

Now is the Future Now? The Urgency of Digital Curation in the Digital Humanities

Alex H. Poole <ahpoole_at_email_dot_unc_dot_edu>, University of North Carolina at Chapel Hill

Abstract

In their seminal report, *Our Cultural Commonwealth* (2006), the American Council of Learned Societies underscored the need for scholars engaged in digital humanities work to leverage their access to data both to expand their audience to the general public and to generate new research questions. “Now is the Future Now?” argues that the progress made in digital humanities toward these goals has depended and will depend not only on digital data, but also on their appropriate curation. The article defines digital humanities, data, so-called Big Data, and digital curation. Next it examines digital curation initiatives in the sciences and in the humanities that occurred before the release of *Our Cultural Commonwealth*. It then considers and evaluates the digital curation work undertaken in the sciences and in the humanities after the report’s publication. In theory and in practice digital curation has benefited substantially from practices developed and tested first in the natural sciences and subsequently adapted for and extended in the humanities. Finally, the piece explores the future work necessary to facilitate symbiosis between digital curation and digital humanities. Collaboration and cooperation, transcending geographical, disciplinary, and institutional boundaries, data sharing, policies and planning, education and training, sustainability — all remain pressing issues in 2013.

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. (*Our Cultural Commonwealth* [American Council of Learned Societies 2006])

Humanists do not lack for questions. (Amy Friedlander, [Friedlander 2009])

This much is clear: “big data” are not just for scientists anymore. (Christa Willford and Charles Henry, [Willford and Henry 2012])

The challenges humanities data stakeholders faced as of the mid-2000s seemed legion: the possible loss of, the fragility of, and the inaccessibility of the cultural record; the cultural record’s intricacy and complexity; vexing intellectual property restrictions; the dearth of incentives to experiment with cyberinfrastructure; uncertainty regarding the future mechanisms and economics of publishing and scholarly communication; and insufficient resources, will, and leadership [American Council of Learned Societies 2006]. But in its sixth report, *Our Cultural Commonwealth: The Report of the American*

Council of Learned Societies Commission on Cyberinfrastructure for the Humanities and Social Sciences (2006), the American Council of Learned Societies proclaimed, “digital technology can offer us new ways of seeing art, new ways of bearing witness to history, new ways of hearing and remembering human languages, new ways of reading texts, ancient and modern” [American Council of Learned Societies 2006, 16]. The report lobbied for increased investment in infrastructure, for policies that fostered openness and accessibility, for public and private sector cooperation, for invigorated leadership, for more scholarly workshops and fellowships, for more national centers, for consensually reached and open standards and tools, and for more extensible and reusable digital collections.

Most important, *Our Cultural Commonwealth* forecasted that if stakeholders adhered to its recommendations, the next five to six years would see, first, an expanded audience among the general public: “All parties should work energetically to ensure that scholarship and cultural heritage materials are *accessible to all* — from a student preparing a high-school project to a parent trying to understand the issues in a school-board debate to a tourist wanting to understand Rome’s art and architecture” [American Council of Learned Societies 2006, 31]. Digital information was “inherently democratizing” and represented a public good [American Council of Learned Societies 2006, 27]. As one of the report’s authors, John Unsworth, later reflected, the general public remains the most important audience for the humanities, digital and conventional [Unsworth 2009]. Willard McCarty (2012) rightly extended Unsworth’s point, noting that “Arguing for economic benefits is a long reach for the humanities, but the ‘well-being of citizens’ is not” [McCarty 2012, 119].

Second, a larger number of scholars would ask newfound research questions. There would be “new patterns and relations to be discerned, and deep structures in language, society, and culture to be exposed and explored” [American Council of Learned Societies 2006, 11]. Neither disciplinary boundaries nor individual institutions nor national borders would constrain digital cultural heritage materials. Scholars could see artifacts in new ways through digital imaging, performance footage, and mapping technology; they could bring together works from physical collections scattered in space and time and study across them; they could collaborate with distant colleagues; and they could engage in data mining, simulations, game play, role play, and virtual worlds.

Our Cultural Commonwealth crystallized the unprecedented urgency of digital data curation in the humanities. Many stakeholders since have embraced the importance of promoting the digital humanities through democratized access to and an expanded audience for cultural heritage materials and through posing new research questions — indeed, they have done so at an accelerating rate. Moreover, many stakeholders recognize the indispensability of digital curation in underpinning not only those specific goals, but also the more general aims of digital humanities scholarship. Despite its stakeholders’ marked advances on multiple fronts, however, *Our Cultural Commonwealth’s* specific recommendations remain of pressing importance in 2013. Ultimately, the digital humanities cannot thrive without digital data curation.

First, this paper defines and situates the digital humanities and both data and Big Data. Next, it probes digital curation, considering it both in the sciences and in the humanities. More specifically, it discusses the professionals who curate data, the key issues in data curation and how best to approach them, the importance of a lifecycle approach, the machinations of sharing and reusing data, and the role of data management planning. Third, it explores reports on and case studies of digital curation undertaken in the United States and United Kingdom prior to the release of *Our Cultural Commonwealth*. Fourth, it considers the trajectory of digital curation efforts in the United Kingdom and United States following *Our Cultural Commonwealth*. In particular, it examines more recent reports and case studies and juxtaposes these findings with those of earlier stakeholders. Finally, the paper assesses the state of digital curation in the humanities in 2013.

In 2009, John Unsworth reflected that the humanities scholars involved in *Our Cultural Commonwealth* found it “very difficult to say exactly why the work they do should matter to the general public; in fact, they often did not believe that it would” [Unsworth 2009]. But the humanities seemed “much better off” than the sciences: “the public might want the results of scientific research, but they are not all that interested in the actual content and conduct of that research; in the humanities research does have a general audience.” [Unsworth 2009]. Digital curation ensures that research and readership in the humanities will be maximized.

I.

2

3

4

5

6

Like digital curation, the digital humanities represent “a hybrid domain, crossing disciplinary boundaries and also traditional barriers between theory and practice, technological implementation and scholarly reflection” [Flanders, Piez and Terras 2007]. Even three years later, the definition and scope of the digital humanities remained “under negotiation” [Svensson 2010]. But such equivocation obscured a pivotal shift: the digital humanities, argued Matthew Kirschenbaum of the Maryland Institute for Technology in the Humanities (MITH), were coalescing into “something like a movement” armed with an “unusually strong sense of community and common purpose” [Kirschenbaum 2010].

7

A year later, Rafael Alvarado of the University of Virginia’s Sciences, Humanities, and Arts Network of Technological Initiatives (SHANTI) thought the digital humanities constituted a “genealogy,” viz. “a network of family resemblances among provisional schools of thought, methodological interests, and preferred tools, a history of people who have chosen to call themselves digital humanists and who in the process are creating that definition” [Alvarado 2011]. Still, “persistent anxiety” about the “richness and strangeness” of the digital humanities lingered [Piez 2011]. More pragmatically, digital humanities scholarship remained a “backwater” [Borgman 2009] regarding hiring, tenure, and teaching and younger scholars often felt “ghettoized and even disadvantaged” as a result [Friedlander 2009]. As such, “alternative” or “para-academic” jobs have served as a frequent recourse [Flanders 2012].

8

Belying such concerns, however, recent scholarship indicates a “visionary and forward-looking sentiment” in the digital humanities, not least because of a salutary increase in size and diversity in the field over the past half-dozen years [Svensson 2012]. Optimally, the digital humanities will serve as “a laboratory, innovation agency, portal and collaborative initiator for the humanities, and as a respectful meeting place or trading zone for the humanities, technology and culture, extending across research, education and innovation” [Svensson 2012]. Indeed, work in the digital humanities frequently “better serves values such as pluralism and innovation than do the professional values of the traditional academic humanities, which often seem to be crouched in a defensive position” [Spiro 2012, 20]. Fulfilling such an ambitious agenda in the digital humanities depends upon digital data and even more important, upon its curation. As historian Dan Cohen (2012) suggests, “*Curation* becomes more important than publication once publication ceases to be limited” [Cohen 2012, 321].

9

The digital humanities pivot around data. The Digital Curation Center defines data as “A reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing.”^[1] Data may be valuable as a public good, as evidence, or as part of the legal record [Rusbridge 2007]. *Our Cultural Commonwealth* report characterized digital data as “notoriously fragile, short-lived, and easy to manipulate without leaving obvious evidence of fraud” [American Council of Learned Societies 2006, 18]. Worse, much collected data were neither curated nor published whatsoever; numerous “data iceberg[s]” resulted [Hey, Tansley and Tolle 2009]. Even more challenging, just as the humanities depend upon context and a critical mass thereof, so too do many humanities data objects maintain intricate structures predicated upon numerous structural and semantic internal relationships. Such objects, therefore, are exceedingly contextual themselves [Blanke, Hedges and Dunn 2009].

10

The notion of data as a vehicle for new scholarship or more rigorous scholarship or both in the natural sciences, social sciences, and humanities accrued unprecedented cachet with the emergence of “Big Data.” Big Data amalgamates technology, analysis, and mythology. Ideally amenable to study at all levels, it undergirds new forms of analysis or enriches existing ones, and nonetheless remains accessible to non-experts [National Science Board 2005]. Harnessing computing power and algorithmic accuracy, researchers may exploit large data sets not only to tease out patterns, but also to inform economic, social, technological, or legal arguments. As Associate Dean for Research Data Management at Johns Hopkins University Sayeed Choudhury asserts (2010), “Fundamentally, there is a shift from a document-centric view of scholarship to a data-centric view of scholarship” [Choudhury 2010, 194]. Scholarship in this vein, moreover, shows that “Technology and creativity are not dichotomous, but are mutually dependent” [Blanke, Hedges and Dunn 2009, 477]. Amy Friedlander of the Council of Library and Information Resources elaborates: “if the infrastructure answers the question, how?, the research program answers the questions what? and why?” [Friedlander 2009].

