Foundations of Data Curation: The Pedagogy and Practice of "Purposeful Work" with Research Data

Carole L. Palmer Professor, Graduate School of Library and Information Science; Director, Center for Informatics Research in Science & Scholarship (CIRSS) – University of Illinois at Urbana-Champaign

Nicholas M. Weber PhD Student, Graduate School of Library and Information Science – University of Illinois at Urbana-Champaign

Trevor Muñoz Associate Director, Maryland Institute for Technology in the Humanities (MITH); Assistant Dean of Digital Humanities Research, University Libraries – University of Maryland

Allen H. Renear Interim Dean and Professor, Graduate School of Library and Information Science – University of Illinois at Urbana- Champaign

Abstract

Increased interest in large-scale, publicly accessible data collections has made data curation critical to the management, preservation, and improvement of research data in the social and natural sciences, as well as the humanities. This paper explicates an approach to data curation education that integrates traditional notions of curation with principles and expertise from library, archival, and computer science. We begin by tracing the emergence of data curation as both a concept and a field of practice related to, but distinct from, both digital curation and data stewardship. This historical account, while far from definitive, considers perspectives from both the sciences and the humanities. Alongside traditional LIS and archival science practices, unique aspects of curation have informed our concept of "purposeful work" with data and, in turn, our pedagogical approach to data curation for the sciences and the humanities.

The Genesis of *Data Curation*

Natural history museums have long been devoted to the curation of scientific data in the form of physical specimens. Their curatorial roles and responsibilities help explain how the term *curation* came to stand for the current conception of the management and preservation of digital research data. Moreover, their focus on making data usable for research over the long term is tightly aligned with our approach to data curation education, to be discussed below, which has been driven by the principle of "purposeful curation."[1]

Before scholarly practices shifted to a digital realm or a big data paradigm, natural history museums were extending their concept of curation in anticipation of the demand for the management and enhancement of digital data. In particular, William Fry, a curator at the British Museum of Natural History, made very early observations about the need to optimize data collections for research in a digital environment. His 1965 *Nature* article, "Methods in Taxonomy," asserts, "what is urgently needed, if numerical taxonomy is to become an accepted tool of routine identification and classification, is proof that the labour of collecting the vast quantities of data required for the statistics is rewarded by greatly increased usefulness of the results."[2]

Fry questioned how "mathematical and computational" methods would affect the practices of curators and collection managers, forecasting the kinds of concerns we hear today about the data deluge: "While collections, the items of which can be counted in hundreds, can be adequately controlled and used by means of written guides and indices, these time-honoured methods cannot be efficient where collections must be numbered in tens or hundreds of thousands."[3]

The notion of *curation* in relation to digital research data became more common in the 1990s, although most of what we now think of as *data curation* was often labeled *informatics* in both the humanities and the sciences. A 1994 Department of Energy (DOE) report on genome informatics is notable in its attention to data access and management, as well as its description of "database curators" in both laboratories and natural history museums. However, the first published application of the term *data curation* may be found in Diane Zorich's 1995 paper on future collections management in *Museum Management and Curatorship*. She elaborates on the concept of database curators from the 1994 DOE report, arguing for an entirely new field of information work in museums, libraries, and traditional scientific laboratory settings:

> *Data sets need to be examined for consistency, long-term quality and relevance over time, and new sources of data must be identified and assessed. Changes or updates to data require authentication and verification. Tools which support object databases, such as authority lists, thesauri, data dictionaries and other documentation resources, need to be maintained, updated and distributed at regular intervals, while data security and access must be considered. All these concerns constitute*

the discipline of data curation.[4]

In the humanities and cultural heritage communities, the curation of digital research data has been an important topic since at least the late 1980s. Again, the museum community, especially those engaged with the early web, was at the center of curation discourse. In 1987, in the first issue of *Archival Informatics Newsletter* (the external house organ for David Bearman's company), Bearman explicitly drew the link between informatics as "an emerging usage in biomedicine" and "a philosophy of looking at the cultural information missions of archives and museums."[5] In 1993, the second International Conference on Hypermedia and Interactivity in Museums (ICHIM) featured a paper by Susan Hockey, then Director of the Center for Electronic Texts in the Humanities (CETH) at Rutgers and Princeton Universities, reporting on the Text Encoding Initiative (TEI)

standard for machine-readable texts initiated that year. Hockey explains the significance of efforts like the TEI to the goals of those working on "museum documentation and information handling systems," in part because of the need for "the usefulness, development and longevity of . . . data."[6]

