

Beyond the Scanned Image: A Needs Assessment of Scholarly Users of Digital Collections

Harriett E. Green and Angela Courtney

This paper presents an analysis of how humanities scholars use digital collections in their research and the ways in which digital collections could be enhanced for scholarly use. The authors surveyed and interviewed humanities faculty from twelve research universities about their research practices with digital collections and present analysis of the resulting responses. The paper also analyzes a sample of qualitative responses from the Bamboo Technology Project's workshops with faculty, librarians, and technologists about the use and functionalities of digital materials for humanities research. This paper synthesizes these data analyses to propose the critical need for interoperability and data curation in digital collections to increase their scholarly use, and the importance of user engagement in development of digital collections.



Humanities scholarship integrates the analog with the digital more fully than ever before, as new approaches to humanities scholarship incorporate digital materials, from comprehensive archival projects that require the gathering of materials from around the world to research that uses analytic tools for text mining and social network analysis. This evolution in humanities scholarship emerging from the rise of digital humanities and deeper interdisciplinary orientations for humanities scholars prompt us to ask critical questions about the digital collections that frequently play a significant role in these new scholarly methodologies: specifically, how effectively are digital collections meeting the research needs of scholars, and how should digital collections evolve to sustain and strengthen their value to digital humanities research?

This paper presents the results of a study that examines how humanities scholars make use of digital collections and the ways in which digital collections could be enhanced for scholarly research. Through analyses of the survey and interviews of humanities scholars conducted for this study, combined with analyzed responses from focus groups of scholars at the Bamboo Technology Project workshops, this paper argues that libraries, museums, and archives should look beyond discovery

Harriett E. Green is English and Digital Humanities Librarian and Assistant Professor, University Library at the University of Illinois at Urbana-Champaign, e-mail: green19@illinois.edu; Angela Courtney is Head, Arts and Humanities Department and Reference Department at the Indiana University at Bloomington Libraries, e-mail: ancourt@indiana.edu. © 2015 Harriett E. Green and Angela Courtney, Attribution-NonCommercial (<http://creativecommons.org/licenses/by-nc/3.0/>) CC BY-NC.

and access issues when evaluating the use and usability of their digital collections. Content providers must also consider the complex user experience issues that emerge in examination of research practices of humanities scholars—such as interoperability and data curation—that critically influence how digital collections are integrated into humanities scholarship.

Background

Digital collections are defined for the purposes of this study as dynamic and coherent aggregations of thematic digital content that provide a “dense unit for exploration or study.”¹ The Library of Congress established the National Digital Library Program in 1994 to pioneer the standards and processes for digitization of research archives; in the following two decades, digital collections have been developed by librarians and scholars for resource preservation and research use. Today there are thousands of terabytes of digitized material, and digital collections encompass a diverse range of sizes, scopes, and topical foci. Humanities scholars’ use of digital resources, a category that comprises commercial databases as well as thematic digital collections, has been explored in a number of studies during the past several decades.

Much of the early research on digital collections analyzes how humanities scholars incorporated early electronic resources into their research, not features and functionality that they need for those collections to be most useful. Gilmore and Case describe the potential of hypertext and large electronic collections in libraries to bridge the digital divide for historians by facilitating connections between texts, but they note the limitations of those electronic resources for historians’ research practices.² Duff and Cherry examine researchers’ use of *Early Canadiana Online/Notre Memoire En Lingne (ECO)* to compare user satisfaction across the digital, print, and microform version of the collection, and their findings focused primarily on issues of speed and perceived authenticity.³ Early research was conducted on humanities scholars’ use of online search systems and databases, such as the series of studies by Bates, Siegfried et al. on the Getty End-User Online Searching Project.⁴ In their observation studies of end user behavior in DIALOG, Bates et al. conclude that the search system did not offer enough benefit to the scholar to be more than a supplement to their research, but they note, “one of the most exciting challenges in the development of future information systems is to identify what resources and search features are needed to provide an automated information resource that the humanities scholar will come to feel is an indispensable part of his or her laboratory.”⁵

As electronic resources evolved in complexity, subsequent studies examined if and how humanities researchers incorporated more digital materials into their workflows. Brockman et al. discuss the growing expectation of humanities researchers to be able to locate and access much of their research materials through their computers, yet this study also noted a lack of certainty and confidence of researchers in their ability to make online resources work effectively.⁶ Spiro and Segal examine citations in humanities publications to ascertain whether digital humanities projects were being cited in published scholarship.⁷ The study of online resource use and potential sustainability done by Warwick et al. suggests that the growing use of a digital resource increases its likely sustainability.⁸ Sukovic extensively analyzes humanities scholars’ research practices with electronic texts; and, while her study maps researchers’ research practices in depth, the study is limited to historians and literary scholars that work primarily with text.⁹

The effectiveness of digital collections for scholarly use has been the focus of several recent studies. Meyer explores the impact of digital resources on scholarship in two studies—which ultimately led to the development of the Toolkit for the Impact of Digi-

tised Scholarly Resources (TIDSR)—and cites the increasing frequency of citations of digital projects in scholarly humanities research.¹⁰ Proffitt and Schaffner focus predominantly on content when they find that scholars are concerned about sustaining digital resources into the future, suggesting that librarians need to increase their involvement in academic departments to prioritize the digitization of library collections.¹¹ Building upon his 2012 impact study, Sinn analyzes scholars' use of digital collections, drawing upon analysis of ten years of citations from the *American Historical Review* to map the usage of a diverse range of collections, but this study does not engage scholars for their input on how they would like to see collections created in the future.¹² Gertz's recent exploration of the prioritization of rare books and special collections for digitization suggests that value and demand must be complemented by high-quality digitization, strong and effective metadata, robust searching capacity, and concern for sustainability.¹³ Her assessment considers a varied user audience, but it does not suggest asking the target users what they need from a digital collection.

