“A Greatly Unexplored Area”: Digital Curation and Innovation in the Digital Humanities

Alex H. Poole
Department of Information Science
College of Computing and Informatics
Drexel University
Rush Building
30 North 33rd Street
Philadelphia, PA 19104
Ahp56@drexel.edu
Abstract

New forms of digital data, tools, and methods, for instance those that cross academic disciplines and domains, those that feature teams of scholars instead of single scholars, and those that involve individuals from outside the academy, can enable new forms of scholarship and teaching in the digital humanities. Such scholarship can promote reuse of digital data, provoke new research questions, and cultivate new audiences. Digital curation, the process of managing a trusted body of information for current and future use, can help maximize the value of research in the digital humanities.

Predicated upon semi-structured interviews, this naturalistic case study explores the creation, use, storage, and planned reuse of data by 45 interviewees involved with nineteen Office of Digital Humanities Start-Up Grant (SUG) projects. The study seeks to determine what digital curation skills had been employed in these projects. Interviewees grapple with challenges surrounding data, collaboration and communication, planning and project management, awareness and outreach, resources, and technology. Overall, this study hopes to understand the existing digital curation practices and needs of those engaged in innovative digital humanities work and to discern how closely these practices and needs align with the digital curation literature.
Data seems to be the word of the moment for scholarship. 
-Owens (2011)

How do you know what you don’t know? 

Introduction

The American Council of Learned Societies (2006) contends, “The emergence of the Internet has transformed the practice of the humanities and social sciences—more slowly than some may have hoped, but more profoundly than others may have expected” (p. 1). New forms of digital data, tools, and methods can enable new forms of scholarship and teaching, for instance those that cross academic disciplines and domains, those that feature teams instead of single scholars, and those that involve individuals from outside the academy. Such scholarship and teaching can promote reuse of digital data, provoke new research questions, and cultivate new audiences. Both digital curation and digital humanities work can play a central role in these explorations.

Digital curation in the digital humanities typically focuses on a range of work products such as scholarly editions, text corpora, marked-up text, thematic research collections, data accompanied by annotation or analysis, and finding aids or information maps such as bibliographies (Flanders & Munoz, 2012). Tasks may include translating or migrating data into new formats, adding contextual information or markup, or linking datasets (Flanders & Munoz, 2012). Digital curation in the digital humanities as in the sciences helps ensure long-term access, facilitates discovery, retrieval, use, and reuse, and maximizes the usefulness of the curated digital content (Harvey, 2010). By making data as functional as possible digital curation can enable better research. After all, data can be interrogated using the same sorts of questions traditionally applied to texts and other humanistic foci of inquiry (Owens, 2011).

Digital curation and the digital humanities share concerns, practices, and objectives. First, both digital curation professionals and digital humanists add value to digital assets: they stimulate reuse, encourage the development of new research questions, and attract new audiences. Second, both areas rely upon interdisciplinarity and collaboration. Third, work in both areas often depends upon short-term, project-based grant funds. Fourth, digital curation and digital humanities struggle with similar challenges: sustainability, project management, institutional position, and the valuation of their work as scholarship. Finally, digital curation and digital humanities are both areas of practice and research.

By dint of its richness, humanities data offers substantial challenges for digital curation (Henry, 2014). Digital curation researchers in this area need a holistic grasp not only of present workflows, but also of stakeholders’ opinions regarding data’s future fit-for-purpose (Palmer, Tefteau, & Pirmann, 2009). Therefore, this study illuminates the existing digital curation practices of and challenges faced by those engaged in innovative digital humanities work. This study seeks “a nuanced, expressive and information-rich understanding of scholarly practices and needs” (Benardou, Constantopoulos, & Dallas, 2013, p. 106). It centers on the following research questions:

- What types of data have digital humanists (whether faculty, “alternative-academics” [1], (alt-acs) or graduate students) created, reused, stored, and planned to reuse in their SUG project?
• What challenges did they face and what lessons did they learn during their project work?

These questions are addressed by an examination of selected awardees of Start-Up Grants from the National Endowment for the Humanities’s Office of Digital Humanities.

The Office of Digital Humanities

The National Endowment for the Humanities’s Office of Digital Humanities (ODH) Start-Up Grants exemplify innovative digital humanities research that incorporates digital curation. ODH began offering Start-Up Grants of the “high risk, high reward” variety in 2007; its head Brett Bobley (2007) remarks, “I chose the name ‘start-up grant’ because it reminded me of the tech world, a tech start-up like the two Apple Computer guys in their garage.”