11

Big Data evinces other important characteristics besides size. As the Coalition for Networked Information’s Cliff Lynch insists, “Data can be ‘big’ in different ways”: stakeholders must consider its size, but its lasting significance and the challenges of describing it as well [Lynch 2008]. As such, Big Data may be “less about data that is big than...about a

12

capacity to search, aggregate, and cross-reference large data sets” [boyd and Crawford 2012, 663]. More problematic, thought the Aspen Institute’s David Bollier, “One of the most persistent, unresolved questions is whether Big Data truly yields new insights — or whether it simply sows more confusion and false confidence” [Bollier 2011, 14]. Big Data engenders seminal challenges for stakeholders.

First, Big Data revamps the definition of knowledge epistemologically and ethically. Second, it facilitates unprecedented and possibly unwarranted claims to objectivity and accuracy. Third, bigger data are not ipso facto tantamount to better data; methodological concerns must not be given short shrift. Fourth, Big Data loses meaning when denuded of context. Fifth, ethical issues revolving around accountability, power, and control must be weighed. Finally, Big Data may reinforce familiar or create new digital divides: the richest and most prestigious institutions can purchase the best data [boyd and Crawford 2012]. Data, in short, may rupture the status quo in the natural sciences, in the social sciences, or in the humanities [Bollier 2011]. Disruptive or not, data requires curation to remain usable.

13

II.

Though the term was coined in 2001 in the United Kingdom, the array of concerns animating digital curation emerged in the middle of the 1990s and engaged a variegated cohort of stakeholders [Higgins 2011]. The Digital Curation Center posits that “digital curation is about maintaining and adding value to a trusted body of digital information for current and future use.” [2] It constitutes an “umbrella term for digital preservation, data curation, and digital asset and electronic records management” and brings together the scientific, educational, and professional communities with governmental and private organizations [Yakel 2007]. Associate Dean of Libraries at California Polytechnic State University Anna Gold (2010) notes that “the activities of curation are highly interconnected within a system of systems, including institutional, national, scientific, cultural, and social practices as well as economic and technological systems” [Gold 2010, 3]. Digital curation “involves the management of digital objects over their entire lifecycle, ranging from pre-creation activities wherein systems are designed, and file formats and other data creation standards are established, through ongoing capture of evolving contextual information for digital assets housed in archival repositories” [Lee and Tibbo 2007]. It amounts to “a central challenge and opportunity” for any data-intensive organization [Hank and Davidson 2009]. Neither its complexity nor its importance for humanities data can be overstated. Historian Mark Kornbluh (2008) insists, “Digital humanities content requires curation” [Kornbluh 2008]. Indeed, cultural information is “a privileged domain” for digital curation [Constantopoulos and Dallas 2007, 5]. Put simply, curation adds value to digital assets.

14

Curators of data comprise many stakeholders: individuals using their hard drives or networked drives, departments or groups using shared or separate drives, institutions, communities of institutions either formal or informal, disciplines, publishers, national services or national data services, or other third parties [Rusbridge 2007]. Key issues in effecting curation include the size of the data, the number of objects to be curated and their complexity, the interventions needed, ethical and legal concerns, policies, practices, standards, and incentives [Rusbridge 2007]. More pointedly, a digital curation program must have a flexible and scalable infrastructure to ingest content, an economically and a technologically sustainable system to provide for data integrity checking, reversioning, and other open-ended tasks, and human and machine interfaces that offer multiple appropriate access points. Provisions must be made for creating or capturing metadata, for recording data provenance, for providing unique identifiers, for hewing to intellectual property rights laws, for drawing up appropriate policies regarding, for instance, submission and use, and finally, for presenting data collection in a cogent and useful context [Witt 2009].

15

An optimal approach to curation involves four steps. First, curators should build curation or re-usability into their workflow. This allows the easiest capture of provenance information and associated metadata. Second, curators should retain the ability to process data, not merely the data themselves. Standard data formats and file types processed with standard programs are preferable, though in some case open source options are advantageous. Third, curators should render transparent any questions about ownership and allowable use. Last, curators should make data citable, adhering to standard formats and to discipline-specific practices [Rusbridge 2007].

16

Digital curation depends upon a lifecycle approach: in other words, all stages and actions are identified, planned, and implemented in the appropriate order. A lifecycle approach implicates multiple processes: appraisal, ingestion,

17

classification, indexing, and cataloging, knowledge enhancement, presentation, publication, and dissemination, use experience, repository management, preservation, goal and usage modeling, domain modeling, and authority management [Constantopoulos and Dallas 2007]. This approach ensures “the maintenance of authenticity, reliability, integrity and usability of digital material” [Higgins 2008, 135]. As Jillian Wallis and her colleagues (2008) contend of ecological sensing data, “Shifting the practices of archiving such as appraisal, curation, and tracking provenance into earlier stages of a given material’s lifecycle can increase the likelihood of capturing reliable, valid, and interpretable data” — and of curating it appropriately [Wallis, Borgman, Mayernik and Pepe 2008, 115].

Christine Borgman (2012) observes that sharing data allows scholars to reproduce or verify research findings, to make findings generated by publicly-funded research available to the public, to permit other researchers to ask new questions about existing data, and to advance research and promote innovation [Borgman 2012]. But stakeholders who consider sharing must know which data can be shared, why it should be shared, by whom and with whom, under what conditions, and to what effect [Borgman 2012]. Rationales for sharing differ, however, by the arguments advanced in its favor, the motivations of its beneficiaries, and the not invariably compatible incentives of stakeholders [Borgman 2012].

18

Conversely, disincentives to share data persist. For example, researchers may fear that they will fail to receive appropriate credit for such labors or that others will “scoop” them. Second, documenting data in a reusable form necessitates much labor. Third, creators of data may worry about re-users misusing or misinterpreting the original data or about a related concern, intellectual property control. Fourth, confidentiality or privacy concerns, legal or otherwise, may motivate scholars to restrict access [Borgman 2012]. Not to be overlooked, though, sharing is “only of use if there are others to share with” [King 2007, 186]. Sharing is purportedly a common practice only in the natural sciences, astronomy and genomics prominent among them [Borgman 2012]. But other fields are following; momentum for data sharing in the social sciences is “evident and growing” [Crosas 2011].

19

Sharing data presages that data’s reuse. To be reused, data must be translatable and thus visible and coherent. Appropriate mechanisms must ensure that data quality and provenance can be trusted [Carlson and Anderson 2007]. The ability to contextualize and document both data and pertinent processes hinges on the discipline’s history and on the configuration of its particular research community [Carlson and Anderson 2007]. Indeed, in *all* disciplines “researcher practices around data are always highly specific and qualitative, even within quantitative disciplines, and that the data are always ‘cooked’ ” [Carlson and Anderson 2007, 144]. Providing for reuse thus requires “making explicit their [data’s] context of production and setting up appropriate systems of quality checks and assessment” [Carlson and Anderson 2007, 644]. To this end, the National Institutes of Health mandated that researchers deposit peer-reviewed, NIH-funded articles in PubMed Central as early as 2008.

20

But ensuring data management plans are created, let alone followed, has been challenging; indeed, merely ensuring that planning represents a systematic and continuous management activity remains a hurdle [Becker 2009]. More recently, the National Science Foundation stipulated that each grant proposal include a data management plan explaining how the project intends to disseminate and share its research results. The NSF noted, “Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. Grantees are expected to encourage and facilitate such sharing.”^[3] Yet good data management plans are as important in the humanities as they are in the natural sciences.

21

Humanists wisely followed the lead of their brethren vis-à-vis data management plans. In 2012, the National Endowment for the Humanities mandated that grant applicants submit a data management plan that addressed four broad issues. The Office of Digital Humanities deliberately aligned its guidance with the NSF’s, assuming grantees could exploit extant or emerging data management initiatives at their home institutions.^[4] First, applicants would describe the types of data their project would generate and subsequently share, the ways in which they would manage and maintain their data, the legal and ethical restrictions that might affect their ability to manage their data, and the mechanism(s) by which they would share or make their data accessible. Second, applicants would address the period of data retention: based on disciplinary norms and best practices, how long would applicants retain their data before sharing it? Third, applicants would describe their data formats and how to render those formats most amenable to

22

dissemination. Finally, applicants would describe the resources and facilities to be used for storing their data and preserving its accessibility. The NEH planned to monitor awardees, though primarily through the awardees' interim and final reports.^[5] More practically, the NEH plans to conduct workshops in 2013 and 2014 to help participants embrace a lifecycle approach to data curation, to model data, to calculate and manage risk, to learn about salient tools and systems, to leverage data curation skills, and to stay current with developments in the field.^[6]

Despite the long term importance of digital curation, however, researchers tend often to postpone it as “that extra burden, the one just beyond what is currently possible, in the queue behind meeting the conference deadline and writing the grant application” [Rusbridge 2007]. A 2002 United Kingdom study found, too, that “sticks are less effective than carrots — people must want to provide their primary research data and be given incentives to undertake the curation work which benefits the wider research community rather than the individual data creators themselves” [Lord and Macdonald 2003, 37]. Information scientist Michael Lesk urges digital curation stakeholders to “focus on *good enough*, on *when needed*, and on *getting help*” [Lesk 2010]. Sundry researchers have focused on just these sorts of issues.