A few recent reports have engaged scholars in their use of digital collections: Zorich examines the state of digital scholarship in art history through interviews with art historians and art history research centers.¹⁴ Bulger et al. present case studies of research practices with digital collections that are based upon interviews with humanities scholars from multiple disciplines: several of their case studies only focus on the use of specific digital humanities resources such as the *Digital Image Archive of Medieval Music*, however, and other cases consist of respondent pools limited to highly specialized fields or one institution's department.¹⁵ Yet their findings are a useful guide: Bulger et al. note that, while humanities researchers have begun to embrace digital tools and to work in increasingly collaborative ways, the scholars still need training to effectively use advanced tools for data analysis.¹⁶ They also propose a "complexity continuum" graph that maps the myriad ways humanities scholars use digital resources; from this, they observe that, because scholars merge digital resources into a hybrid workflow still dominated by print resources, this highlights "the difficulty of assessing and documenting the impact digital resources are having on scholarship and learning."¹⁷

Thus, while there is a rapidly expanding body of research on the use of and infrastructure for digital collections, less has been written on functionalities of digital collections critical for the research needs for humanities scholars, and even less attention has been given to learning what humanities scholars think is important or missing from these collections. The results of this needs assessment study begin to connect the expressed needs of humanities scholars to the infrastructural aspects of digital collections. This prompts a deeper investigation into how scholar engagement and partnerships can enable creators of digital collections to translate needs into realistic enhancement of digital collections' content and functionalities.

Methods and Study Design Method

Research Design

A mixed methods research design was used for this study, as this study sought to conduct a needs assessment of scholarly research with digital collections, and a mixed methods design provided the needed "variety of data sources and analyses to completely understand complex multifaceted institutions or realities."¹⁸ This approach also "allows researchers to address issues more widely and more completely than one method could, which in turn amplifies the richness and complexity of the research findings."¹⁹ A web survey of English and history faculty at 12 research universities was conducted in conjunction with interviews of 17 fine arts faculty at the same universities to gather in-depth and holistic responses on uses of digital collections by humanities scholars in multiple disciplines. The methodological approach drew upon the compara-

tive protocols used in research studies by Brockman et al., Bulger et al., and Palmer and Neumann of humanities scholarly research practices.²⁰ The validity of this study derives from its primarily qualitative orientation, and that qualitative research can collect valid data and generate valuable research insights from a purposefully created sample. The findings can be generalized in the sense of being “extrapolated in their relation to their theoretical application.”²¹ The criteria applied to build this study’s purposeful sample of humanities scholars consisted of restriction to teaching faculty members who were officially confirmed to be working in arts and humanities departments at recognized North American public and private research universities. As such, the carefully created sample for this study’s subject pool, as well as the methodological approach, is comparable to previously cited studies on humanities scholars’ practices.

The core data analysis focused on the open-ended survey responses and the interview responses; through the subsequent comparative analysis with focus groups’ responses from Phase One of the Bamboo Technology Project, this study seeks to document new insights into scholarly practices with and ongoing needs for digital collections.

Participants

The subject population was selected through a mix of random purposive sampling and snowball sampling: The targeted population consisted of humanities faculty members selected from 12 research universities that are members of the Committee for Institutional Cooperation (CIC) consortium. The 12 research universities are of representative size and breadth for comprehensive U.S. institutions of higher education, especially given that 7 of the universities are members of the American Association of Universities (AAU) and all institutions are members of the Association of Research Libraries (ARL).²² The faculty targeted for the survey came from the English and history departments, which are among the largest humanities departments in institutions of higher education, according to the 2012 National Study of Postsecondary Faculty.²³ The performing and fine arts department faculty were selected as the third humanities disciplinary area and were interviewed given their smaller numbers and distinctly different research and teaching practices.²⁴ The faculty members were identified via the publicly available departmental lists of teaching faculty and lecturers on university websites. The authors worked with a selection of librarians and academic technologists at the surveyed institutions to identify specific faculty members at each university who were recruited for their interest or expertise in using digital collections for research. Of the full list of all English and history faculty members at the 12 campuses, a random one-third was selected via the `RANDBETWEEN` formula in Excel to ensure an equally distributed random selection, and the sample then was manually adjusted to ensure representative distribution among gender and tenure status.

Survey

The survey contained a mix of multiple choice questions and open-ended responses, which enabled a richer set of responses on the research topic and further supported the multiple types of data collection best suited for an exploratory needs assessment. The survey instrument was developed in consultation with the University of Illinois’s Survey Research Lab, and the final version was created and distributed via Qualtrics software. The drafted survey was pretested with 10 selected scholars, academic technologists, and librarians from the surveyed institutions. The final version was distributed via Qualtrics web survey software and revised accordingly. The survey was distributed via direct individual e-mails to the identified random pool of one-third of the English and history faculty members at each institution and was conducted from October 2011 through February 2012. The survey was conducted with English and history faculty

due to the significant numbers of faculty in these disciplines on the surveyed campuses, and the similarity of their predominantly text-based research methods. The number of recruited subjects for the survey totaled 367 faculty and teaching lecturers, and the number of received responses was 75, resulting in a total cooperation rate for the survey of 20.4 percent. Of the respondents, 43 percent were tenured full professors, 29 percent were tenured associate professors, and 19 percent were tenure-track assistant professors, with the remaining single-digit percentages representing non-tenure-track lecturers, emeritus faculty, postdoctorates, and unknown.

The survey asked respondents to describe their research work with text-, image- and media-based digital collections, the functionalities that would improve each of these types of digital collections for scholarly research, and the benefits and disadvantages of digital materials for their own research agendas (see appendices A and B). Survey respondents were provided with a precise definition of digital collections as “curated collections of digital materials... that are accessed online or with computing software,” and then asked a series of questions on how they used digital collections.

Interviews

The responses to the survey informed the development of the interview protocol, which asked arts faculty from the same 12 research universities how they used digital materials from digital collections in their research and scholarly practices. Semistructured interviews were chosen as the second data collection instrument for the study, as they provided a “more holistic picture of people’s understandings than a conventional survey analysis would provide and elucidate the meanings that research participants attribute to their practices and actions.”²⁵ The interviews captured in-depth data by a set of purposely selected respondents that fit criteria similar to those of the survey respondents, with the difference residing in the discipline. The fine arts faculty subjects were selected as interview participants due to the smaller sample size and the anticipated differing responses due to their multimedia based research. The authors conducted these interviews from January through August 2012 via e-mail and telephone, and a random one-third of teaching lecturers and faculty members from fine arts departments were recruited for interviews via individual e-mails. The randomized sample enabled a relatively balanced selection of scholars with diverse experiences in using digital resources and tools, ranging from low technical competencies to scholars with almost fully digital workflows.

