individual sources in archives and libraries.

Mapping the Republic of Letters

The visualization project, completed in 2009, focuses on the networks of scholars existing before professional associations, colloquia, and other organized events allowed networking (Stanford University 2013). In the landscape of print sources, correspondence was the only way to document the relationships of acquainted scholars. This project aimed to map the correspondence, travels, and publications

organized around case studies of historic scholars. The visualizations include maps, bar charts, and digitized images from archives and libraries, which can help researchers analyze, verify, clarify, or refute historically accepted facts in print literature. For instance, the use of interactive layered maps for Voltaire's places of publication between 1712 and 1800 can help researchers verify discrepancies in the historical record about the places were Voltaire's works were and were not published. Another map presents Voltaire's correspondence network with each location linked to metadata in the Electronic Enlightenment catalog. Again, this visualization tool can provide further context for Voltaire's publishing record. Another case study (Travels) has made heavy use of A Dictionary of British and Irish Travellers in Italy, 1701–1800, where the information was organized into structured data rendered by a visualization tool into a timeline, map, or bar chart that enable researchers to identify travelers frequenting (or even staying at) European locations by using a search box. Yet another visualization tool on this site presents a social space based on the network of relationships of specific individuals in a specific date range in specific locations. This approach may not only answer research questions but also spark new ones based on concrete primary data from archival sources.

Conclusion

The digital environment has redefined the humanities, archives, and the practice of curation. Archives have supplied primary source materials to historians and other humanists for centuries since the formal organization of public archives in the eighteenth century. Since then, the reformation of academic disciplines and institutions and the professionalization of archives have prepared the landscape for collaboration between the DH and archives in the digital environment. Archivists are highly specialized in the appraisal, selection, description, arrangement, preservation, and curation of unique collections as evidence for historical and humanities research; hence this process is not bureaucratic as much as it is a matter of time needed for documentation and preparation for use by researchers expecting reliability and credibility almost always guaranteed with print collections. The reproduction of digital collections in archives, however, has met with skepticism from humanists and historians suspicious of digital content, but archivists can eliminate these concerns by documenting the scanning and metadata creation processes as a way to assure the quality and reliability of the digital resources contributed to the institutional repositories. A digital certificate issued by the archives should be sufficient to alleviate researchers' concerns about the authenticity of scanned documents and fragile primary sources. The role of archivists in the digital environment as preservationists, therefore, presents a key argument for representation and against disintermediation, the latter of which is favored in some DH communities. The evolution of the archival framework since the French Revolution showed that archivists have become a less bureaucratic factor and more of a

supportive element in the digital landscape where communities and individuals are curating their own histories and cultural heritage. The models for collaboration between DH researchers and archives have also transformed into a more mutually supportive and agreeable model where these two communities have begun to cross over into each other's domain.

The use of social media in DH curation practice—or social curation—is a recent development that merits a separate analysis that is beyond the scope of this study. In this environment, both researchers and curators are highly visible to the archives community, which has focused its attention on the preservation of textual data and media. Given the ephemeral nature of digital content in social media—which is due to the short life cycle of contents in social networks—archivists and librarians have begun promoting the idea of personal digital archiving, which supports the philosophy of preserving not just media and data but also the identity, memory, and heritage of individuals, families, and communities.

In this digital landscape of abundance and ephemeral content, archivists and curators can still play a crucial role in collaborating with researchers. However, the culture of mutual inclusiveness must prevail. Humanists should realize that DH scholarship (data, information, and knowledge) must be preserved professionally by professionally trained archivists, while the latter should expand their scope of curation by accommodating humanists, who, in turn, focus their work on digital scholarly editions, digital thematic research collections, digital archives, and databases. The mutual consideration of each other's domains, practices, and priorities can only prolong collaboration.

References

Ayers, Edward L. 2007. "The Story behind the Valley Project." *The Valley of the Shadow*. http://valley.lib.virginia.edu/VoS/usingvalley/valleystory.html

Bearie, Neil. 2004. "The Digital Curation Centre." *Learned Publishing* 17(1): 7–9. https://doi.org/10.1087/095315104322710197.