In the same year that Zorich's article coined the term "data curation," the Joint Information Systems Committee (Jisc)[7] in the UK helped to establish the Arts and Humanities Data Service (AHDS), with the following mission: it "acquires, curates, preserves and provides access to complex digital resources created by or supporting research and teaching in Higher and Further Education and life-long learning."[8] Shortly thereafter the aims of the AHDS were further refined to "collect, describe, and preserve the electronic resources which result from scholarly research in the humanities" and to further "develop a generalized and extensible framework for digital resource creation, description, preservation, and location."[9]

Diffuse Trajectory in the Humanities

Initiatives in the humanities, such as the AHDS, were undoubtedly responding to data-curation imperatives similar to those in bioinformatics and other sciences. Clear parallels are the exception, however, in a longer and more discontinuous and diffuse trajectory as the *digital humanities* progressed. Through 2001, digital research in the humanities was sometimes referred to as humanities informatics,[10] but *humanities computing* became the more common term for the growing field now known as digital humanities.

Several significant strands of early computational research in humanities computing involved the development of indices, annotated linguistic corpora, and digitally encoded texts—in other words, the preparation, collection, organization, and maintenance of datasets. The widely accepted origin story of

the digital humanities as a field centers on Father Roberto Busa's project, beginning in 1949, to create a concordance to the complete works of Thomas Aquinas.[11] Busa's reflection on the genesis of his own work contains an echo of Fry's call for a curatorial response to "mathematical and computational" methods—with the challenge being the accelerated production of textual data rather than biological specimens. As Busa states, it was clear "that to process texts containing more than ten million words, I had to look for some machinery."[12]

Two decades of work in concordance and index building followed from Busa's initial achievement, as documented by Dolores M. Burton,[13] and a cognate track of scholarship was devoted to digital dictionaries and corpora, or "electronic lexicography."[14] Databases became important in the 1980s and 1990s, evidenced by the 1992 inaugural CETH seminar devoted to a project on "Building a Humanities Database."[15] Although less pronounced than with genomics or the natural museum community, the value of databases and their many applications to scholarship were beginning to be recognized.

Compared to the science realm where curation activities generally remained distinct from research, humanities computing encompassed work with data that was curatorial in nature. The focus of the research community, however, tended toward information retrieval, processing, and publishing more than data curation, even with the earliest case, when Busa referred to the need "to process" data for the Aquinas concordance. There was no comparable modern tradition of stewardship or collection management as with natural history museums. Thus, humanists did not adopt the terminological framework of curation for their work with data. Instead, they applied vocabulary familiar from philology, textual editing, quantitative history, and linguistics. In digital

literary studies, the development of methods for digitally encoding textual variants, for example, demonstrates an engagement with curatorial concerns such as provenance but using discipline-specific terms like "diplomatic" transcription strategies, and "editions."[16] Many of the debates over findings from quantitative history related to the sufficiency or insufficiency of information collections created by scholars pursuing those analyses and to ideas of historical "evidence."[17]

While the principles and many processes involved in data curation are common across the science and humanities, there are still important differences in how and when curation work is integrated into, or provided as, a service for research. As with other areas of digital scholarship, flagship initiatives are often harbingers of future trends. The highly visible AHDS lost funding in 2008. Since that time, data curation in the humanities returned to being a more diffuse, *ad hoc* enterprise rather than a part of governmental and institutional research programs, as is more typical in the sciences.[18]

Continuity in the Sciences

The 2002 paper, "Online Scientific Data Curation, Publication, and Archiving," by computer scientist Jim Gray and collaborators was instrumental in projecting the term data curation beyond the museum and bioinformatics domains. Gray and his colleagues were developing architecture and storage environments for the Sloan Digital Sky Survey in astronomy, and they had also begun writing about a new paradigm of scientific research associated with the emergence of grid computing and a deluge of digital data.[19] The "fourth paradigm" referred to a new data-driven method for scientific inquiry, where analysis of expertly curated, open access data collections would complement and extend traditional methods like simulation, observation, and theoretical experimentation.[20] Data curation, central to this vision, was described as the annotation, preservation, and expert description of datasets, to be carried out, by and large, by librarians.[21]

At the same time, the UK was embracing the potential digital transformation of a national research agenda. What had been called an informatics approach in North America and continental Europe was being labeled *eResearch* and *eScience* in the UK.[22] A 2003 Jisc-commissioned report on the infrastructures needed to support eScience was one of the first attempts to clarify differences between preservation, archiving, and curation in relation to digital scholarship.[23] The report identifies data curation as the "activity of, managing and promoting the use of data from its point of creation, to ensure it is fit for contemporary purpose, and available for discovery and re-use." It specifies that "archiving" is needed for "curation," and "preservation" is "an aspect of archiving," with all three "involved in managing change over time."[24] To this day, the distinctions and applications of these three concepts remain imprecise, as is also the case with the terms *digital curation*, *data curation*, and *data stewardship*. In the next section we clarify some of the differences between these terms, their use, and the overall importance in understanding the unique contribution of data curation to science and scholarship.