Data Analysis

This study originated as a subinvestigation for the Bamboo Technology Project (also known as “Project Bamboo”), a multi-institutional research initiative funded by the Andrew W. Mellon Foundation from 2008 through 2012 that sought to build an e-research platform for digital humanities research.²⁶ While this study initially sought to gather need requirements from scholars how collections should be prepared for use on humanities e-research platforms such as Project Bamboo, it rapidly acquired a broader goal of investigating how humanities scholars used digital materials in a variety of research environments and their needs for making effective use of digital corpora in scholarly research.

The gathered qualitative data—the open-ended survey responses and interview responses—were analyzed based on the grounded theory of coding, in which themes arise from the responses, rather than a hypothesis.²⁷ The open-ended survey responses and the interview responses were separately hand-coded for themes, with the two authors and a graduate research assistant conducting double coding of the open-ended survey responses and interview data to ensure intercoding reliability. The quantitative survey

responses were analyzed for basic percentages and averages; and, as the sampling was not scientifically probabilistic, the reported results are limited in their generalization.

These data, gathered from humanities scholars, reveals critical ways that the scholars incorporate digital materials into their research and the potential research impacts of enhancing functionalities of digital collections.

Results

The content of digital collections was a critical factor in survey responses about the use of digital collections: The most frequently used materials were texts at 100 percent and images at 94 percent, followed by maps at 58 percent, video at 42 percent, and audio at 39 percent. The respondents were then asked to identify the most useful functionalities for collections of digital materials per certain formats.

When asked to identify the most important needed functionalities for a digital collection of texts, images, and multimedia, the leading survey responses addressed the areas of metadata, content, and exportability, as displayed in table 1.

Text	Images	Multimedia
Highly Searchable Content	Downloading Capabilities	Detailed Metadata
Download and Export Functionality	Editing tools	Searchability
Detailed Metadata	Metadata	Download Capabilities
Quality of the Text	Usability and User-friendly Interfaces	Editing Tools
Breadth of Content	Searchability	

Features and functionalities in digital collections that survey respondents cited as being useful enough to “induce” them to use digital collections for research ranged in scope and type. While one survey respondent retorted “I don’t need to be induced,” the leading responses could be divided in the categories of download and export functionalities; searchability; increased quality and breadth of content; and interoperability of tools and content between collections.

Of the types of use with digital collections, the most frequent use was access to primary source material, followed by searching materials and verification of sources. One respondent described his/her research work as encompassing “every way conceivable: searching large collections, looking up and finding otherwise inaccessible articles.”

When broken down by academic status from postdoctorate to full professors, all levels used digital collections most frequently for access to primary sources and searching. But while full professors overwhelmingly used digital collections for searching and accessing materials, tenured associate professors also had notably high use of digital collections for research analysis.

One of the most prominent challenges cited by respondents in their use of digital collections was the inability to search effectively through the collection materials. As one respondent noted, “The ability to conduct corpus wide inquiries is still severely limited. Search tools are good for finding needles in haystacks, terrible for extracting data for subsequent manipulations.” Other notable challenges expressed by respondents were access to materials, copyright restrictions, and difficulty in downloading and

exporting materials. While the survey was focused on scholars who pursued heavily text-based scholarship, the interviews with fine arts scholars who use primarily image- and media-based collections generated similar responses for needs.

The interviews allowed for more detailed discussion of needs and use, and the interviewed scholars' leading needs for digital collections included:

- Better search functionality
- Need for annotation and editing tools
- Improvement in user interface design
- Expanded completeness of the collection's content

Scholars also were asked what were the benefits of originals and benefits of digital content. The stated benefits of original materials included completeness and immediate sensory interaction with the material. Yet even in their responses, the respondents alluded to the need to integrate the print/analog and digital. As one interviewee explained,

I still must and do consult originals, but for teaching I have always had to rely on surrogates, and for my research, the availability of databases particularly for periodicals, historical dictionaries, and other online primary and secondary sources generally has been an enormous boon.

As such, additional stated benefits of the digital included broader access to materials, portability in research environments that stretch across multiple locations and devices; preservation of rare objects; and the ways in which digital materials supported innovative teaching strategies.

The scholarly use of digital collections and challenges in using such collections resonate with another set of responses gathered from scholars, technologists, and librarians who participated in the Bamboo Technology Project workshops. An analysis of these qualitative data reveals similar uses and challenges for scholarly use of digital collections.

Bamboo Technology Project Workshops

From 2008 through 2009, a series of Bamboo Technology Project workshops brought together disciplinary faculty, librarians, academic technologists and students to discuss their uses of digital materials and needs for pursuing digital humanities scholarship. A comparative analysis of the anonymized qualitative data from these workshop focus groups and interviews, initially analyzed by Quinn Dombrowski in the *Project Bamboo Scholarly Report*, reveals interesting parallels with the survey and interviews on the needs of scholars.²⁸

Modes of scholarly practice emerged in the workshop responses as a key theme of how scholars framed their needs for digital resources. The Bamboo scholarly practices identified in the report are listed in table 2.

An analysis of the Bamboo workshop responses that generate these scholarly practices reveals complementary insights into the needs of scholars for digital collections.

Data Summary

For this paper, a sample of 1,055 anonymized responses from the Bamboo Technology Project workshops were analyzed in the same code rubric and analytical framework as the survey and interviews, using the approach of secondary supplementary analysis.²⁹ In this analysis, distinctive nuances emerged in the analysis of the Bamboo workshop responses when compared to the survey and interviews—perhaps due to the broader demographic of librarians, technologists, and scholars that participated

TABLE 2
Scholarly Practices as Identified by the Bamboo Technology Project

Gathering/Foraging	Synthesizing/Filtering
Contextualizing	Conceptualizing
Refining	Critiquing
Documenting Methods	Managing Data
Annotating/Documenting	Modeling/Visualizing
Teaching/Research	Sharing/Publishing
Funding	Collaborating
Citation/Credit/Peer Review	

in the workshops. Three core areas of need emerged: tools and analysis, content, and search/discovery and interoperability.