23

III.

Reports and specific projects both before and after *Our Cultural Commonwealth* show how stakeholders — in a variety of situations and from a variety of perspectives — have responded to the prospects of digital curation. Well before *Our Cultural Commonwealth*, scholars turned their attention to digital curation generally and to specific curation initiatives. By the early 2000s, scientists and humanists faced similar problems, namely electronic sources and datasets too large for traditional analysis and materials that demanded contextual knowledge outstripping what an individual researcher could master. But unlike scientists, humanists lacked the resources to construct the new requisite scholarly infrastructure. Scientists have been “remarkably effective” in making their arguments for funding to administrators, legislatures, funding agencies, and the general public [Borgman 2009]. Thus investments remained “highly uneven” by field [Waters 2007, 8]. In no small measure because of their superior resources, stakeholders in the natural sciences took the lead in addressing the curation needs of Big Data in the early 2000s.

24

Scholarship produced by the National Science Foundation and the National Science Board in the United States introduced a set of concerns that remain relevant — indeed, pressing — a decade later. Underwritten by the National Science Foundation Blue Ribbon Panel on Cyberinfrastructure, the “Atkins Report” of 2003 announced, “a new age has dawned in scientific and engineering research, pushed by continuing progress in computing, information, and communication technology, and pulled by the expanding complexity, scope, and scale of today’s challenges” [Atkins 2003, 1]. Such developments triggered considerable optimism about addressing priorities such as climate change and natural disasters, national security, and public health. Feedback from research communities, meanwhile, suggested that such projects necessitated federated resources (namely data and facilities), multidisciplinary expertise, and an international reach. The NSF pledged to lead the effort [Atkins 2003].

25

Two years later, the National Science Foundation’s Cyberinfrastructure Council revisited the importance of interdisciplinarity and collaboration in supporting new research possibilities. The Council queried, “What answers will we find — to questions we have yet to ask — in the very large datasets that are being produced by telescopes, sensor networks, and other experimental facilities?” [National Science Foundation 2005, 4]. Despite “converging advances” in numerous areas from networking to data systems, still more collaborative partnerships were needed on national and international fronts and among government agencies, private sector organizations, and educational institutions [National Science Foundation 2005, 4]. Also released in 2005, the National Science Board’s report on “Long-Lived Digital Data Collections” stressed long-lived digital data’s role in spurring democratization in science and education. The report advocated for an “agency-wide umbrella strategy” in service of this goal [National Science Foundation 2005, 11].

26

In no small measure due to the leadership of the NSF and the NSB in the natural sciences, by 2006 fields such as astronomy, particle physics, and bioinformatics were grappling with the research possibilities of Big Data. Industries ranging from banking to pharmaceuticals, medicine to aerospace, also sought to use unprecedented amounts of data, albeit commercially [Beagrie 2006]. Such possibilities captured — and in some cases galvanized — public attention. But on the other hand, Big Data in the humanities seemed to generate less fanfare among scholars or the public. But

27

stirrings in various digital humanities arenas belied observers' assumption of stasis.

Since the early 2000s, for instance, digital humanities centers have been a "driving force" for digital scholarship [Zorich 2009, 70]; [Fraistat 2012]. The Digital Curation Center has shown particularly vital leadership since 2004 [Beagrie 2004]; [Rusbridge et al 2005]; [Hockx-Yu 2007]. These "hubs" have helped transform humanities scholarship and teaching, advocated for the humanities' continuing usefulness in a digital environment, served as intellectual "sandboxes," offered sites for training, fostered interdisciplinarity, attracted new audiences, engaged with various professional communities, encouraged collaborations among numerous communities, and extended otherwise unavailable operational services to scholarly communities [Zorich 2009]. For example, the Roy Rosenzweig Center for History and New Media at George Mason University pledges "to incorporate multiple voices, reach diverse audiences, and encourage popular participation in presenting and preserving the past." [7] Though siloing, redundancies, and non-integrated digital production may undercut the effectiveness of such digital humanities centers, their importance for digital curation and thus for recruiting new audiences and addressing new scholarly questions cannot be gainsaid [Zorich 2009].

28

Similarly, the emergence and increased visibility of institution repositories (IRs) beginning in the early 2000s generated new and stimulated ongoing digital curation efforts. Institutional repositories both extend the reach of scholarly communication by spurring innovation in a decentralized publishing system and represent tangible indicators of an institution's prestige and public value socially, scientifically, and economically [Crow 2002]. A "mature" IR, Clifford Lynch proposed in 2003, would contain faculty and students' research and teaching materials. It would document the institution itself, namely its events and performances. Most important, it would hold experimental and observational data [Lynch 2003]. As with centers, the importance of institutional repositories for digital curation, digital humanities, and their commingling cannot be overstated.

29

Notwithstanding the leadership evinced by the National Science Foundation and the National Science Board, early data-intensive research projects tested the reports' assertions at the grassroots and provided salutary lessons for digital curation stakeholders. For instance, the Biological Sciences Collaboratory (BSC) at Pacific Northwest National Laboratory sought both to offer tools and capabilities to facilitate collaboration and sharing and to capture the context(s) in which sharing occurred. The BSC enabled biological data and analyses to be shared through metadata capture, electronic lab notebooks, data organization views, data provenance tracking, analysis notes, task management, and scientific workflow management. But successful sharing also required the provision of overall contexts regarding total data space, applications, experiments, projects, and the scientific community. Such provision of context occurred frequently in one to one situations, whether face to face or through email [Chin and Lansing 2004]. In short, standards and best practices were conspicuously lacking.

30

Also in the early 2000s, the Collaboratory for Multi-scale Chemical Science (CMCS) cultivated an informatics-based approach to synthesizing multi-scale information that in turn supported systems-based research. One group of researchers drew two important conclusions. First, they argued, "As knowledge grids lower barriers to discovering, analyzing, and generating chemical information, technologies and research processes will need to co-evolve" [Myers et al 2005, 251]. In other words, researchers must avoid letting technology outrun research agendas. Second, Myers and his colleagues called for flexibility: "sub-communities will need to be able to independently develop and evolve their domain resources while contributing to multi-scale goals" [Myers et al 2005, 251].

31

Established by the National Science Foundation in 1980, the Long Term Ecological Research (LTER) network by 2006 hosted 26 sites locally and globally, supporting disciplines ranging from soil chemistry to stream flows to forest ecology. A 2006 study of LTER called for further study of actual curation practices over the long term to counter the "technical overemphasis inherent in near-term planning and with increased computing power, middleware, and shared grid capabilities" [Karasti, Baker and Halkola 2006, 324]. Second, LTER's work underscored that "growing attention to informatics, education, and social sciences initiates an interdisciplinary coordination within which jointly framed questions create new types of data needs and an arena within which data integration can be explored" [Karasti, Baker and Halkola 2006, 325]. Third, LTER showed that "it is the process of creating standards that is informed by practice and a likely determining factor of success of whether a deployed or adopted standard is enacted in practice" [Karasti, Baker and Halkola 2006, 343]. Fourth, open access to publicly-funded research seemed attractive but had not been

32

implemented or tested. In tackling these issues, ultimately, research communities must be involved from the ground up and from the project's germination.

Early reports and case studies in natural sciences in the United States both evaluated previous work and pushed for expanded and innovative future work. Reports by the National Science Foundation and by the National Science Board underscored the indispensability of collaboration, namely in sharing resources and strategies across geographic and disciplinary boundaries. Similarly, the reports stressed the democratic potential inherent in Big Data in the sciences. Early case studies, meanwhile, also foregrounded collaboration and interdisciplinarity. But this work contributed new findings as well. Perhaps most important, sharing required the provision of appropriate context. Second, these cases demonstrated the need for balance and flexibility: between new technical advances and new research questions and between disciplinary (and even sub-disciplinary) differences and large-scale common goals. Third, early cases showed the need for consensually developed standards and best practices. Finally, they considered the possibility of open access to publicly-funded research data. Subsequent digital curation efforts in the United Kingdom and United States, especially in the humanities, built upon and refined these priorities while allowing them to be tested empirically.

33

IV.

Despite the attention given to developments in the natural sciences, curation in the humanities was also progressing, albeit in less high-profile fashion. The United Kingdom's grassroots strategy of the mid-2000s laid important groundwork. For instance, the University of York-based Archaeological Data Service (ADS)'s Archaeotools: Data Mining, Faceted Classification, and e-Archaeology made available 40,000 reports of gray literature. Oxford University's Image, Text, Interpretation: e-Science, Technology and Documents deciphered fragmentary, stained, or damaged classical manuscripts. Finally, Birmingham's Medieval Warfare on the Grid: The Case of Manzikert permitted a virtual reenactment of the 1071 battle. Such projects not only hinted at the potential use of crowdsourcing (and thus democratized knowledge) to support data integration for research in the humanities, but also indicated a "clear trend" toward the development and use of new scholarly methodologies [Blanke, Hedges and Dunn 2009, 479].