The ODH sponsors two types of Start-Up Grants, each of which may run up to eighteen months in duration. Disbursing up to $30,000, Level I grants underwrite brainstorming sessions, workshops or conferences, alpha-level prototypes, or initial planning. Likely outcomes include reports, position papers, or plans for future work. Level II grants support more mature projects, generally those that prepare for implementation or that show proof of concept. Receiving up to $60,000, these grants produce more concrete deliverables than Level I grants, for instance prototypes, test beds, or demonstration projects. Between the first round of awards in 2006 and the ninth round in 2014, 2,057 scholarly teams have applied for Start-Up Grants and 280 (13.6%) have garnered awards.

In 2011, the ODH began to mandate the inclusion of a data management plan in each application. Like other funding agencies, the ODH increasingly stresses preservation over ephemerality vis-à-vis its projects, especially of their data (Reed, 2014). The required Data Management Plans (DMPs) focus on the types of data the project will produce and on the grantee’s plan for managing that data both during and after the grant. The ODH defines data broadly as project materials generated or collected such as citations, code, algorithms, digital tools, documentation, databases, geospatial coordinates, reports, and articles. The ODH expects grantees to disseminate their data as well as their findings to scholarly and public audiences.

Humanists increasingly realize that they create and reuse “data” (though consensus on what “data” are in the digital humanities remains frustratingly elusive, the ODH’s definition notwithstanding) and thus can profit by making this data sharable and reusable, especially when their work depends upon public monies. Despite the promise of Start-Up Grants, however, practical and methodological challenges abound. The full scholarly potential of work in the digital humanities may depend upon digital curation. The Start-Up Grant projects examined in this study illustrate the potential for communication, coordination, and collaboration among digital curation professionals and digital humanists. They provide a useful way of studying the practices and attitudes of key stakeholders.

Method

Interviews were conducted with 45 persons. (See Appendix A for interview questionnaire.) They worked in a variety of departments and campus units and had similarly diverse job titles (Tables 1 and 2).

<table>
<thead>
<tr>
<th>Department/Campus Unit</th>
<th>Number of Interviewees</th>
</tr>
</thead>
<tbody>
<tr>
<td>University Library</td>
<td>6</td>
</tr>
<tr>
<td>Computer Science</td>
<td>5</td>
</tr>
<tr>
<td>Center or Institute</td>
<td>4</td>
</tr>
<tr>
<td>English</td>
<td>4</td>
</tr>
</tbody>
</table>
This study employs a naturalistic approach (Lincoln & Guba, 1985). A case study (Yin, 2009), it is predicated upon semistructured interviews (Rubin & Rubin, 2005; Weiss, 1994). (The study questionnaire is available in Appendix A.) Further, it relies upon purposive (specifically, snowball) sampling (Atkinson & Flint, 2004). Finally, in its coding and interpretation strategies the study borrows heavily from Constructivist Grounded Theory (Charmaz, 2014; Corbin & Strauss, 1990; Saldana, 2013).

To maximize the likelihood of interviewee recall, this study centered on the 23 SUGs that began in May of 2013 and were slated to be completed by the end of 2014 (no-cost extensions notwithstanding). The Principal Investigators (34 in total) of these projects were approached as initial contacts. Representatives from nineteen of those 23 projects agreed to participate. Each initial participant was asked for referrals to other project stakeholders. PIs most frequently referred me to their graduate students, though some referred me to librarians or to others involved in the project. Forty-five scholars (44 based at institutions in the United States) were interviewed. Interviews (they ranged between 30 and 90 minutes in duration) took place between August and November of 2014 using the telephone and Skype. (See Appendix B for details on the SUG projects examined.)

Through my interviews with PIs, I gained a top-down—“five thousand foot level” view, as interviewee 3, a PI, put it—of each project. By the same token, I also received an “in the trenches” perspective that this same PI’s Project Manager (interviewee 4) offered.
Findings

Findings center on the digital curation practices adopted by project team members. More specifically, interviewees discussed the types of data they reused, created, stored, and intended to reuse for future projects.