In the responses that mapped to the area of tools and analysis, expressed needs not only included analytical tools, but also there was repeated desire for educational infrastructure to teach usage of the tools and the methodologies of digital humanities research. As one respondent explained, “seamless access to the technology to quickly get into the heart of the research” and “easy usable tooling and support” would enable “a fertile playground to improve the research cycle.” Detailed metadata and sourcing also emerged as leading enhancements for digital collections, and the workshop responses also specified the need to know the data format and structures for exposing digital material to the larger scholarly community.

The needs and functionalities related to content encompassed not only the importance of a high quality of digital material, but it also included the incorporation of peer review as a guiding factor in the development of digital collections and as a method for evaluating collections’ contents. When discussing the need for expanded content, the respondents reflected on how digital collections needed to justify why the collection’s content would be useful for scholarship. As one respondent noted, “If anyone’s going to use anything besides databases as hyped up way of finding material, [there] has to be a convincing reason to answer intellectual question that you couldn’t answer otherwise.” The ability to export and download content included the ability to transfer material into the scholars’ own workspaces and having the capability to import the material into the digital tools for their particular research work. This exportability of digital content also highlighted the need for access and discovery in digital collections.

Search and discovery of content is core to the function of being able to use the content, and responses included the need to filter the immense amounts of content on the web to manageable browsing. Responses included suggestions such as “interface to the data that allows us to ‘go fishing,’” “interfaces that teach people how to search,” and “pop-up menus on the search bars to give users additional information.” The expressed needs for search capabilities also included being able to customize aggregations of content—as one respondent put it, “Isn’t it about tagging and bagging? Bagging into conceptual bins, maybe one object in multiple bins.” On a deeper level, the responses revealed that the enhancement of discovery and access functionalities in digital collections enables scholars to engage with data in new ways and integrate their new datasets into the research workflow. As one respondent noted, “They want to trace arguments through data. They are willing to make data available even if not published. Maybe in a decade’s time that these objects will be ‘published’ contribu-

tions." The potential future contributions of datasets also points to the importance of a community of scholars in digital scholarship.

The need for a collaborative research community around digital humanities research was another major theme that emerged from the responses. As one respondent noted, "We don't have a space where we can talk about the process in process.... [We] need a place where you can throw out, where the annealing starts, where the argument takes form." Another respondent argued that there was a need to "find collaborators, find & form invisible colleges, talking about/giving feedback about research, organize/share research, engage in thoughtful writing and conversation with colleagues, participate in several interest groups." Respondents envisioned that this collaborative environment could incorporate the teaching and student research that was a critical part of their experiences in digital scholarship. As one respondent explained, "Students using mixed media, we have to teach them how to do it, need machinery, way to share it and store it."

It becomes evident from the responses that digital content is the heart of this evolution in the nature of humanities scholarship, and the malleability of digital content is important to this change. As one respondent observed,

Disc space on any given computer is being filled with digital media (not text). What does the "replacement" of text as a primary area of focus mean to humanities scholarship? What tools are available for searching on, annotating, extracting from, publishing, etc.—these are at a less-developed stage than tools for text.

This fluidity in the definition and role of "text" in humanities scholarship is critical to librarians' assessment of digital collections. In light of these responses, how can we prepare our digital collections to reflect these new notions of content and new modes of scholarly synthesis in the humanities?

Discussion: Curation and Interoperability

In the analysis of these responses, two primary needs emerged: having sustained access and effective discovery mechanisms for digital collections, and the ability to mix and reuse digital materials. These needs map to the areas of digital curation and interoperability, which are two critical issues facing the development of digital collections to meet the research needs of today's humanities scholars.

Analysis: Digital Curation

In the humanities, data is a complex mixture of materials and types, as noted in the American Council of Learned Societies' 2006 report *Our Cultural Commonwealth*:

"The complexity of the record of human cultures—a record that is multilingual, historically specific, geographically dispersed, and often highly ambiguous in meaning—makes digitization difficult and expensive. Moreover, a critical mass of information is often necessary for understanding both the context and the specifics of an artifact or event, and this may include large collections of multimedia content: images, text, moving images, audio."³⁰

This diverse archive of digital material requires a multifaceted approach to digital curation. As Maron and Pickle note, "collaboration between the Special Collections department and the library IT team is critical to integrate the digital curation, preservation, and access activities of Special Collections into the institutions' overall digital infrastructure."³¹ The responses of scholarly users in the study attested to the need for digital curation.

Functionalities that enabled scholarly use and reuse of digital materials was prevalent in survey and interview responses, including detailed metadata, accessible textual and

image files, temporal coverage of content, transcriptions, the inclusion of nontextual sources in collections, and access to broader content. As a survey respondent noted, "The easier objects are to repurpose, remix, and reuse the better." In their Council of Library and Information Resources report on the work of Digging Into Data grant awardees, Williford and Henry note that "scholars often consider the resulting "clean" data to be just as important in potential impact as are the final research products the data make possible," and they cite Richard Healey's observation that "It has become clear that 'making data diggable' or providing 'keys' that unlock future digging potential may be just as important from a scholarly viewpoint, especially at this very early stage of the overall digging game."³²

The challenges of the curation of digital collections can be complex: In their examination of developing an all-digital research library, Spiro and Henry prioritize many of the digital curation needs, such as a critical mass of resources, user-friendly tools, resource interoperability, and reliable cyberinfrastructure, and note that "understanding data-based collaborations will have an impact on the design and development of digital library services and architectures."³³ Henry and Smith note that a major challenge in ensuring the quality of large-scale digital collections are the quality of the scanned content, "uneven collections," and missing content and recommend that libraries establish "standards of quality control, access policy, audit frequency, and terms of sustainability [that] are necessary to convince scholars that a digital object is an adequate, acceptable, and trusted surrogate for a printed book or article."³⁴

A few models exist for preparing and curating humanities digital collections: The 2003 *Sustaining Digital Scholarship* report published by the University of Virginia's Institute of Advanced Technology in the Humanities discusses curatorial methods for preparing digital humanities projects and digital collections for preservation and data curation and proposes multiple levels of preservation.³⁵ The Digital Curation Centre Digital Curation Lifecycle charts the processes that content providers should keep in mind as they curate digital collections, such as continuous Appraisal and Selection of the digital material for long-term curation; Ingest of data to a repository or data center as part of Preservation Action; and taking actions for enabling Access, Use, and Reuse of the digital collections' material and usable data. But beyond processes, another important aspect of digital curation is ascertaining the impact of the digital collections: Studies conducted by Ithaka S+R on the sustainability of digital collections observe that the most successful digital resources are ones that are "Creating Value"; as the *Sustaining Digital Resources* report observes,