34

Other case studies in the United Kingdom fleshed out this work. These cases more explicitly addressed sharing, reuse, and data management planning — and their potential ramifications for new audiences and research questions. For instance, a 2007 study addressing four United Kingdom interdisciplinary case studies — SkyProject, SurveyProject, CurationProject, and AnthroProject — illuminated data sharing and reuse practices. These projects suggested two correctives to conventional wisdom about data-intensive scholarship. First, knowledge could not be easily extracted either from its creators or from its original contexts and be facilely reused. Numbers and raw data could never be self-explanatory: how much context was "enough"? Second, Carlson and Anderson found the presumed binary divide between quantitative and qualitative sciences spurious. Rather, project team members constructed "socio-technical hybrids" through collecting, processing, annotation, release, and reuse of data [Carlson and Anderson 2007, 636].

35

Also addressing sharing and reuse and conducted between 2007 and 2009, the United Kingdom's Sharing, Curation, Reuse and Preservation (SCARP) case studies "aimed to understand expectations, risks and constraints, and find appropriate ways to build on current capabilities" in digital curation [Lyon, Rusbridge, Neilson and Whyte 2009] The research groups involved in the SCARP cases — Curating Brain Images in a Psychiatric Research Group: Infrastructure and Preservation Issues; Curating Atmospheric Data for Long Term Use: Infrastructure and Preservation Issues for the Atmospheric Sciences Community; Clinical Data from Home to Health Centre: the Telehealth Curation Lifecycle; Curated Databases in the Life Sciences: The Edinburgh Mouse Atlas Project; Roles and Reusability of Video Data in Social Studies of Interaction; Digital Curation Approaches for Architecture; and Curation of Research Data in the Disciplines of Engineering — lacked formalized curation practices. Still, they showed commonalities. First, researchers protected their own data. Second, they framed reuse as a way to advance their own research efforts. Finally, researchers thought interdisciplinary work pivotal in addressing data integration, schema development, quality assessment, and pooled storage [Lyon, Rusbridge, Neilson and Whyte 2009].

36

A 2009 United Kingdom study returned to the natural sciences, specifically the life sciences at the University of Edinburgh. It analyzed seven case studies: Animal Genetics and Animal Disease Genetics; Transgenesis in the Chick

37

and Development of the Chick Embryo; Epidemiology of Zoonotic Diseases; Neuroscience; Systems Biology; Regenerative Medicine; and Botanical Curation. All seven cases examined humans, animals, and plants but did so in a variety of research environments: analytical laboratory-based, field, and in-silico. The cases produced data ranging from field to image, clinical to laboratory-derived.

Each group customarily worked in a culture of data exchange in which use and generation is “recognizably participative, with most groups exhibiting complex levels of identifiable and routine data exchange” [Pryor 2009, 74]. On the other hand, these researchers shared their methods and tools more freely than their experimental data, remaining “naturally reluctant to share data that comprise the main means of adding value to their own research and...their careers” [Pryor 2009, 76]. Personal relationships loomed large in researchers’ willingness to share their data externally; conversely, they felt apprehensive about cyber-sharing. The Edinburgh study confirmed that national strategies and policies must take root in the practices of specific research communities. Input from below is as important as input from above.

38

Notwithstanding data sharing and reuse, stakeholders also began to think more carefully about data management, especially its planning component. Assessing the Rural Economy and Land Use program (RELU) (established in 2004) and the longitudinal, qualitative Timescapes program (established in 2007) the Economic and Social Research Council (ESRC) in the United Kingdom discerned that researchers needed more information about how to plan data management better. They particularly needed assistance with implementing informed consent procedures and with ensuring anonymization. Beyond data management, the ESRC emphasized that “Planning data management does not guarantee its implementation, and research funders need to consider how to ensure that good data management intentions are indeed implemented and revisited” [Eynden, Bishop, Horton and Corti 2010, 3]. Unfortunately, data management plans, much less successfully implemented and enforced ones, remain few in number and far from uniform in content, especially in the humanities, as of 2013.

39

Perhaps the most important United Kingdom digital curation case study, over more than a decade (1997-2008) the Arts and Humanities Data Service Performing Arts subject center (AHDS Performing Arts) in the United Kingdom safeguarded the digital products of more than 60 projects and provided digital resources (music, theater, dance, radio, film, television, and performance) to the United Kingdom research and teaching community. The AHDS web portal made information about these projects, as well as the knowledge of how best to create, to manage, and to preserve such digital content, freely accessible. The project ultimately offered “a national approach to developing best practice in digital curation, whilst maintaining the subject-based expertise so important for offering appropriate strategies and advice in domains with very specific needs, such as Performing Arts” [Abbott, Jones and Ross 2008, 2]. Moreover, it helped create and subsequently nurture a variety of research and practice communities and effected knowledge transfer to and among them about how to increase the long-term value of their performances. Initiatives such as the AHDS Performing Arts and its lessons both inspired and complemented digital curation work in the United States.

40

As in the United Kingdom, curation work in the United States in the second half of the 2000s accrued momentum in the humanities and retained it in the natural sciences. Digital curation efforts revealed both change and continuity. Bolstering earlier research, new case studies stressed the importance of coordination-cum-collaboration, an interdisciplinary or multidisciplinary approach, and the need for common standards. The studies also emphasized challenges such as the expense of curation and the recruitment of new audiences. But these studies highlighted progress in attracting new audiences and in addressing new research questions as well. In the same vein, other projects demonstrated the potential payoff of crowdsourcing, democratized access to and scholarship based on such opportunities, and how these possibilities related to ever-expanding computing power. Last, the first “Digging into Data” challenge inaugurated in 2009 represents perhaps the most promising development yet vis-à-vis new research possibilities and new audiences by dint of digital curation.

41

First, a 2007 workshop underwritten by the National Science Foundation and the Joint Information Systems Committee embraced the sciences, the social sciences, and the humanities and attracted American and European stakeholders from government, higher education, and industry. Participants agreed that unprecedented amounts of digital content necessitated a new and qualitatively different form of research and scholarship: “cyberscholarship.” But prospective scholars needed to develop national and international coordination, interdisciplinary research and development efforts,

42

and consensual standards [Arms and Larsen 2007].

Three contemporary projects in the United States showed cyberscholarship's nascent possibilities. The National Science Foundation-funded National Virtual Observatory (NVO) brought together disparate sets of astronomical data, coordinated access to this distributed data, and allowed users to select data extracts and download them to personal computers. Second, the National Center for Biotechnology Information (a division of the National Library of Medicine) developed Entrez, which pulled together sources ranging from PubMed citations and abstracts to content from databases such as Genbank. Moreover, Entrez provided cross-domain search capacity across its 23 databases and allowed researchers to use their own machines to explore data. Third, Cornell University's Web Lab (WL) copied large chunks of the Internet Archive's content to the Lab, mounted it on their computer system, organized it, and offered effort-saving tools and services to researchers [Arms 2008]. Ultimately, these three projects enabled new types of research and broadened the potential audience for producing and consuming such research.

43

Meanwhile, American scholars in the liberal arts also came to realize the research and scholarly potential of large quantities of data — and how that potential ramified into questions of audience [Green and Roy 2008, 36]. Cyberscholarship supported two new analytical approaches. First, data-driven scholarship depended upon algorithmic selecting and sorting. A second type of scholarship explored the culture of computer and social networking. In either case, as the Perseus Project and the Institute for Advanced Technology in the Humanities (IATH) at the University of Virginia showed, liberal arts cyberscholarship “takes a village”; in these cases, cyberscholarship depended upon collaborators ranging from faculty members to software programmers, designers to project managers, digitization specialists to copyright lawyers.

44

Cyberscholarship in the liberal arts as elsewhere faced obstacles. Its sheer expense could exacerbate the “digital divide.” One promising way of democratizing services was to develop templates to help with the creation of scholarship, as at the Institute for the Future of the Book's Sophie or the New Media Consortium's Pachyderm. Second was a problem of audience: how could stakeholders seed projects, get them germinate, and finally facilitate their spread nationally and internationally? Potential options included privatization, open source and thus “pay as you say,” or transinstitutional associations like the National Institute for Technology and Liberal Education (NITLE) [Green and Roy 2008, 36].

45

A specific example of fruitful cyberscholarship emerged with the Quilt Index, a project that gestated in the late 1990s. A National Endowment for the Humanities planning grant awarded to Michigan State University allowed the conversion of quilts into digital representations. Collaboration among scholars and curators then yielded a standardized vocabulary and standardized database fields to capture core information. The Quilt Index therefore achieved maximum flexibility and pointed toward future growth and cross-institutional collaborations.

46

The NEH subsequently funded the creation of Michigan State University's MATRIX: The Center for the Humane Arts, Letters, and Social Sciences On-line. Partnering with the Alliance for American Quilts and four collecting institutions, MATRIX created a searchable database and a web interface usable across diverse institutions. Next, a second-generation digital repository financed by the NEH and the Institute of Museum and Library Services both provided for long-term preservation of data in the Quilt Index and developed crosswalk tools to assist institutions in formatting data and in ingesting quilt materials from their own records. After another round of development funded by the Institute of Museum and Library Services (IMLS), any individual or institution could contribute to the Quilt Index. Supplementary materials accumulated: journals about quilts, pictures and photographs, published quilt patterns, and oral histories. Most recently, the Index has added 2.0 capabilities, including tools that facilitate using the product pedagogically.

47

Ultimately, the Quilt Index allowed contributors to build new content, to publish new scholarship, and to critique quilts and exhibitions. The project cultivated new and enlarged audiences and engendered new research questions. As historian Mark Kornbluh noted, “My ultimate goal for the Quilt Index is to be able to ask questions in a way that no one has been able to ask before” [Kornbluh 2008].