Reuse of Previously-Generated Data

Reusability constitutes a central concern of digital humanities work (Hockey, 2006). Seventeen of the nineteen projects (89.5%) reused existing data, whether their own or others.’ Of those seventeen projects, nine relied upon web-based information available in the public domain. Thus they implicitly trusted the quality of this information and assumed that they or others could revisit it subsequently. Three projects reused data previously generated by project personnel: survey materials, grant application materials, and field data. A final category of data reuse was more nebulous. Project personnel in these cases reused metadata gleaned from third-party institutions, film prints and ephemera, a video game and its associated materials, original literary texts, and data collected from a historical society. Hence those responsible for curating such data need a range not only of technological skills, but also of media-specific and domain skills.

Data Created

In the ODH’s definition of the term, all nineteen projects created data. Notwithstanding the mandatory White Paper and interim and final reports, the projects created at least 36 different types of data: heterogeneity appeared the norm. Certain types of data were created more frequently than other types, however. Notably, the two most frequently generated types of data were conventional (peer-reviewed) publications and presentations at professional events. Nonetheless, six projects generated code and five generated a database of some kind. Four projects generated conference or meeting records, while three projects produced software, blog posts, datasets, or models. Finally, data created by two projects included algorithms, curriculum materials, documentation, GIS or map files, photographic images, interfaces, metadata, or videos. Based on this sample, it seems that a paradigm shift toward the creation of less traditional types of data has yet to occur, even among innovative digital humanities scholars. Thus as Evans and Rees (2012) insist, digital humanists need to bolster the impact of their “wonderful whizz-bangy tools” with peer-reviewed research articles in key journals (p. 186).

Data Management Plans (DMPs)

Interviewees’ attitude toward and involvement in the ODH’s data management plan requirement ranged considerably. Despite the ODH’s mandate that each applicant include a DMP, one of the nineteen projects neglected to do so, an omission the ODH reviewers overlooked. Personnel from the other eighteen projects expressed widely varying views on the process and its value. One PI (interviewee 18) found the process of writing a DMP “really annoying” and did the “minimum.” Another PI (interviewee 38) characterized her effort as not “the most shining example of the genre.” Interviewee 33 suggested that many DMPs seem “somewhat boiler plate.” Persons from three other projects (interviewees 5, 12, and 25) struggled even to remember the process of writing their SUG DMPs.

Other project personnel felt differently about data management planning. One PI (interviewee 29) called the development of the DMP a “big process.” Another (interviewee 42) thought the DMP the “hard part.” She and her colleagues were “still feeling our way.” Perhaps most striking, interviewee 43 characterized the DMP process as akin to being “in a box trying to describe another box.” Despite the
challenges inhering in the DMP process, however, only one PI contacted the ODH for input or feedback.

Conversely, some other projects’ personnel found the DMP process easier. Interviewee 14 called the process “quite simple and straightforward.” Two projects used DMPTool, which provides a structured environment for data management planning and links directly to funders’ stipulations [3]. As a result, interviewee 9 labeled the development of her team’s DMP a “complete no-brainer.” The other persons whose project used DMPTool (interviewee 44) also reused materials from a previously submitted National Endowment for the Humanities grant. Finally, interviewee 3 reused material developed for an Andrew Mellon Foundation grant.

According to personnel from the Office of Digital Humanities, grantees’ DMPs have improved since 2011. Early plans seemed “written in a vacuum, or from a template” (interviewee 35). Interviewee 35 noted that recent DMPs are “much more consistent” and applicants display “at least a baseline sense of what sorts of questions they need to answer.” After four years, moreover, the ODH has numerous examples to share with applicants. “A weak data management plan,” interviewee 25 concluded, “often is a signal of a broader weak work plan or weak project conception as well.”

The ODH continues to educate its constituents about the DMP process. One ODH official (interviewee 36) noted, “We don’t think of [the DMP] as another hoop for them to jump through...It leads to stronger, more sustainable projects.”

Institutional Storage of Data

First, in their data management plans the vast majority of projects (94.8%) made provisions for storing their data institutionally. Second, representatives from ten projects (52.6%) mentioned storing part or all of their project’s data on a personal computer (or computers) during the project. Personnel often stored the data they had created or worked on, not necessarily the entire project’s data. This storage was generally ad hoc; interviewees made no mention of whether they would archive this data or would allow it to fall prey to benign neglect. This worrisome finding mirrors that of Carlson and Stowell-Bracke (2013), that of Jahnke and Asher (2012), and that of (Piorun, Kafel, Leger-Hornby, Najafi, and Martin, 2012). Third, personnel from eight of the nineteen projects (42.1%) stored or planned to store data on GitHub. Yet interviewees who resorted to this method apparently gave little thought to possible sustainability or other issues regarding GitHub. Fourth and finally, project personnel mentioned storage spaces such as Omeka, Zotero, ArcGIS, Scalar, YouTube, HathiTrust, commercial providers, and the cloud. Again, they seemingly devoted little thought to the sustainability of these of these entities.