The value these projects create relates to the extent to which they become important parts of the workflow of their user communities and the extent to which users rely on them to do their work... While the process of digitisation itself creates value for end-users, many projects go further, investing in tools and features to aid users in discovering and using the content in innovative ways.³⁶

As such, a critical component to the curation of digital collections is the input of users: Determining how users will make use of the collections is a cornerstone for a digital curation workflow. Maron and Pickle note in their study of digitized special collections in Association for Research Libraries institutions that "although the ability to offer greater access emerged as a key motivator for digitizing collections, investments in understanding the needs of the audiences are quite low."³⁷ They argue for libraries to invest in qualitative research of users, as "this will provide rich information about the value scholars and students place on the digitized collections."³⁸

Institutional programs that create and successfully sustain digital collections frequently have proactive user assessment and user engagement built into their work-

flows.³⁹ Indiana University's Digital Library Program, for example, offers a range of services to scholars in the planning stages of potential digital projects. Through user consultations, the collaborative teams of digital services librarians and faculty members incorporate user input into the project design, and then the resource is further subjected to usability testing and live research situations.⁴⁰ Among recent research output on institutional programs, the Toolkit for the Impact of Digitised Resources (TIDSR) that resulted from the studies by Meyers et al. provides critical strategies for institutions to employ in assessment of the impact of digital collections on users.⁴¹ Curation of digital materials is core to assessing and expanding the impact of digital collections on users, and one critical conduit for facilitating this use and impact is the enabling of interoperability of content across digital collections.

Analysis: Interoperability

Interoperability is defined in the *Oxford English Dictionary* as "the ability of two or more computer systems or pieces of software to exchange and subsequently make use of data."⁴² For digital libraries, metadata is the core conduit of facilitating interoperability, and Baca frames the functionality of metadata to access content across digital collections as "semantic interoperability," which is "the ability to search seamlessly for digital information across heterogeneous distributed databases as if they were all part of the same virtual repository."⁴³ The interoperability of metadata is core to the needs of humanities scholars to search across multiple digital collections and requires functionalities beyond tools for information organization and retrieval.

Users need digital collections that contain interoperable content or functionality that facilitates comparative analyses of digital materials. As observed by Brockman et al., the research practices of humanities scholars prominently includes the gathering of sources from multiple collections to create a customized corpus that enables them to explore particular research questions.⁴⁴ Henry and Smith acknowledge this issue, noting that "the challenge before scholars now is to make connections among and within huge sets of digitized data and to create new knowledge from them."⁴⁵

As such, digital collections need effective interoperability between content as well as metadata to support scholarly research, and the respondents in this study clearly expressed the need for such functionalities.

Robust search tools across multiple digital collections were a strongly expressed need among interview and survey respondents. Particularly valuable search functionalities included keyword searching, faceted searching, previewing of files, and general browsing of all types of materials. Responses also identified the need for comprehensive metadata in digital collections to enable comparative analysis of collections' content, particularly the identification of specific scholarly editions. The cross-collection use of digital materials results in remixing and reuse of materials for teaching and research, and one respondent explained that ideal digital collections provided means for "exporting files and creating my own text and visual files either for teaching or research purpose."

Baca notes that, while metadata is key and "a metadata standard appropriate to the materials in hand and the intended end-users must be selected," the process of integrating diverse information sources is an enduring challenge:

The dream of integrated access to diverse information resources is still just that—a dream. The dream can become a reality if those responsible for making cultural heritage information available online judiciously select and implement the appropriate metadata schemas, controlled subject vocabularies and thesauri, metadata crosswalks, and information technologies available to us today.⁴⁶

The needed resources to turn the “dream” of interoperability into reality requires collaboration not only across standards and technological infrastructure but also between the information experts who develop tools, curators, and the expert users who make use of cultural heritage digital material.

Discussion: Needs for Digital Collections

This study seeks to highlight the pivotal, deeper needs for enhanced functionality in digital collections that enable humanities scholarship, beginning with digital curation and interoperability as core issues. Libraries, museums, and other content providers of digital collections should reevaluate their approach to building and enhancing digital collections. A number of recent studies—notably the CLIR report *The Idea of Order: Transforming Research Collections for 21st Century Scholarship*—argue for a dramatic shift in research libraries’ conceptualization of collections.⁴⁷ This shift is marked by an active, user-centered perspective toward both digital and physical collections. In particular, the principle of contextual mass, which prioritizes users’ research practices in determining collection content, is more imperative than ever in the development of digital collections.⁴⁸ Thus, as scholarly users demand greater functionality and reliability in digital collections, it is critical that libraries anchor the scope and functionality of their collections in the needs of users, and in doing so, begin establishing deeper research partnerships with users.

This also necessitates that libraries redefine their roles in the campus ecosystem in broader terms that encompasses active integration in the full research lifecycle from creation to curation. Emerging initiatives in academic libraries in research data management, open access publishing, and information literacy seek to embed librarians into the research and teaching workflows, and digital collections can play a critical conduit in helping libraries collaborate with their users. Walters and Skinner note that research libraries must be “repositioned as vibrant knowledge branches that reach throughout their campuses to provide curatorial guidance and expertise for digital content” and highlight digital humanities projects as an area where libraries have “roles to play in helping these initiatives to produce and to manage sustainable resources for present and future generations.”⁴⁹ As such, the reconfiguration of digital collections to reflect the research needs of humanities scholarship is a critical aspect of librarians’ embedment in the research workflows of students and scholars alike.