48

As the Quilt Index suggested, digital curation of data in the humanities garnered new appreciation in the latter half of the 2000s, but it continued to mature in the natural sciences, too. Most notably, the National Science Foundation

49

inaugurated the Sustainable Digital Data Preservation and Access Network Partners (DataNet) in 2007 to support national and international data research infrastructure organizations. DataNet integrated library science, archival science, computer science, information science, domain science expertise, and cyberinfrastructure. “By demonstrating feasibility, identifying best practices, establishing viable models for long term technical and economic sustainability, and incorporating frontier research,” the program solicitation noted, “these exemplar organizations can serve as the basis for rational investment in digital preservation and access by diverse sectors of society at the local, regional, national, and international levels, paving the way for a robust and resilient national and global digital data framework.” [8] Data Conservancy and DataNetONE proved path-breaking projects in just this sense.[9]

DataNet aside, by 2009 projects in the natural sciences had addressed crowdsourcing, democratizing access, and exploiting increased computational power in service of despoiling “needles” in data “haystacks.” For example, through crowdsourcing the Sloan Digital Sky Survey (SDSS) tested the claim that more galaxies rotate in an anticlockwise than in a clockwise direction. Using custom code, project staff created a webpage that provided pictures of galaxies to members of the public willing to play Galaxy Zoo, a game that focused on classifying the “handedness” of the galaxies. The project’s first year drew over 50 million classifications. The work of such “citizen-scientists” was *as accurate* as work done by astronomers, a propitious development for digital curation stakeholders [Goodman and Wong 2009].

In a related project, Microsoft’s WorldWide Telescope (WWT) democratized access to online data stored in the cloud. A user could enlist WWT to pan or zoom around the sky at nearly any wavelength; to examine an observationally-derived three-dimensional model of the universe; to discern correspondences between features at multiple wavelengths at some point(s) in the sky and then examine relevant publications linked to them; to connect a telescope to a computer running WWT and overlay new images atop the existing online images of the same piece of the sky; and to use user-provided narrative “tours” as guides. Most important, WWT surmounted its standalone capabilities, comprising part of “an ecosystem of online astronomy that will speed the progress of both ‘citizen’ and ‘professional’ science” [Goodman and Wong 2009, 41]. WWT’s potential uses in collaborative and educational initiatives appeared “truly limitless” [Goodman and Wong 2009, 42].

Finally, generally increased computational power enabled scalability and introduced new ways of handling, analyzing, and making accessible scientific datasets. Researchers could triage and identify unique objects, events, and data outliers and subsequently route them to citizen-scientist networks for verification. Citizen-scientists’ participation could be increased and enhanced through better-defined interfaces that rendered work into play. These three developments — crowdsourcing, democratized access, and increased computing power — were equally applicable to data-intensive research efforts in the humanities [Goodman and Wong 2009].

Capping more than a decade of evolving digital curation work, the first Digging into Data challenge (2009-2011) demonstrated the “promise of revelatory explorations of our cultural heritage that will lead us to new insights and knowledge, and to a more nuanced and expansive understanding of the human condition” [Willford and Henry 2012, 1]. The Office of Digital Humanities of the National Endowment for the Humanities (NEH-ODH), the National Science Foundation, the Joint Information Systems Committee (JISC), and the Canadian Social Sciences and Humanities Research Council (SSHRC) funded the eight projects. Digging into Data is likely the most important digital curation initiative yet attempted in the humanities; its projects augur well for synthesis of the recommendations and lessons of *Our Cultural Commonwealth*. Using Zotero and TAPOR on the Old Bailey Proceedings: Data Mining with Criminal Intent (DMCI); Digging into the Enlightenment: Mapping the Republic of Letters; Towards Dynamic Variorum Editions (DVE); Mining a Year of Speech; Harvesting Speech Datasets from the Web; Structural Analysis of Large Amounts of Music Information (SALAMI); Digging into Image Data to Answer Authorship Related Questions (DID-ARQ); and Railroads and the Making of Modern America — all showed “previously unimagined correlations between social and historical phenomena through computational analysis of large, complex data sets” [Willford and Henry 2012, 2].

All eight projects grappled with heterogeneous data corpora far larger than what could be exploited by an individual scholar. Additionally, all eight projects applied some form of computational analysis to their corpora, refined their tools and data periodically, and adopted similar research processes. Common concerns also earmarked the projects. Each team struggled with scarce funding, with managing time, with communication, and with the labor-intensive nature of

sharing data or with making it “diggable” or both.

On the other hand, differences arose among the projects. These differences stemmed from varying disciplinary traditions, from the choice of collaborators seemingly most suitable for particular data sets, from the proportion of manual to automated work, from the need for continual adaptation of analytical tools, and from the (un)likelihood of attaining major outcomes in only fifteen months.

55

Digging into Data awardees offered recommendations based on their project experiences. Once again, these recommendations reflected long running concerns and challenges, albeit in new and more sophisticated contexts. Digging into Data participants emphasized the need to increase incentives for collaborative and multidisciplinary work, especially for students and junior faculty, to establish standards for assessing such work, to nurture cross-disciplinary research tools and methods, to underwrite travel expenses, to facilitate inter-institutional sharing of hardware, software, and data, to clarify legal and ethical obligations, to encourage multi-institutional strategies for data management, to increase the range of publication options for data-rich and multimedia products, and to emphasize open access to research data.

56

Most important, the Digging into Data teams vividly showed the possibility of attracting new and larger audiences to digital humanities projects and indicated the emergence of new research avenues. Participants saw computers and their associated technologies as “a moveable and adjustable lens that allows scholars to view their subjects more closely, more distantly, or from a different angle than would be possible without it” [Willford and Henry 2012, 21]. Even so, they chose not to jettison more traditional disciplinary concerns, framing their work as “augmenting and transforming, rather than supplanting, research practice within their disciplines” [Willford and Henry 2012, 32]. Overall, however, it remains unclear to what extent researchers are posing new research questions as a result of the eight projects or to what extent the project have cultivated new or expanded audiences or both. The potential is there.

57

V.

By the end of the 2000s, digital curation stakeholders aiming to develop new research questions and expand audiences found themselves in an ambivalent position despite their considerable investment in curation and its concomitant payoff. Curation could appear a Sisyphean endeavor. Even in 2009, *Nature* inveighed against scientific data’s “shameful neglect” [Nature 2009]. Similarly, *Science* lamented that data-intensive scientific research had been “slow to develop due to the subtleties of databases, schemas, and ontologies, and a general lack of understanding of these topics by the scientific community” [Bell, Hey and Szalay 2009, 1298]. Despite these travails, the “most obvious and profound impact” of data-intensive research lay in the natural sciences [Ogburn 2010, 241]. By implication, then, digital humanists were hamstrung further.

58

For their part, digital humanists needed to supplant “boutique” projects with innovative collaborative strategies; the outstanding question was “whether and how to stimulate large-scale coherence without stymieing individual enterprise, frustrating existing self-organization, or threatening... individualism.” [Friedlander 2009, 12]. Indeed, one recent study suggested that collaboration was not proceeding as smoothly as hoped; it noted that “Although sharing with close, trusted collaborators happened regularly, sharing with anyone outside this inner circle, sometimes including other members of a project team, took place through ‘just in time’ negotiations” [Cragin, Palmer, Carlson and Witt 2010, 4036]. Too, researchers held “primarily speculative” views on sharing data with the public — most had shared only within collaborations or by request [Cragin, Palmer, Carlson and Witt 2010, 4036]. Last, the data most commonly shared were those either easiest to share or the most “presentable” — but not always those most valuable for curation, particularly for researchers in other disciplines [Cragin, Palmer, Carlson and Witt 2010]. Clearly much work remains to be done in delineating the best mechanisms for sharing.

59

Reports released in 2009 and 2010 highlighted both advances and continuing challenges. The National Academy of Science concluded that researchers were in fact using data to probe new research questions. Simulations could steer theoretical approaches or validate new experimental ones; interdisciplinary and international teams could capitalize on myriad intellectual perspectives; and scholars could use data generated by others to supplement their own data or to address research questions earlier researchers did not. Such approaches could benefit researchers in the humanities

60

as well as those in the sciences. According to the Blue Ribbon Task Force on Sustainable Digital Preservation and Access, however, obdurate challenges for digital curation stakeholders such as time considerations, diffused stakeholders, misaligned or weak incentives, and lack of clearly defined roles and responsibilities persisted [Berman et al 2010].

The National Academy of Science report's recommendations, however, reiterated familiar priorities — priorities applicable to curation in the humanities as well as in the sciences. The report foregrounded data integrity, proper training, professional standards developed consensually, appropriate recognition for contributions, public accessibility of data and results, data sharing, clear policies regarding management of and access to data, and the importance of data management plans developed at the project's inception [National Academy of Science 2009]. Similarly, the Blue Ribbon Task Force urged stakeholders to make the case for use, to create incentives to preserve data in the public interest, and to define explicitly stakeholder roles and responsibilities throughout the lifecycle not only to ensure the efficient use of resources, but also to minimize free riding [Berman et al 2010].

61

Ultimately, the overlapping digital humanities and digital curation communities must collaborate even more extensively in the future among themselves and among other professional communities such as librarians, curators, and archivists and experts in law, business, and science. Such collaborations must traverse geographical, disciplinary, and institutional boundaries. Indeed, the United States federal government should serve “as a reliable and transparent partner and as a coordinating entity,” as should the government in the United Kingdom [Interagency Working Group on Digital Data 2009, 16].