Even those persons who planned to store their data in institutional spaces were not always aware of relevant institutional policies governing data retention. Those who were aware, moreover, did not always know the length of retention to which their institution committed. Institutions committed to specific time periods did so for amounts of time ranging from three years to ten years. Some institutions meanwhile committed to preserving the data for the life of the project, while others committed “in perpetuity” or “permanently,” though it remained unclear what those commitments might mean in actuality. Data storage and backup provisions seem at best a patchwork. As the work of Alexogiannopoulos, McKenney, and Pickton (2010) and Steinhart, Chen, Arguillas, Dietrich, and Kramer (2012) suggests, without oversight researchers are likely to store and backup...
their data in ways that will not maximize the data’s sharing or reuse potential. Funders may have an important role to play in encouraging researchers to keep their data fit for future use.

**Plan to Reuse Project Data**

In accord with ODH policy, all projects committed to sharing their data barring confidentiality or privacy issues. But the vast majority of projects had not at the time of their personnel’s interviewing produced sharable data; what was more, none knew of any other person(s) reusing their data. Therefore, a crucial future priority should be tracking reuse over time.

Nearly two-thirds of the interviewees (68.3%) planned to reuse project data themselves. Interviewee 43 characterized reuse as the project’s central goal—the team hoped to “do it over, better.” Similarly, interviewee 27 enthused, “oh God, yes,” and claimed it would be foolish not to reuse her data. Overall, despite their professed interest in and optimism toward reusing their data, interviewees’ plans for reuse tended to remain speculative. The ODH requires only that data be made available and in a format conducive to sharing; it does not require applicants to spell out planned or possible future reuse(s) by themselves or by others outside the original project team. But suffice it to say, availability and usability are far from synonymous (Wallis, Rolando, & Borgman, 2013). As Palmer et al. (2009) suggest, “Many fields have yet to develop the common practices needed for data sharing to succeed, in part because there may be no straightforward approach to gathering and coordinating data” (p. 33). This study supports their argument.

The 28 interviewees who planned to reuse their data hazarded fourteen diverse data reuses. Yet only three possible reuses were mentioned more than once: five interviewees planned to reuse their code, two intended to reuse their approach to or method of using their data, and two hoped to reuse their data in their dissertations (in some way).

Two notable contrasts emerged: the maturity of the grant project (Level I versus Level II) and the rank of the project team member (namely faculty member as opposed to student). Interviewees working on Planning (Level I) Grants found it more difficult to suggest possible reuses for their data, perhaps because of the perceived limited utility of workshop and meeting records as “data.” Nonetheless, such materials may provide a foundation for future grant applications.

Nearly all Principal Investigators thought they could reuse their project data in some capacity. Because of their relative job security (ten of the seventeen have tenure), these scholars may have more freedom to think beyond the span of the immediate project and to fit their SUG projects into their career research trajectories. On the other hand, students whether Master’s or doctoral have a clear educational end point (namely their graduation or the expiration of their funding). It may prove difficult to maintain ties with PIs subsequently. Graduate students might reuse project data at least for publications, Master’s theses, or dissertations. Data may be most amenable to students’ their reuse when employed for these conventional purposes. These conventional products may also prove more attractive to employers than more innovative products that some of the projects produced.

**Discussion**

Interviewees offered 102 free-form comments that addressed challenges they faced and lessons they learned. These comments comprised six categories: data, collaboration and communication, awareness and outreach, planning and project management, resources, and technological issues. Despite interviewees’
emphasis on data issues, they were almost equally concerned with non-technical issues.

Data Use Issues

Interviewees most frequently mentioned data issues (32 comments), but concerns about data were quite dispersed (Table 2). Versioning was the most commonly mentioned issue, but it figured in only four of 32 comments (12.5%). Other concerns included the sheer size of the data (9.4%), disciplinary translation issues (9.4%), documentation (9.4%), data cleaning (6.3%), data complexity (6.3%), file format (6.3%), and developing training data (6.3%). Given these challenges, it is no wonder that stakeholders struggled to suggest possible reuses for their data, as suggested above.