Burdick, Anne, Johanna Drucker, Peter Lunenfeld, Todd Presner, and Jeffrey Schnapp. 2012. *Digital Humanities*. Cambridge, MA: MIT Press. https://mitpress.mit.edu/sites/default/files/titles/content/9780262018470_Open_Access_Edition.pdf.

Clement, Tanya, Wendy Hagenmaier, and Jennie Levine Knies. 2013. "Toward a Notion of the Archive of the Future: Impressions of Practice by Librarians, Archivists, and Digital Humanities Scholars." *The Library Quarterly* 83(2): 112–30. https://doi.org/10.1086/669550

Cohen, Daniel J., Federica Frabetti, Dino Buzzetti, and Jesus D. Rodríguez-Velasco. 2011. "Research without Borders: Defining the Digital Humanities." Columbia University Academic Commons. [YouTube video]. http://hdl.handle.net/10022/AC:P:14226.

Cook, Terry 2013. "Evidence, Memory, Identity, and Community: Four Shifting Archival Paradigms." *Archival Science* 13(2): 95–120. https://doi.org/10.1007/s10502-012-9180-7

Cox, Richard J. 2005. *Archives & Archivists in the Information Age*. New York: Neal-Schuman.

Cragin, Melissa H., Bryan P. Heidorn, Carole L. Palmer, and Linda C. Smith. 2007. "An Educational Program on Data Curation." Poster presented at the American Library Association Conference, Science and Technology Section, Washington, DC (June 25).

Cunningham, Leigh 2010. "The Librarian as Digital Humanist: The Collaborative Role of the Research Library in Digital Humanities Projects." *Faculty of Information Quarterly* 2(21): 1–11. http://fiq.ischool.utoronto.ca/index.php/fiq/article/view/15409/12437.

Dallas, Costis. 2007. "An Agency-Oriented Approach to Digital Curation Theory and Practice." In *International Cultural Heritage Informatics Meeting (ICHIM07): Proceedings,* edited by J. Trant and D. Bearman. Toronto: Archives & Museum Informatics. http://www.archimuse.com/ichim07/papers/dallas/dallas.html.

Emory University. 2013. "*History of the Project*." Voyages: The Trans-Atlantic Slave Trade Database. http://www.slavevoyages.org/about/history.

Flanders, Julia, and Trevor Muñoz. 2011. "An Introduction to Humanities Data Curation." In *DH Curation Guide*. https://web.archive.org/web/20141022125715/http://guide.dhcuration.org/intro/.http://guide.dhcuration.org/intro/.

Foley, Ronan, and Rachel Murphy. 2015. "Visualizing a Spatial Archive: GIS, Digital Humanities, and Relational Space." *Breac: A Digital Journal of Irish Studies*. https://breac.nd.edu/articles/61505-visualizing-a-spatial-archive-gis-digital-humanities-and-relational-space/.

Folsom, Ed, and Kenneth M. Price (eds.). 2017. "About the *Archive*." *The Walt Whitman Archive*. http://whitmanarchive.org/.

Gold, Matthew K., and Lauren F. Klein. 2012. *Debates in the Digital Humanities*. Minneapolis: University of Minnesota Press http://dhdebates.gc.cuny.edu/debates/text/40.

Gorzalski, Matthew J. 2016. "Archivists and Thespians: A Case Study and Reflections on Context and Authenticity in a Digitization Project." *The American Archivist* 79(1): 161–85. http://opensiuc.lib.siu.edu/morris_articles/65/.http://opensiuc.lib.siu.edu/cgi/viewcontent.cgi?article=1080&context=morris_articles.

Hedstrom, M. Margaret. 2001. "Exploring the Concept of Temporal Interoperability." *Proceedings of the Third DELOS Network of Excellence Workshop on Interoperability and Mediation in Heterogeneous Digital Libraries*. http://www.ercim.eu/publication/ws-proceedings/DelNoe03/10.pdf.

Jones, Maggie, and Neil Bearie. 2000. *Preservation Management of Digital Materials: Workbook: Pre-Publication Draft*. London: Resource.

Klein, Julie Thompson. 1994. *Interdisciplinarity: History, Theory, and Practice*. Detroit: Wayne State University Press.

Lee, Christopher A., and Helen Tibbo. 2007. "Digital Curation and Trusted Repositories: Steps Toward Success." *Journal of Digital Information*, 8(2). https://journals.tdl.org/jodi/index.php/jodi/article/view/229/183.