62

Ideally, a symbiotic and even synergistic partnership will mature between digital curation and digital humanities. This partnership must be nurtured both top-down and bottom-up. All the same, stakeholders must remember that “collaborative approaches are far from a panacea; success requires good faith and investment from all the players” [Repository Task Force 2009, 25]. In this vein, digital curation projects have been developed at an “alarmingly fast rate, producing a useful but bewildering array of theoretical frameworks, diagrams, software and services” [Prom 2011, 142]. Nor can stakeholders afford not to engage with the human factor. As Gunther Weibel contends, “The social engineering of incentives and services will be as critical to success as the business models and cost structures” [Weibel 2009].

63

At the highest level, stakeholders must focus on long-term sustainability. “Sustainability is not merely about money; it is about organizational commitment and the ability to build persistent collaborations to address the ongoing needs for repository services and infrastructure” [Repository Task Force 2009, 8]. Long-term sustainability in turn hinges on policies and planning and on compliance. The National Science Foundation's and National Institute of Health's policies for data planning constitute a “major strategic move”; on the other hand, planning requirements are not particularly specific and provisions for accountability remain nebulous [Buckland 2011, 34]. As Paul Schofield and his colleagues (2009) point out, “It is one thing to encourage data deposition and resource sharing through guidelines and policy statements, and quite another to ensure that it happens in practice” [Schofield et al 2009, 171].

64

Such high-level concerns notwithstanding, at the grassroots digital curation is also a pressing concern. Perhaps most important in addressing technological issues and the human factor in tandem is education and training. Professionals engaged in digital curation often end up in these roles by accident and thus tend to “skill up” on the job. Ideally, digital curation professionals have “a research background together with a technical aptitude and finely-tuned advocacy and interpersonal skills” [Swan and Brown 2008, 28]. As Youngseek Kim and his colleagues (2011) observe, a “significant demand will arise for individuals with eScience professional skills in terms of data curation and cyberinfrastructure, that numerous other institutions of higher education will need to join the process of educating them, and that a significantly expanded supply of students to join these programs will be required” [Kim, Addom and Stanton 2011, 134–135]. Indeed, debate continues over the feasibility of integrating digital curation skills into undergraduate curricula [Swan and Brown 2008]. Of the 58 accredited Library and Information Science programs in North America, moreover, merely 13 (22%) offer one or more courses in data management or curation [Creamer et al 2012]. Furthermore, approximately half of these data-related courses are offered only online. Suffice it to say, LIS graduate programs have a substantial opportunity to engage more aggressively with data curation as a lodestone of the curriculum [Creamer et al 2012].

65

These educational endeavors facilitate the spread of digital curation initiatives, which have clustered in a handful of research universities. But these universities constitute only 297 of 1,832 four-year institutions; therefore, stakeholders have an opportunity to integrate curation education into Master's and Baccalaureate institutions [Shorish 2012]. Yasmeen Shorish enjoins, "Smaller institutions can engage with data curation on some level, however minimal, to ensure that the research data of teaching institutions are not lost or hidden" [Shorish 2012, 271]. Liberal arts colleges may prove well-suited for digital humanities and digital curation projects [Green and Roy 2008]; [Pannapacker 2013].

66

But even research universities such as the University of Minnesota and Cornell University struggle still with operationalizing digital curation. A "large unmet need" for assistance with data curation persists [Johnston, Lafferty, and Petsan 2012, 79]. In late 2010, Minnesota inaugurated a workshop on data management planning for grant applications. Scalable and flexible, the workshop exerted an "overwhelmingly positive impact" [Johnston, Lafferty, and Petsan 2012, 85]. Meanwhile, a full 62% of National Science Foundation Principal Investigators at Cornell wanted assistance in crafting their data management plans. Gail Steinhart and her associates determined "a great deal of uncertainty among PIs about what the new NSF requirement means and how to meet it, and that researchers welcome offers of assistance — both with data management planning, and with specific components or data management NSF asks them to address in their plans" [Steinhart et al. 2012, 77].

67

Campus libraries have a pivotal role to play in educating researchers about curation. They must evolve into "vibrant knowledge branches that reach throughout their campuses to provide curatorial guidance and expertise for digital content" [Walters 2009, 5]. Resembling numerous other institutions wrestling with the creation and implementation of systematic and active curation programs, the Georgia Institute of Technology has found its progress "incremental and characterized by the reallocation of existing library resources to data curation" [Walters 2009, 91]. More specifically, librarians' roles vis-à-vis digital curation will embrace three broad areas. First, as part of a national infrastructure including research libraries, government bodies, professional organizations, and industry, librarians will help establish national curation strategies that include economic models and that will remain viable over the long-term. Second, a robust campus infrastructure will depend on resources created by research library leaders collaborating with campus information technology leaders. Third, librarians will spearhead professional development and education [Gold 2010]. In short, libraries and librarians can increase awareness of digital curation's importance, can provide archiving and preservation services through institutional repositories, and can develop new professional practices suitable for data librarianship [Swan and Brown 2008].

68

Like libraries, institutional repositories, archives, and centers show great leadership potential. The Distributed Data Curation Center of Purdue's University Library, for instance, "integrate[s] librarians and the principles of library and archival sciences with domain sciences, computer and information sciences, and information technology to address the challenges of managing collections of research data and to learn how to better support interdisciplinary research through data curation" [Witt 2009, 191]. Similarly, archives, particularly in tandem with institutional repositories, should be at the forefront of curation education and practice [Prom 2011]. Not to be overlooked, the Digital Curation Center continues to break new ground with the assistance it offers stakeholders, for example with its recent "5 Steps to Research Data Readiness" [Miller 2012]. Overall, campus-wide initiatives or centers or partnerships with domain researchers, computer scientists, and campus information technology at Cornell University, Purdue University, the Massachusetts Institute of Technology, the University of Minnesota, the University of Massachusetts, and the University of Virginia have flourished [Gold 2010]. As Michael Witt (2009) concludes, "a critical mass of similar data that is archived and shared in one place can become fertile ground for the congregation of virtual communities and the emergence of shared tools and formats — perhaps even new standards for interoperability — as researchers come together to use the data and contribute their own data to the collection" [Witt 2009, 194–195]. Digital humanists and digital curators, take note.

69

Ultimately, it remains unclear when a critical mass of case study evidence will be assembled to address these stubborn concerns. How much data has been shared? How much has been reused? What specific audiences have been cultivated and what research questions have been developed? Regardless of what has been done or not done with digital humanities data, digital curation will be indispensable in securing such digital assets for the indefinite future. Stakeholders must avoid the digital humanities community learning about the seminal importance of digital curation only

70

after losses and the hard lessons such losses impart. After all, “Reaching out to determine what data are generated and whether it should be curated requires a cooperative audience and time but no additional infrastructure or financial investment” [Shorish 2012, 270].

In 2009, Christine Borgman asserted that “Digital content, tools, and services all exist, but they are not necessarily useful or usable” [Borgman 2009]. Despite obvious progress in digital curation in the humanities, she issued a “call to action” to stakeholders and insisted the “future is now.” Three years later, we may — we must — ask the same question, lest we are reduced ultimately to exclaiming, along with Michael Buckland, “What a waste!” [Buckland 2011, 35].

71

Acknowledgments

Sincere thanks to Helen R. Tibbo of the University of North Carolina at Chapel Hill, the staff of *Digital Humanities Quarterly*, and the *Quarterly's* anonymous reviewers.

72

Notes

- [1] Digital Curation Center Glossary: <http://www.dcc.ac.uk/digital-curation/glossary#D>
- [2] Digital Curation Centre: <http://www.dcc.ac.uk/digital-curation/what-digital-curation>
- [3] National Science Foundation Dissemination and Sharing of Research Results: <http://www.nsf.gov/bfa/dias/policy/dmp.jsp>
- [4] Data Management Plans for NEH Office of Digital Humanities Proposals and Awards: http://www.neh.gov/files/grants/data_management_plans_2012.pdf
- [5] Data Management Plans for NEH Office of Digital Humanities Proposals and Awards: http://www.neh.gov/files/grants/data_management_plans_2012.pdf
- [6] Digital Humanities Data Curation: <http://www.neh.gov/divisions/odh/institutes/digital-humanities-data-curation>
- [7] Roy Rosenzweig Center for History and New Media: <http://chnm.gmu.edu/>
- [8] Sustainable Digital Data Preservation and Access Network Partners (DataNet): <http://www.nsf.gov/pubs/2007/nsf07601/nsf07601.htm>
- [9] The DataNet Partners: Sharing Science, Linking Domains, Curating Data: <http://www.asis.org/Conferences/AM09/panels/41.pdf>

Works Cited

- Abbott, Jones and Ross 2008** Abbott, D., S. Jones and S. Ross. “Curating Digital Records of Performance”. Presented at *CIDOC (2008)*. http://network.icom.museum/fileadmin/user_upload/minisites/cidoc/ConferencePapers/2008/67_papers.pdf.
- Alvarado 2011** Alvarado, R. *The Digital Humanities Situation*. 2011. <http://transducer.ontoligent.com/?p=717>.
- American Council of Learned Societies 2006** Unsworth, John, et al. *Our Cultural Commonwealth: The Report of the American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities and Social Sciences. The ACLS Report*. 2006. <http://www.acls.org/cyberinfrastructure/ourculturalcommonwealth.pdf>.
- Arms 2008** Arms, W.Y. “Cyberscholarship: High Performance Computing Meets Digital Libraries”. *Journal of Electronic Publishing* 11: 1 (2008).
- Arms and Larsen 2007** Warning: Biblio formatting not applied. WilliamArms. RonaldLarsen. *The Future of Scholarly Communication: Building the Infrastructure for Cyberscholarship. Report of the NSF/JISC workshop on scholarly information repositories*. National Science Foundation and Joint Information Systems Committee. 2007.
- Atkins 2003** Warning: Biblio formatting not applied. Daniel E. Atkins. *Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure. The Atkins Report*. Arlington. National Science Foundation. 2003. <http://www.nsf.gov/od/oci/reports/atkins.pdf>.
- Baker and Yarmey 2009** Baker, K.S., and L. Yarmey. “Data Stewardship: Environmental Data Curation and a Web-of-Repositories”. *International Journal of Digital Curation* 4: 2 (2009), pp. 1-16.