Collaboration and Communication

Collaboration and communication earned 20 comments from interviewees. Collaboration

Humanities scholars traditionally worked alone and rely heavily on personal interpretations of research materials (Stone, 1982; Watson-Boone, 1994). Yet these studies underestimated the extent of collaboration in the humanities, as subsequent work indicated (Brockman, Neumann, Palmer, & Tidline, 2001; Toms & O’Brien, 2008).

Benefits of collaboration among digital humanities scholars may include drawing upon a wider range of expertise, sharing costs and pooling resources, accessing new tools, developing standards and best practices, and raising awareness (Harvey, 2010). Collaborations occupy “a continuum of engagement, from basic consultation to fully integrated teamwork, and project management may be loosely coordinated or highly structured and closely administered” (Palmer et al., 2009, p. 25).

But collaboration remains challenging in the humanities and in digital curation alike (Abbott, 2015; Alexogiannopoulos et al., 2010; Benardou et al., 2013; Maron & Pickle, 2014; McCarty, 2012; Molloy, 2012; Waters, 2012). It depends on achieving common understandings of language and terminology, methods and research styles, theories, outputs and publications, and values (Siemens et al., 2012). More prosaically, teams need to allocate tasks, share information and documents, reach consensus on decisions, and schedule meetings (Siemens L., 2010). Many collaborative efforts run aground: “Individuals working in such interdisciplinary teams often find they are the ‘Other’...as if they come from foreign climes” (Terras, 2012, p. 213). Indeed, a report covering the 2007-2010 SUG projects found internal and external collaboration the most intractable problem Principal Investigators faced (National Endowment for the Humanities, Office of Digital Humanities, 2010).

Doctoral students, PIs, and Office of Digital Humanities officials reflected on the collaborative process. Doctoral students played key collaborative roles in the SUG projects. One PI (interviewee 38) characterized her doctoral student (also her PM) as the project’s “linchpin.” Often PIs were doctoral students’ supervisors or mentors as well as their de facto employers. Additionally, in some cases graduate students played liaison roles between PIs and other project workers such as undergraduates or Master’s students. But doctoral students could encounter unfamiliar challenges. One PM (interviewee 15) found it “awkward” and “a little strange” managing other graduate students for the first time not least because she lacked preparatory project management training.

Principal Investigators also reflected upon their collaborative efforts. Interviewee 27 described her work with the other PI on the project as the epitome of their longstanding “intellectual kinship.”
Interviewee 6 asserted that collaboration, specifically through co-writing “helped us tremendously.” On the other hand, interviewee 36 said she “assumed too much” of her graduate student worker regarding motivation and work ethic.

Finally, officials from the Office of Digital Humanities underlined the importance of collaboration. It appears “too much to ask,” interviewee 26 observed, to find a scholar who owns the requisite data curation, disciplinary/domain, and outreach expertise. The ODH promotes “strength in numbers” (interviewee 25) and tells applicants not to “go it alone” (interviewee 26).

In particular, interviewees mentioned problems with terminology (4), clarity (3), frequency (3), geographical distance (2), and timing (2). Two also noted the challenge of securing feedback from their collaborators.

Interdisciplinarity


Interviewees reflected upon interdisciplinarity as part and parcel of collaboration. One PM (interviewee 18) hailed the “tremendous amount” of interdisciplinarity in her project. Another project participant (interviewee 45) suggested that there was “a lot of skill involved” in negotiating with personnel from different academic units. According to one PI (interviewee 38), interdisciplinarity remains “incredibly difficult” and can come about serendipitously. In this vein, another PI (interviewee 19) described the gestation of her collaboration with another campus scholar as a “complete fluke.” Obstacles to interdisciplinary collaboration arose, however. One PI (interviewee 29) stressed the need to grow a “thick skin” and show willingness to ask experts from other disciplines for explanation or clarification.

“Navigating the literature and research practices of another field,” conclude Palmer et al. (2009), “requires developing familiarity with new terminology, concepts, theories, and methods” (p. 31).

Interdisciplinarity could pay dividends. One project member (interviewee 43) noted that interdisciplinarity strengthened their project group of four persons: the team proved both “flexible” and “nimble.” One PM (interviewee 15) even described one of her PIs as a “rare bird” able to traverse disciplines. According to one Office of Digital Humanities representative (interviewee 25), grantees may evolve into “wonderful bridge people.”