The results of this study suggest that engaging with users is an important strategy in enhancing digital collections: Libraries and content providers must reevaluate their approach to building and expanding digital collections with an intensely user-centered focus. Humanities scholars are using digital collections in complex ways: not simply to find the existence of a document or to supplement images for an article but as primary source material, as analytical tools, and as resources that broaden the scope and depth of their analysis and scholarly knowledge. And ultimately, we must consider how scholars’ needs should shape the development of humanities cyberinfrastructure, as the ALCS’s *Our Cultural Commonwealth* report argues that “extensive and reusable digital collections are at the core of the humanities and social science cyberinfrastructure. Scholars must be engaged in the development of these collections.”⁵⁰ Thus, to support e-research environments for the humanities, libraries must undertake collaborative investment in the functionalities, infrastructures, and content for digital collections.

Collaboration is essential both internal to the library and outwardly with the scholarly community: Librarians can develop innovative and multipronged methods of user engagement through collaborations between specialists such as digital services librarians, subject specialists, and metadata librarians that ensure the broadest outreach to scholarly communities that use their digital collections to produce scholarship. If

libraries and content providers are to provide digital collections for digital humanities research, continuous dialogue with humanities scholars on their research practices and needs is critical to enabling humanities datasets that are powerful enough for the new types of analyses being done today.

Conclusion

This study presents an initial examination of the needs of humanities researchers when they are using digital collections. While there are vast differences among the scholarly needs of individual researchers, this study begins to reveal that libraries must work to build collections that exist in a sustainable, networked, and iterative environment and that the content is responsive to the evolving needs for digital humanities research. Questions asked of researchers focused not only on how they use existing resources, but also, importantly, how features and functions would better enable effective future use.

Librarians have rich opportunities to insert themselves into the evolving, increasingly digitally oriented research workflows of humanities students and scholars through their existing and emergent expertise in information retrieval, digital curation, and facilitating the interoperability of resources for discovery and access—all areas that surveyed scholars indicated as important to effective research use of digital collections. Librarians should reach out not only to the scholars who are working with digital materials, but build lines of communication with a wide breadth of researchers to assess how their digital collections can be more effectively discovered, accessed, and used for research work. As digital humanities initiatives and the resultant research data proliferate exponentially, the demand will increase for content providers such as libraries and museums to collaborate with humanities scholars when building and curating digital collections.

Acknowledgments

Special thanks go to Nicole Saylor for her collaborative research work on earlier versions of this study. The authors also thank graduate assistant Elizabeth Gonzalez for her assistance with data analysis. The authors also would like to thank the following colleagues for their assistance with the protocol development and subject recruitment for the study: Kate Brooks, Bill Brockman, Charlotte Cabbage, Timothy Cole, John Cullars, Quinn Dombrowski, Michael Hancher, Patricia Hswe, Trevor Muñoz, Michael Rodriguez, Carrie Roy, Mary Stuart, John Unsworth, Kay Walters, and Ethan Watrall.

Appendix A: Survey Questions

Survey was distributed via Qualtrics software.

1. Digital collections are curated collections of digital materials such as digitized texts, digital images, digitized maps, and/or multimedia objects that are accessed online or with computing software. Do you use materials from digital collections in your research and scholarship, as distinct from your teaching?
 - Yes (1)
 - No (2)

2. Materials from digital collections can cover a range of topics and resources from the specialized William Blake Archive (<http://www.blakearchive.org/blake/>) to the Library of Congress's American Memory collections (<http://memory.loc.gov/ammem/index.html>). Have you ever used any of the following types of digital materials from digital collections in your research and scholarship, apart from your teaching? (Please select all that apply.)
 - None (1)
 - Texts (2)
 - Images (3)
 - Audio (4)
 - Video (5)
 - Maps (6)
 - Other: (7) _____

3. How often do you use digital materials from digital collections in your scholarship, regardless of whether a print/analog version is available?
 - Never (1)
 - Less than half of the time (2)
 - Equally print and digital resources (3)
 - More than half of the time (4)
 - Always (5)

4. Please list at least three (3) important attributes for a digital collection of TEXTS to have for the collection to be an effective research resource (that is, capability to export files, detailed metadata, and so on).

5. Please list at least three (3) important attributes for a digital collection of IMAGES to have for the collection to be an effective research resource (that is, viewing and zooming tools, ability to tag files, and so on).

6. Please list at least three (3) important attributes for a digital collection of MULTIFORMAT MEDIA to have for it to be an effective research resource (that is, annotation tools, detailed metadata, and so on).

7. What are at least two (2) tools or functionality features that would induce you to use digital collections more frequently in your research? (that is, interoperability with other digital collections, ability to export files, and so on)

8. How do you specifically use digital materials from digital collections in your research and scholarship, as distinct from your teaching?

9. What challenges, if any, have you encountered because of your use of digital collections in your scholarship?
10. Do you have any additional comments about humanities scholarly research with digital collections?
11. Why do you not use digital collections in your scholarship and research? (If answered "No" to the second question)
12. May we contact you for a follow-up interview?
 - Yes (1)
 - No (2)

Appendix B: Interview Questions

1. In your words, what are the primary characteristics of "corpora" for humanities research?
2. What functionalities and features would compel you to use digital collections more frequently in your research?
3. Why would these features improve your research experience with digital collections?
4. How do you specifically use digital materials from digital collections in your research? With what frequency?
5. What are the benefits of digital collections for your research?
6. What are the benefits of using a print object over its digital surrogate or another alternative?
7. What challenges have you encountered because of your use of digital collections in research?
8. How do you foresee humanities research changing in light of the rapid increase in digital collections and access?