- Beagrie 2004** Beagrie, N. "The Digital Curation Centre". *Learned Publishing* 17: 1 (2004), pp. 7-9.
- Beagrie 2006** Beagrie, N. "Digital Curation for Science, Digital Libraries, and Individuals". *International Journal of Digital Curation* 1: 1 (2006), pp. 4-16.
- Becker 2009** Becker, C., et al. *Systematic Planning for Digital Preservation: Evaluating Potential Strategies and Building Preservation Plans*. 2009. <http://www.ifs.tuwien.ac.at/~becker/pubs/becker-ijdl2009.pdf>.
- Bell, Hey and Szalay 2009** Bell, G., T. Hey and A. Szalay. "Beyond the Data Deluge". *Science* (March 6 2009), pp. 1297-1298.
- Berman et al 2010** Berman, F., et al. *Sustainable Economies for a Digital Planet: Ensuring Long-Term Access to Digital Information*. Blue Ribbon Task Force on Sustainable Digital Preservation and Access, 2010.
- Blanke, Hedges and Dunn 2009** Blanke, T., M. Hedges and S. Dunn. "Arts and Humanities e-Science--Current Practices and Future Challenges". *Future Generation Computer Systems* 25: 4 (2009), pp. 474-480.
- Bollier 2011** Bollier, D. *The Promise and Peril of Big Data*. 2011. <http://dx.doi.org/10.1080/1369118X.2012.678878>.
- Borgman 2009** Borgman, Christine L. "The Digital Future is Now: A Call to Action for the Humanities". *Digital Humanities Quarterly* 3: 4 (2009). <http://www.digitalhumanities.org/dhq/vol/3/4/000077/000077.html>.
- Borgman 2012** Borgman, C. "The Conundrum of Sharing Research Data". *Journal of the American Society for Information Science and Technology* 63: 6 (2012), pp. 1059-1078.
- Buckland 2011** Buckland, M. "Data Management as Bibliography". *Bulletin of the American Society for Information Science and Technology* 37: 6 (2011), pp. 34-37.
- Carlson and Anderson 2007** Carlson, S., and B. Anderson. "What are Data? The Many Kinds of Data and Their Implications for Data Re-use". *Journal of Computer-Mediated Communication* 12: 2 (2007), pp. 301-317.
- Carlson, Fosmire, Miller and Nelson 2011** Carlson, J., M. Fosmire, C. Miller and M.S. Nelson. "Determining Data Information Literacy Needs: A Study of Students and Research Faculty". *portal: Libraries and the Academy* 11: 2 (2011), pp. 629-657.
- Chin and Lansing 2004** Chin, G., and C.S. Lansing. "Capturing and Supporting Contexts for Scientific Data Sharing via the Biological Sciences Collaboratory". Presented at ACM (2004).
- Choudhury 2010** Choudhury, S. "Data Curation: An Ecological Perspective". *College and Research Libraries News* 71: 4 (2010), pp. 194-196.
- Cohen 2012** Cohen, D. "The Social Contract of Scholarly Publishing". In M.K. Gold, ed., *Debates in the Digital Humanities*. Minneapolis: University of Minnesota Press, 2012. pp. 319-321.
- Constantopoulos and Dallas 2007** Constantopoulos, P., and C. Dallas. *Aspects of a Digital Curation Agenda for Digital Heritage*. 2007. http://panteion.academia.edu/CostisDallas/Papers/967349/Aspects_of_a_Digital_Curation_Agenda_for_Cultural_Heritage.
- Cragin, Palmer, Carlson and Witt 2010** Cragin, M., C.L. Palmer, J.R. Carlson and M. Whitt. "Data Sharing, Small Science and Institutional Repositories". *Transactions of the Royal Society* 368: 1926 (2010), pp. 4023-4038.
- Creamer et al 2012** Creamer, A.T., et al. "A Sample of Research Data Curation and Management Courses". *Journal of eScience Librarianship* 1: 2 (2012), pp. 88-96.
- Crosas 2011** Crosas, M. *The Dataverse Network: An Open-Source Application for Sharing, Discovering, and Preserving Data*. 2011. <http://www.dlib.org/dlib/january11/crosas/01crosas.html>.
- Crow 2002** Crow, R. *The Case for Institutional Repositories: A SPARC Position Paper*. Washington D.C.: The Scholarly Publishing and Academic Resources Coalition, 2002.
- Economic and Social Research 2010** Economic and Social Research Council. *Research Data Policy*. Swindon: Economic and Social Research Council, 2010.
- Eynden, Bishop, Horton and Corti 2010** Eynden, V.V.d., L. Bishop, L. Horton and L. Corti. *Data Management Practices in the Social Sciences*. Essex: UK Data Archive, 2010.
- Flanders 2012** Flanders, Julia. "Time, Labor, and 'Alternate Careers' in Digital Humanities Knowledge Work". In M.K. Gold, ed., *Debates in the Digital Humanities*. Minneapolis: Minnesota University Press, 2012. pp. 292-308.

- Flanders, Piez and Terras 2007** Flanders, Julia, Wendell Piez and Melissa Terras. "Welcome to Digital Humanities Quarterly". *Digital Humanities Quarterly* 1: 1 (2007). <http://digitalhumanities.org/dhq/vol/1/1/000007/000007.html>.
- Fraistat 2012** Fraistat, N. "The Function of Digital Humanities Centers at the Present Time". In M.K. Gold, ed., *Debates in the Digital Humanities*. Minneapolis: University of Minnesota Press, 2012. pp. 281-291.
- Friedlander 2009** Friedlander, Amy. "Asking Questions and Building a Research Agenda for Digital Scholarship". In *Working Together or Apart: Promoting the Next Generation of Digital Scholarship*. pp. 1-15. <http://www.clir.org/pubs/reports/reports/pub145/pub145.pdf>.
- Gold 2010** Gold, A. *Data Curation and Libraries: Short-Term Developments, Long-Term Prospects*. San Luis Obispo: California Polytechnic State University, 2010.
- Goodman and Wong 2009** Goodman, A., and C. Wong. "Bringing the Night Sky Closer: Discoveries in the Data Deluge". In T. Hey S. Tansley and K. Tolle, eds., *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Redmond: Microsoft, 2009. pp. 39-44.
- Green and Roy 2008** Green, D., and M. Roy. "Things to Do While Waiting for the Future to Happen: Building Cyberinfrastructure for the Liberal Arts". *EDUCAUSE Review* (2008), pp. 35-48.
- Hank and Davidson 2009** Hank, C., and J. Davidson. *International Curation Education Education (IDEA) Working Group: A Report from the Second Workshop of the IDEA*. 2009. <http://www.dlib.org/dlib/march09/hank/03hank.html>.
- Hey, Tansley and Tolle 2009** Hey, T., S. Tansley and K. Tolle. "Jim Gray on eScience: A Transformed Scientific Method". In T. Hey S. Tansley and K. Tolle, eds., *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Redmond: Microsoft, 2009.
- Higgins 2008** Higgins, S. "The DCC Curation Lifecycle Model". *International Journal of Digital Curation* 3: 1 (2008), pp. 134-140.
- Higgins 2011** Higgins, S. "Digital Curation: The Emergence of a New Discipline". *International Journal of Digital Curation* 6: 2 (2011), pp. 78-88.
- Hockx-Yu 2007** Hockx-Yu, H. "Digital Curation Centre--Phase Two". *International Journal of Digital Curation* 2: 1 (2007), pp. 123-127.
- Interagency Working Group on Digital Data 2009** Interagency Working Group on Digital Data. *Harnessing the Power of Digital Data for Science and Society*. Washington D.C.: National Science and Technology Council, 2009.
- Johnston, Lafferty, and Petsan 2012** Johnston, L., M. Lafferty and B. Petsan. "Training Researchers on Data Management: A Scalable, Cross-Disciplinary Approach". *Journal of eScience Librarianship* 1: 2 (2012), pp. 79-87.
- Karasti, Baker and Halkola 2006** Karasti, H., K.S. Baker and E. Halkola. "Enriching the Notion of Data Curation in E-Science: Data Managing and Information Infrastructuring in the Long Term Ecological Research (LTER) Network". *Computer Supported Cooperative Work* 15: 4 (2006), pp. 321-358.
- Kim, Addom and Stanton 2011** Kim, Y., B.K. Addom and J.M. Stanton. "Education for eScience Professionals: Integrating Data Curation and Cyberinfrastructure". *International Journal of Digital Curation* 6: 1 (2011), pp. 125-138.
- King 2007** King, G. "A Introduction to the Dataverse Network as an Infrastructure for Data Sharing". *Sociological Methods and Research* 36: 2 (2007), pp. 173-199.
- Kirschenbaum 2010** Kirschenbaum, Matthew G. *What is Digital Humanities and What's It Doing in English Departments?* 2010. <http://mkirschenbaum.files.wordpress.com/2011/03/ade-final.pdf>.
- Kornbluh 2008** Kornbluh, M. "From Digital Repositories to Information Habitats: H-Net, the Quilt Index, Cyber Infrastructure, and Digital Humanities". *First Monday* 13: 8 (2008). <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/rt/prinFRIENDLY/2230/2019>.
- Lee and Tibbo 2007** Lee, C.A., and H.R. Tibbo. *Digital Curation and Trusted Repositories: Steps Toward Success*. 2007. <http://journals.tdl.org/jodi/index.php/jodi/article/view/229/183>.
- Lesk 2010** Lesk, M. *Data Curation: Just in Time or Just in Case?* West Lafayette: International Association of Scientific and Technological University Libraries, 2010.
- Lord and Macdonald 2003** Lord, P., and A. Macdonald. *Data Curation for e-Science in the UK: An Audit to Establish Requirements for Future Curation and Provision*. Twickenham: JISC Committee for the Support of Research (JCSR), 2003.