Communication

Translation issues proved most important to participants in this study, but concerns about clarity and about frequency of communication also emerged. One PI (interviewee 35) observed the difficulty of initiating and then sustaining robust communication among groups of scholars who had never met before and who worked apart from one another geographically.

Other types of divides could apply. For example, one PI (interviewee 3) emphasized the challenge of translation among technically literature and domain-knowledgeable persons as part of negotiating “institutional culture”: she found such translation “really important and very difficult.” Similarly, one PM (interviewee 23) spoke about even more specific translation issues, noting the challenges of achieving stakeholder agreement on the
definition and use of certain metadata categories such as genre.

Not surprisingly, project personnel relied heavily on email to communicate. Email remains the “lifeblood of communication” among collaborators (Palmer et al., 2009). Email conveys benefits: the sender and receiver do not have to be in the same place at the same time; the method is also inexpensive. On the other hand, asynchronous communication can be both effortful and time consuming. Moreover, details and nuance may be sacrificed (Siemens L., 2010).

Perhaps more surprising, the vast majority (84.2%) of projects’ personnel (at least some members of each team) met face to face at least once, the richest communication medium because it allows the apprehension of body and facial cues as well as immediate feedback. This medium nurtures trust and commitment, but also presupposes that stakeholders can travel to see one another (Siemens L., 2010). Last, nearly two-thirds (63.2%) of the nineteen projects employed Skype or video chat. These mechanisms surmount problems of geographical distance; they also allow real time feedback and convey at least some visual cues. Conversely, time delays and transmission problems can obtrude (Siemens L., 2010).

Nearly a third (31.6%) of projects relied on other modes of communication as well. Two projects communicated via telephone and two used Google Docs. One project each capitalized on co-writing, social media, a wiki, an IRC channel, and Google Chat. More projects might exploit these venues to ensure frequent, high-touch collaboration.

Email and other text-based asynchronous communication vehicles such as blogs and wikis offer important benefits: the sender and receiver do not have to be in the same place at same time and these mechanisms are inexpensive. On the other hand, details and nuance may be sacrificed and time lag may also prove problematic (Siemens L., 2010). Thus project personnel should likely rely on multiple communication channels. Put simply, digital communication simply cannot replace face-to-face in collaborative projects (Siemens L., 2010).

Librarians and Archivists

Librarians or personnel affiliated with the University Libraries took part in eight of the nineteen projects (42.1%). The PIs of two of the nineteen projects relied heavily on librarians. One (interviewee 30) called the library a “key player”; the other (interviewee 42) answered “Good Lord, yes,” when asked if librarians played a collaborative role in the project. On the other hand, one PI (interviewee 3) admitted that she did not work with librarians on her SUG project, but nonetheless believed such involvement “essential.” Similarly, this PI’s PM (interviewee 4) saw “lots of potential for collaboration” with librarians. Both claimed their SUG project was not yet mature enough to require librarian involvement, likely a misconception.

Archivists’ involvement in the nineteen projects was rare: they participated in only two projects (10.5%). Once again, this likely represents a waste of human resources and a growth area for archives and archivists (Jahnke & Asher, 2012; Noonan & Chute, 2014; Poole, 2015; Prom, 2011; Redwine et al., 2013).

Planning and Project Management
Project management skills include organization, planning and follow up, prioritization, grant administration, human resources, and conflict resolution. Often characterized as “soft” or implicit skills, they can resist measurement, which can contribute to their remaining overlooked (Reed, 2014). Planning, content creation, technical development, technical maintenance, preservation, dissemination, and storage—all need to be managed (Maron & Pickle, 2014). Project sustainability, finally, remains the “elephant in the room” (McGann, 2010, p. 5).

Planning and project management issues elicited 20 comments. One PI (interviewee 29) likened working on a digital humanities project to “wrestling an octopus.” One PM (interviewee 23) noted that being skilled at digital humanities work and at project management were two very different things.

A predominant concern of interviewees (11 comments) was the challenge of approaching projects with a long-term perspective. Other issues mentioned more than once included project management skills (3 comments), time management (3), and providing for early involvement in the project (2). Ultimately, project management may be an overlooked weak point of digital curation work in the digital humanities.

Duration of Grant
The Start-Up Grant award period lasts up to eighteen months, but twelve projects received no-cost extensions. It is no wonder a key concern of interviewees was project planning and management. Of the seven that did not request extensions, four were Planning Grants and all four concluded with workshops.