Lord, Phillip, and Alison Macdonald. 2003. "Data Curation for E-Science in the UK: An Audit to Establish Requirements for Future Curation and Provision " Report for The JISC Committee for the Support of Research. https://web.archive.org/web/20050825130427/http://www.jisc.ac.uk/uploaded_documents/e-ScienceReportFinal.pdf.

McCarty, Willard 1998. "What Is Humanities Computing? Toward a Definition of the Field." https://www.researchgate.net/https://www.researchgate.net/publication/248421565_What_is_Humanities_Computing_Toward_a_Definition_of_the_Field.

Owens, Trevor 2011. "Defining Data for Humanists: Text, Artifact, Information or Evidence?" *Journal of Digital Humanities* 1(1). http://journalofdigitalhumanities.org/1-1/defining-data-for-humanists-by-trevor-owens/.

Oxford English Dictionary. 1971. "Curation." *The Compact Edition of the Oxford English Dictionary: Complete Text Reproduced Micrographically*. Oxford: Clarendon Press.

Pearce-Moses, Richard. 2012. "Archives." Society of American Archivists. https://www2.archivists.org/glossary/terms/a/archives.

Perseus Digital Library. 2017. "About the Perseus Digital Library." http://www.perseus.tufts.edu/hopper/about.

Preece, Emily, and Christine Zepeda. 2009. "The Perseus Digital Library: A Case Study." https://repositories.lib.utexas.edu/bitstream/handle/2152/6836/perseus_case_study.pdf;sequence=5.

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

Robertson, Stephen 2016. "The Differences between Digital Humanities and Digital History". In *Debates in the Digital Humanities*, edited by M. K. Gold and L. F. Klein. Minneapolis, MN: University of Minnesota Press. http://dhdebates.gc.cuny.edu/debates/text/76.

Rosenzweig, Roy 2003. "Scarcity or Abundance? Preserving the Past in a Digital Era." *American Historical Review* 108(3): 735–62. https://chnm.gmu.edu/digitalhistory/links/pdf/introduction/0.6b.pdf.

Schnapp, Jeffrey, and Todd Presner. 2008. "A Digital Humanities Manifesto." http://manifesto.humanities.ucla.edu/2008/12/15/digital-humanities-manifesto/.

Schnapp, Jeffrey, and Todd Presner. 2009. "The Digital Humanities Manifesto 2.0." http://manifesto.humanities.ucla.edu/2009/05/29/the-digital-humanities-manifesto-20/.

Schreibman, Susan, Ray G. Siemens, and John Unsworth. 2004. "The Digital Humanities and Humanities Computing: An Introduction." In *A Companion to Digital Humanities*, edited by Susan Siemens, Ray Siemens, and John Unsworth. Malden, MA: Blackwell. http://www.digitalhumanities.org/companion/.

Stanford University. 2013. "Mapping the Republic of Letters." http://republicofletters.stanford.edu/.

Sternfeld, Joshua. 2010. "Thinking Archivally: Search and Metadata as Building Blocks for a New Digital Historiography." http://dh2010.cch.kcl.ac.uk/academic-programme/abstracts/papers/html/ab-747.html.

Sternfeld, Joshua. 2011. "Archival Theory and Digital Historiography: Selection, Search, and Metadata as Archival Processes for Assessing Historical Contextualization." *The American archivist* 74(2): 544–75.

Theimer, Kate. 2012. "Archives in Context and as Context." *Journal of Digital Humanities* 1(2). http://journalofdigitalhumanities.org/1-2/archives-in-context-and-as-context-by-kate-theimer.

Travis, Charles. 2017. "Digital Literary Atlas of Ireland, 1922–1949." https://web.archive.org/web/20161228025736/http://www.tcd.ie//trinitylongroomhub/digital-atlas/writers/peadar-odonnell/.

Wikipedia. n.d. "Digital Humanities." https://web.archive.org/web/20131207093955/http://en.wikipedia.org/wiki/Digital_humanities.

Yakel, Elizabeth. 2007. "Digital Curation." *OCLC Systems & Services: International Digital Library Perspectives* 23(4): 335–40. https://doi.org/10.1108/10650750710831466