- Lynch 2003** Lynch, Clifford A. "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age". *ARL Bimonthly Report* 226 (2003).
- Lynch 2008** Lynch, C. "Big Data: How Do Your Data Grow?". *Nature* (September 4 2008), pp. 28-29.
- Lyon, Rusbridge, Neilson and Whyte 2009** Lyon, L., C. Rusbridge, C. Neilson and A. Whyte. *Disciplinary Approaches to Sharing, Curation, Reuse and Preservation*. Edinburgh: Digital Curation Centre, 2009.
- McCarty 2012** McCarty, W. "A Telescope for the Mind?". In M.K. Gold, ed., *Debates in the Digital Humanities*. Minneapolis: University of Minnesota Press, 2012. pp. 113-123.
- Miller 2012** Miller, K. *5 Steps to Research Data Readiness*. Edinburgh: Digital Curation Centre, 2012.
- Myers et al 2005** Myers, J., et al. "A Collaborative Informatics Infrastructure for Multi-scale Science". *Cluster Computing* 8: 4 (2005), pp. 243-253.
- National Academy of Science 2009** National Academy of Science. *Ensuring the Integrity, Accessibility, and Stewardship of Research Data*. Washington D.C.: National Academy of Science, 2009.
- National Science Board 2005** National Science Board. *Long-Lived Data Collections*. Arlington: National Science Board, 2005.
- National Science Foundation 2005** National Science Foundation. *NSF's Cyberinfrastructure Vision for 21st Century Discovery*. Arlington: National Science Foundation, 2005.
- Nature 2009** Nature. "Data's Shameful Neglect". *Nature* (September 10 2009), pp. 145-145.
- Ogburn 2010** Ogburn, J. "The Imperative for Data Curation". *portal: Libraries and the Academy* 10: 2 (2010), pp. 241-246.
- Pannapacker 2013** Pannapacker, W. "Stop Calling It Digital Humanities and 9 Other Strategies to Help Liberal-Arts Colleges Join the Movement". *Chronicle of Higher Education* (February 18 2013).
- Piez 2011** Pietz, W. "Impractical Applications". *Digital Humanities Quarterly* 5: 1 (2011).
<http://digitalhumanities.org/dhq/vol/5/1/000095/000095.html>.
- Prom 2011** Prom, C. "Making Digital Curation a Systematic Institutional Function". *International Journal of Digital Curation* 6: 1 (2011), pp. 139-152.
- Pryor 2009** Pryor, G. "Multi-Scale Data Sharing in the Life Sciences: Some Lessons for Policy Makers". *International Journal of Digital Curation* 4: 3 (2009), pp. 71-82.
- Repository Task Force 2009** Repository Task Force. *The Research Library's Role in Digital Repository Services: The Final Report of the ARL Digital Repository Issues Task Force*. Washington D.C.: Association of Research Libraries, 2009.
- Research Information Network and the British Library 2009** Research Information Network and the British Library. *Patterns of Information Use and Exchange: Case Studies of Researchers in the Life Sciences*. London: British Library, 2009.
- Rusbridge 2007** Rusbridge, C. "Introduction". In *Create, Curate, Re-use: The Expanding Life Course of Digital Research*. Educause Australia, 2007. pp. 1-11.
- Rusbridge et al 2005** Rusbridge, C., et al. "The Digital Curation Centre: A Vision for Digital Curation". *IEEE Computer Society* (2005), pp. 31-41.
- Schofield et al 2009** Schofield, P., et al. "Post-Publication Sharing of Data and Tools". *Nature* (September 10 2009), pp. 171-173.
- Shorish 2012** Shorish, Y. "Data Curation is for Everyone! The Case for Master's and Baccalaureate Institutional Engagement with Data Curation". *Journal of Web Librarianship* 6: 4 (2012), pp. 263-273.
- Spiro 2012** Spiro, Lisa. "This is Why We Fight". In Matthew K. Gold, ed., *Debates in the Digital Humanities*. Minneapolis: University of Minnesota Press, 2012. pp. 16-35.
- Steinhart et al. 2012** Steinhart, G., et al. "Prepared to Plan? A Snapshot of Researcher Readiness to Address Data Management Planning Requirements". *Journal of eScience Librarianship* 1: 2 (2012), pp. 63-78.
- Stodden 2009** Stodden, V. "The Legal Framework for Reproducible Scientific Research: Licensing and Copyright".

Computing in Science and Engineering 11: 1 (2009), pp. 35-40.

Svensson 2010 Svensson, Patrik. "The Landscape of Digital Humanities". *Digital Humanities Quarterly* 4: 1 (2010).
<http://digitalhumanities.org/dhq/vol/4/1/000080/000080.html>.

Svensson 2012 Svensson, Patrik. "Envisioning the Digital Humanities". *Digital Humanities Quarterly* 6: 2 (2012).
<http://www.digitalhumanities.org/dhq/vol/6/1/000112/000112.html>.

Swan and Brown 2008 Swan, A., and S. Brown. *The Skills, Role and Career Structure of Data Scientists and Curators: An Assessment of Current Practices and Future Needs*. Truro: Key Perspectives Ltd., 2008.

Uhlir 2010 Uhlir, P. "Information Gulags, Intellectual Straightjackets, and Memory Holes: Three Principles to Guide the Preservation of Scientific Data". *Data Science Journal* 9 (2010), pp. ES1-ES5.

Unsworth 2009 Unsworth, J. "The Making of Our Cultural Commonwealth". *Digital Humanities Quarterly* 3: 4 (2009).
<http://www.digitalhumanities.org/dhq/vol/3/4/000073/000073.html>.

Vardigan and Whiteman 2007 Vardigan, M., and C. Whiteman. "ICPSR Meets OAIS: Applying the OAIS Reference Model to the Social Science Archive Context". *Archival Science* 7: 1 (2007).

Wallis, Borgman, Mayernik and Pepe 2008 Wallis, J., C. Borgman, M. Mayernik and A. Pepe. "Moving Archival Practices Upstream: An Exploration of the Life Cycle of Ecological Sensing Data in Collaborative Field Research". *International Journal of Digital Curation* 1: 3 (2008), pp. 114-126.

Walters 2009 Walters, T.O. "Data Curation Program Development in U.S. Universities: The Georgia Institute of Technology Example". *International Journal of Digital Curation* 4: 3 (2009), pp. 83-92.

Walters and Skinner 2011 Walters, T., and K. Skinner. *New Roles for New Times: Digital Curation for Preservation*. Washington D.C.: Association of Research Libraries, 2011.

Waters 2007 Waters, D.J. "Doing Much More Than We Have So Far Attempted". *Educause Review* 42: 5 (2007), pp. 8-9.

Weibel 2009 Weibel. *Weibel Lines Blog*. 2009. <http://weibel-lines.typepad.com/weibelines/2009/03/are-data-repositories-the-new-institutional-repositories.html>.

Whitlock 2010 Whitlock, M.C., et al. "Data Archiving". *The American Naturalist* (2010), pp. 145-146.

Whyte and Wilson 2010 Whyte, A., and A. Wilson. *How to Appraise and Select Research Data for Curation*. 2010.
<http://www.dcc.ac.uk/sites/default/files/documents/How%20to%20Appraise%20and%20Select%20Research%20Data.pdf>.

Willford and Henry 2012 Willford, C., and C. Henry. *One Culture: Computationally 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 on Library and Information Resources, 2012.

Witt 2009 Witt, M. "Institutional Repositories and Research Data Curation in a Distributed". *Library Trends* 57: 2 (2009), pp. 191-201.

Yakel 2007 Yakel, E. "Digital Curation". *OCLC Systems and Services* 23: 4 (2007), pp. 338-339.

Zimmerman 2008 Zimmerman, A. "New Knowledge from Old Data: The Role of Standards in the Sharing and Reuse of Ecological Data". *Science, Technology & Human Values* 35: 5 (2008), pp. 631-652.

Zorich 2009 Zorich, D.M. *Working Together or Apart: Promoting the Next Generation of Digital Scholarship*. Washington D.C.: Council on Library and Information Resources, 2009.

boyd and Crawford 2012 boyd, d., and K. Crawford. "Critical Questions for Big Data". *Information, Communication, and Society* 15: 5 (2012), pp. 662-679.



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.