It appeared easier temporally to plan and to keep workshops or meetings than more experimental explorations on track. Additionally, the tendency of projects to train their personnel only as the project unfolds may not be the most time-efficient—or productive—strategy. Certainly optimal digital curation planning and practice requires early and frequent intervention. Project management skills such as organization, planning, prioritization, monitoring, and human resources seem of crucial but frequently overlooked importance in digital humanities work (Reed, 2014).

Awareness and Outreach
Interviewees identified awareness and outreach issues (19 total comments). Concerns suggested more than once in this area included demonstrating the relevance or legitimacy or both of digital curation to stakeholders (4 comments), researcher and faculty concerns (4), user issues (4), data preparation and preservation mindedness (2), and keeping abreast of the current state of the field (2).

Resources
Resources were mentioned, though perhaps not as frequently as one might expect (seven times). Beyond the need for more resources of all types, interviewees offered several more specific remarks. They stressed financial and temporal resources, appropriate personnel (interviewee 12 concluded, “Talent is expensive”), technological resources (the need for sufficient server space), and educational resources (those that delivered core concepts and best practices).
Technology

Technological issues were mentioned only four times and included user-friendly software, general technological shortcomings and technological “immaturity,” and a shortage of institutional technical support. Nonetheless, these projects evidently give credence to Brockman et al.’s (2001) that humanists continue to adopt technology to supplement their tried and true work practices.

Future Research

Ultimately, challenges surrounding data, collaboration and communication, planning and project management, awareness and outreach, resources, and technological issues hampered project personnel in their efforts to curate their data. In short, there persists a disjuncture between the types of innovative scholarly projects in which these scholars are engaging and their ability to curate the data underpinning these projects for future data reuse.

The findings of this study have three implications. First, the study suggests that awareness of the value of digital curation among digital humanists is increasing, even if many such scholars remain unsure about how best to implement optimal digital curation practices. Second, challenges in translating terminology between digital curation and digital humanities persist. Third, the study indicates that so-called soft skills such as communication and collaboration and planning and project management merit increased attention in digital curation education.

Three sets of questions are promising for future research; each one builds upon the work of this study. As Palmer and Cragin (2008) conclude, “Studies focused on data will be imperative to improve our understanding of both the epistemological bases and the actual practices that arise from new forms of collaboration and novel approaches to data management” (pp. 193-194).

One set of questions concentrates on sampling from other groups of digital humanists. Thirty-eight of 38 of 3 (two interviewees do not work at universities) interviewees (84.4%) work at a Carnegie RU/VH or Carnegie RU/H institution. Therefore, the sample included only four of 29 Carnegie categories. What are the digital curation practices (and education practices) for digital humanities scholars at liberal arts colleges, traditionally women’s colleges, Historically Black Colleges and Universities (HBCUs), community colleges, and Master’s level institutions? What digital curation skills are employed by personnel pursuing digital humanities projects more advanced (i.e. creating and using more data and data of greater complexity and variegation, as in the Digging into Data projects) than the Start-Up Grants?

A second set of questions takes a longitudinal approach. Have the nineteen SUG projects proved sustainable? Has their data been reused by project personnel, by those outside the original project team, or both? Have librarians’ and archivists’ roles in digital curation increased or evolved?

A third set of questions centers on digital curation education. Virtually none (93.3%) of the 45 interviewees had received any formal education in digital curation. Interviewee 38, a PI, summed up: stakeholders in digital humanities are “not doing a good enough job” teaching digital curation. Are digital humanists interested in learning skills to curate their data better? If so, what sort of educational program or framework would help them do so?

Despite manifold challenges, it seems clear that digital curation, as interviewee 16 asseverated, constitutes a “well-established part of the university of the future.” Concordantly, challenges in digital curation will become only more
urgent and more complex (Council on Library and Information Resources, 2013). In the end, digital curation in the digital humanities remains a “greatly unexplored area,” as interviewee 15 maintained.

Appendix A: Interview Questionnaire

Participation is voluntary. You may stop at any time, or decline to answer any question for any reason. Do I have your permission to audio record this interview? May I quote you directly in the final report? Do I have your consent to participate?