**Additional questions may be asked depending on the respondents' answers during the interview.

Notes

1. Carole Palmer, "Thematic Research Collections," in *A Companion to Digital Humanities*, eds. Susan Schreibman, Ray Siemens, and John Unsworth (Oxford: Blackwell, 2004), available online at www.digitalhumanities.org/companion/view?docId=blackwell/9781405103213/9781405103213.xml&chunk.id=ss1-4-5&toc.depth=1&toc.id=ss1-4-5&brand=default [accessed 22 May 2015]; Carole Palmer, Oksana Zavalina, and Katrina Fenlon, "Beyond Size and Search: Building Contextual Mass in Digital Aggregations for Scholarly Use," *Proceedings of the American Society for Information Science and Technology* 47, no. 1 (2010): 2.
2. Matthew B. Gilmore and Donald O. Case, "Historians, Books, Computers, and the Library," *Library Trends* 40, no. 4 (1992): 667–86, available online at https://www.ideals.illinois.edu/bitstream/handle/2142/7802/librarytrendsv40i4g_opt.pdf?sequence=1 [accessed 18 June 2014].
3. Wendy M. Duff and Joan M. Cherry, "Use of Historical Documents in a Digital World: Comparisons with Original Materials and Microfiche," *Information Research* 6, no. 1 (2000), available online at <http://informationr.net/ir/6-1/paper86.html> [accessed 22 May 2015].
4. Marcia J. Bates, Deborah N. Wilde, and Susan L. Siegfried, "An Analysis of Search Terminology Used by Humanities Scholars: the Getty Online Searching Project Report Number 1," *Library Quarterly* 63, no. 1 (1993): 1–39; Susan L. Siegfried, Marcia J. Bates, and Deborah N. Wilde, "A Profile of End-User Searching Behavior by Humanities Scholars: The Getty Online Searching Project Report No. 2," *Journal of the American Society for Information Science* 44, no. 5 (1993): 273–91; Marcia J. Bates, "The Design of Databases and Other Information Resources for Humanities Scholars: The Getty Online Searching Project Report No. 4," *Online & CDROM Review* 18, no. 6 (1994): 331–40; Marcia J. Bates, Deborah N. Wilde, and Susan L. Siegfried, "Research Practices of Humanities Scholars in an Online Environment: The Getty Online Searching Project Report No. 3," *Library & Information Science Research* 17 (1995): 5–40; Marcia J. Bates, "Document Familiarity, Relevance, and Bradford's Law: The Getty Online Searching Project Report No. 5," *Information Processing & Management* 32, no. 6 (1996): 697–707; Marcia J. Bates, "The Getty End-User Online Searching Project in the Humanities: Report No. 6: Overview and Conclusions," *College & Research Libraries* 57, no. 6 (1996): 514–23.
5. Siegfried, Bates, and Wilde, "A Profile of End-User Searching Behavior by Humanities Scholars," 288.
6. William S. Brockman et al., *Scholarly Work in the Humanities and the Evolving Information Environment*, CLIR Publication No. 104 (Washington, D.C.: Digital Library Federation and Council on Library and Information Resources, 2001), available online at www.clir.org/pubs/reports/pub104/contents.html [accessed 18 June 2014].
7. Lisa Spiro and Jane Segal, "The Impact of Digital Resources on Humanities Research," available online at <http://library.rice.edu/services/dmc/about/projects/the-impact-of-digital-resources-on-humanities-research> [accessed 22 May 2015].
8. Claire Warwick et al., "If You Build It Will They Come? The LAIRAH Study: Quantifying the Use of Online Resources in the Arts and Humanities through Statistical Analysis of User Log Data," *Literary and Linguistic Computing* 23, no. 1 (2008): 85–102.
9. Suzana Sukovic, "Convergent Flows: Humanities Scholars and Their Interactions with Electronic Texts," *Library Quarterly* 78, no. 3 (2008): 263–84.
10. Eric T. Meyer, *Splashes and Ripples: Synthesizing the Evidence on the Impacts of Digital Resources*, Joint Information Systems Committee Report (London: JISC, 2011), available online at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1846535 [accessed 22 May 2015]; Eric T. Meyer, Kathryn Eccles, and Christine Madsen, "Digitisation as e-Research Infrastructure: Access to Materials and Research Capabilities in the Humanities" (paper presented at the 5th International Conference on e-Social Science, Cologne, Germany, June 2009), available online at http://www.researchgate.net/profile/Eric_Meyer/publication/255658936_Digitisation_as_e-Research_infrastructure_Access_to_materials_and_research_capabilities_in_the_Humanities/links/00b7d53847b7cfbcfbf000000.pdf [accessed 22 May 2015].
11. Merrilee Proffitt and Jennifer Schaffner, *The Impact of Digitizing Special Collections on Teaching and Scholarship: Reflections on a Symposium about Digitization and the Humanities* (Dublin, Ohio: OCLC Programs and Research, 2008), available online at <http://www.oclc.org/content/dam/research/publications/library/2008/2008-04.pdf?url=162913> [accessed 22 May 2015].
12. Donghee Sinn, "Impact of Digital Archival Collections on Historical Research," *Journal for the American Society of Information Science and Technology* 63, no. 8 (2012): 1521–37.
13. Janet Gertz, "Should You? May You? Can You? Factors in Selecting Rare Books and Special Collections for Digitization," *Computers in Libraries* 33, no. 2 (2013): 6–11.
14. Diane Zorich, *Transitioning to a Digital World: Art History, Its Research Centers, and Digital Scholarship* (New York: Samuel H. Kress Foundation and Roy Rosenzweig Center for History and

New Media, 2012), available online at www.kressfoundation.org/research/Default.aspx?id=35379 [accessed 22 May 2015].

15. Monica Bulger et al., *Reinventing Research? Information Practices in the Humanities* (London: Research Information Network, 2011): 16.

16. *Ibid.*, 8.

17. *Ibid.*, 14–15, 71.

18. Charles Teddlie and Abbas Tashakkori, "Major Issues and Controversies in the Use of Mixed Methods in the Social and Behavioral Sciences," in *Handbook of Mixed Methods in Social & Behavioral Research*, eds. Abbas Tashakkori and Charles Teddlie (Thousand Oaks, Calif.: Sage Publications, 2003), 16.

19. Raya Fidel, "Are We There Yet? Mixed Methods Research in Library and Information Science," *Library and Information Science Research* 30, no. 4 (2008): 266.

20. Brockman et al., *Scholarly Work in the Humanities and the Evolving Information Environment*, 36; Bulger et al., *Reinventing Research?* 16; Carole L. Palmer and Laura J. Neumann, "The Information Work of Interdisciplinary Humanities Scholars: Exploration and Translation," *Library Quarterly* 72, no. 1 (2002): 91.

21. Robert Grover and Jack Glazer, "Implications for Application of Qualitative Methods to Library and Information Science Research," *Library and Information Science Research* 7 (1985): 252; Julia Brannen, "Combining Qualitative and Quantitative Approaches: An Overview," in *Mixing Methods: Qualitative and Quantitative Research*, ed. Julia Brannen (Brookfield: Avebury, 1992), 9.