1) Do you consider yourself a digital humanist? Please elaborate.
2) Please describe your involvement in the project.
3) What were your goals in undertaking the project?
4) Did you use already existing data?
   a. If so, please describe what data you used and how you used it
5) What type(s) of data did you create?
6) Did your project involve interdisciplinarity? Please discuss.
7) What methods did your project involve? Qualitative? Quantitative? A blend?
8) What type(s) of software did you use in the project?
9) How did you develop your data management plan?
   a. Were you assisted in developing or implementing it by anyone such as a student, librarian, archivist, or other project team member?
   b. Did your data management plan change over the course of the project?
10) Did you encounter any challenges using your data during the project?
11) Was there any sharing or reuse of data during the project?
12) Where is your data now?
13) When and where do you plan to transfer or deposit your data?
   a. How long would you like to keep your data?
14) Have you shared your data?
   a. If so, with whom and under what conditions?
   b. Is it ready to be shared?
15) Are you planning to reuse your data? In what way(s)?
16) What were the deliverables of the project?
   a. Did those deliverables differ from those you expected?
17) Was any education or training involved for you or for others involved in the project?
   a. What types of education or training do you think would be most helpful?
18) Did you have collaborators, e.g. librarians or others on or off campus?
   a. Division of labor in project?
19) Did you work with students?
20) Would you do anything differently in managing your data before, during, or after your project? I.e. lessons learned?
21) Who else should I talk to?
## Appendix B: Projects Examined

<table>
<thead>
<tr>
<th>Project</th>
<th>Type of Grant</th>
<th>Discipline (ODH classification)</th>
<th>Duration: Projected/Actual (months)</th>
<th>Deliverable</th>
<th>Number of Personnel Interviewed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Implementation</td>
<td>Archaeology</td>
<td>12/19</td>
<td>Scholarly publication</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>Implementation</td>
<td>Archival Management and Conservation</td>
<td>13/20</td>
<td>Open source tool</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>Planning</td>
<td>Interdisciplinary Studies, General</td>
<td>12/12</td>
<td>Workshop</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>Planning</td>
<td>Library Science</td>
<td>12/12</td>
<td>Workshop</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>Implementation</td>
<td>Asian Languages</td>
<td>21/21</td>
<td>Prototype Platform</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>Planning</td>
<td>Humanities</td>
<td>10/10</td>
<td>Workshop</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>Implementation</td>
<td>Music History/Criticism</td>
<td>19/31</td>
<td>Database</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>Implementation</td>
<td>Arts</td>
<td>14/25</td>
<td>Software</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>Implementation</td>
<td>Interdisciplinary Studies, General</td>
<td>12/12</td>
<td>Best practices</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>Planning</td>
<td>Literature</td>
<td>12/16</td>
<td>Workshop</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>Implementation</td>
<td>Museum Studies/Historical Preservation</td>
<td>12/27</td>
<td>Prototype Platform</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>Implementation</td>
<td>Humanities</td>
<td>13/25</td>
<td>Preservation Strategies</td>
<td>2</td>
</tr>
<tr>
<td>13</td>
<td>Implementation</td>
<td>Archaeology</td>
<td>12/24</td>
<td>Prototype Platform</td>
<td>3</td>
</tr>
<tr>
<td>14</td>
<td>Planning</td>
<td>Art History/Criticism</td>
<td>12/18</td>
<td>Workshop</td>
<td>2</td>
</tr>
<tr>
<td>15</td>
<td>Planning</td>
<td>Humanities</td>
<td>12/12</td>
<td>Workshop</td>
<td>2</td>
</tr>
<tr>
<td>16</td>
<td>Implementation</td>
<td>Humanities</td>
<td>12/25</td>
<td>Web-based Tool</td>
<td>2</td>
</tr>
<tr>
<td>17</td>
<td>Implementation</td>
<td>Humanities</td>
<td>12/20</td>
<td>Model</td>
<td>3</td>
</tr>
<tr>
<td>18</td>
<td>Implementation</td>
<td>Literature</td>
<td>19/31</td>
<td>Software</td>
<td>2</td>
</tr>
<tr>
<td>19</td>
<td>Planning</td>
<td>Media Studies</td>
<td>6/18</td>
<td>Pilot Preservation Project</td>
<td>3</td>
</tr>
</tbody>
</table>

### Works Cited


Molloy, L. (2012). *JISC Research Data MANTRA Project at EDINA, Information Services, University of Edinburgh*. JISC.


3) http://blog.dmptool.org/about-the-dmptool/