22. Surveyed universities included: Indiana University–Bloomington; Michigan State University; Northwestern University; Pennsylvania State University; University of Chicago; University of Illinois at Chicago; University of Illinois at Urbana–Champaign; University of Iowa; University of Maryland–College Park; University of Minnesota; University of Nebraska at Lincoln; and the University of Wisconsin–Madison.

23. American Academy of Arts and Sciences, *Humanities Indicators*, available online at www.humanitiesindicators.org/content/hrcolIID.aspx [accessed 22 May 2015].

24. *Ibid.*

25. Julia Brannen, "Mixing Methods: The Entry of Qualitative and Quantitative Approaches into the Research Process," *International Journal of Social Research Methodology* 8, no. 3 (2005): 175.

26. Harriett Green, "Humanist Scholars' Use of Digital Materials," Project Bamboo Documentation wiki, last modified April 11, 2013, available online at <https://wikihub.berkeley.edu/display/pbamboo/Humanist+Scholars%27+Use+of+Digital+Materials> [accessed 24 January 2014].

27. Barney Glaser and Anselm Strauss, *The Discovery of Grounded Theory: Strategies for Qualitative Research* (New York: Aldine De Gruyter, 1967); Juliet Corbin and Anselm Strauss, *Basics of Qualitative Research (3rd ed.): Techniques and Procedures for Developing Grounded Theory* (Thousand Oaks, Calif.: SAGE Publications, 2008): doi:<http://dx.doi.org/10.4135/9781452230153>.

28. *Project Bamboo Scholarly Practices Report*, last modified March 8, 2013, available online at <https://wikihub.berkeley.edu/display/pbamboo/Project+Bamboo+Scholarly+Practice+Report> [accessed 16 December 2013].

29. Janet Heaton, "Secondary Analysis of Qualitative Data," in *The SAGE Handbook of Social Science Research Methods*, eds. Pertti Alasuutari, Leonard Bickman, and Julia Brannen (New York: Sage, 2008), 506–20.

30. American Council of Learned Societies, *Our Cultural Commonwealth: The Report of the ACLS Commission on Cyberinfrastructure for the Humanities and Social Sciences* (New York: American Council of Learned Societies, 2006), available online at <http://www.acls.org/cyberinfrastructure/ourculturalcommonwealth.pdf> [accessed 22 May 2015].

31. Nancy L. Maron and Sarah Pickle, *Appraising Our Digital Investment: Sustainability of Digitized Special Collections in ARL Libraries* (New York: Ithaca S+R, 2013), 3.

32. Christa Williford and Charles Henry, *One Culture, Computational Intensive Research in the Humanities and Social Sciences: A Report on the Experiences of First Respondents to the Digging Into Data Challenge* (Washington D.C.: Council of Library and Information Resources, 2012), 25.

33. Lisa Spiro and Geneva Henry, "Can a Research Library Be All-Digital?" in *The Idea of Order: Transforming Research Collections for 21st Century Scholarship*, CLIR Publication No. 147 (Washington, D.C.: Council of Library and Information Resources, June 2010), 45.

34. Charles Henry and Kathlin Smith, "Ghostlier Demarcations: Large-Scale Text-Digitization Projects and Their Utility for Contemporary Humanities Scholarship," in *The Idea of Order: Transforming Research Collections for 21st Century Scholarship*, ed. Council of Library and Information Resources, CLIR Publication No. 147 (Washington, D.C.: Council of Library and Information Resources, June 2010), 113–114.

35. Institute of Advanced Technology in the Humanities, *Sustaining Digital Scholarship Final Report* (Charlottesville, Va.: University of Virginia, 2004), available online at http://www.digital-curationservices.org/files/2012/05/SDS_FinalReport2003.pdf [accessed 22 May 2015].

36. Nancy L. Maron, K. Kirby Smith, and Matthew Loy, *Sustaining Digital Resources: An On-Ground View of Projects Today* (London: JISC, July 2009), 15.

37. Nancy L. Maron and Sarah Pickle, *Appraising Our Digital Investment: Sustainability of Digitized Special Collections in ARL Libraries* (New York: Ithaka S+R and Association of Research Libraries, 2013), 10.

38. Maron and Pickle, *Appraising Our Digital Investment*, 2–3.

39. Nancy L. Maron and Sarah Pickle, *Searching for Sustainability: Strategies from Eight Digitized Special Collections* (New York: Ithaka S+R, 2013), 7.

40. Indiana University Libraries Digital Projects & Services, "Interface Design and Usability Services," available online at <http://libraries.iub.edu/library-technologies> [accessed 22 May 2015].

41. Eric T. Meyer, *Splashes and Ripples: Synthesizing the Evidence on the Impacts of Digital Resources*, Joint Information Systems Committee Report (London: JISC, 2011), available online at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1846535 [accessed 22 May 2015]; Meyer, Eccles, and Madsen, "Digitisation as e-Research Infrastructure," 2009.

42. Oxford English Dictionary, s.v. "interoperability," available online at www.oed.com.proxy2.library.illinois.edu/view/Entry/248420 [accessed 6 January 2014].

43. Murtha Baca, "Practical Issues in Applying Metadata Schemas and Controlled Vocabularies to Cultural Heritage Information," *Cataloging and Classification Quarterly* 36, no. 3–4 (2003): 49.

44. Brockman et al., *Scholarly Work in the Humanities and the Evolving Information Environment*, 17–18.

45. Henry and Smith, "Ghostlier Demarcations," 108.

46. Baca, "Practical Issues in Applying Metadata Schemas and Controlled Vocabularies to Cultural Heritage Information," 54.

47. Council of Library and Information Resources, *The Idea of Order: Transforming Research Collections for 21st Century Scholarship*, CLIR Publication No. 147 (Washington, D.C.: Council of Library and Information Resources, 2010), available online at www.clir.org/pubs/reports/pub147/reports/pub147/pub147.pdf [accessed 22 May 2015].

48. Palmer, Zavalina, and Fenlon, "Beyond Size and Search," 3.

49. Tyler Walters and Katherine Skinner, *New Roles for New Times: Digital Curation for Preservation* (Washington, D.C.: Association for Research Libraries, 2011), 5, 71.

50. American Council of Learned Societies, *Our Cultural Commonwealth*, 38.