Co-Evolution of Digital Curation and Data Curation

In the UK, the term digital curation first appeared in a 2001 digital preservation workshop,[25] but later Neal Beagrie more explicitly defined it in his description of the newly formed Digital Curation Centre (DCC):

The term "digital curation" is increasingly being used for the actions needed to maintain digital research data and other digital materials over their entire lifecycle and over time for current and future generations of users. Implicit in this definition are the processes of digital archiving and digital preservation, but it also includes all the processes needed for good data creation and management, and the capacity to add value to data to generate new sources of information and knowledge.[26]

This text is nearly an exact match with the roles outlined in the 2003 Jisc report on *data* curation—evidence that early on the two concepts were thought of synonymously. Other publications advancing the concept of curation in the UK applied the terms loosely for a number of years. For example, a publication formalizing many of the original 2003 Jisc report recommendations states: "data curation concerns the long-term management of data, from its initial collection to its deposition into an archive."[27] The following year, digital curation was more "broadly interpreted" as being "about maintaining and adding value to, a trusted body of digital information for current and future use."[28] Summarizing many different uses and definitions of digital curation, Yakel's definition of *digital curation* is explicit about its place within the information professions. She refers to it as "the active involvement of information professionals in the management, including the preservation, of digital data for future use."[29]

Yakel's publication came out the same year that we offered a definition of data curation for the Data Curation Education Program (DCEP), funded by the Institute of Museum and Library Services (IMLS) at the Graduate School of Library and Information Science (GSLIS), University of Illinois at Urbana-Champaign. The DCEP definition states that data curation is "the active and ongoing management of data throughout its entire lifecycle of interest and usefulness to scholarship."[30] This definition guided development of the Specialization in Data Curation in the GSLIS master's program and has been applied consistently in DCEP initiatives across the sciences and the humanities, including an annual Summer Institute on Data Curation. The distinction between this definition and those of digital curation is subtle, but important. The focus of data curation is on the "interest and usefulness" of data to scholarship, and this emphasizes the traditional mission of research libraries to provide information to support research and the production of new knowledge.

Digital curation is, or has become, a term that better accommodates a broader range of digital material. It does not indicate what is being curated nor does it necessarily imply which communities can be purposefully served by curatorial activities. Data curation, on the other hand, relates directly to data that is produced and used by scholarly communities, and it facilitates the reuse and repurposing of data to meet new research needs. As Renear and Muñoz aptly explained in the introduction to a DCEP workshop held at the annual Digital Humanities conference, "Data curation addresses the challenge of maintaining digital information that is produced in the course of research in a manner that preserves its meaning and usefulness as a potential input for further research."[31] Their emphasis, echoing Cragin et al., is on the scholarship and research enabled by the effective curation of digital data resources.[32]

Although far from complete, this account of the concept of curation provides important contextual background for the next section, where we discuss DCEP's pedagogical approach to data curation education. As noted above, the principle of

"purposeful curation" has been the guiding concept for DCEP and the Specialization in Data Curation. This area focuses exclusively on the curation of research data for scholarly and scientific communities, covering the complex theoretical and practical problems that span eScience and digital humanities.

Purposeful Work with Data: Beyond Stewardship

The "purposeful curation" concept guiding local educational efforts was more formally articulated in response to a paper that asserted, "library science has not demonstrated that it has the theoretical foundations and knowledge base that are capable of providing the framework for handling digital entities."[33] We argued the contrary: Because of its commitment to the "needs of users to access and use information of value over the long term," library and information science (LIS) is uniquely positioned to provide the kind of foundation needed for systems and services to manage digital content, especially research data.[34] Further, we explained that our aim in educating data curators was not only to prepare them to build and maintain data collections but also to be responsible for the "associated indexing systems, metadata standards, ontologies, and retrieval systems" that will make it possible for research data to work in concert with existing digital libraries, archives, and repositories. Data curation efforts that embrace this orientation to user communities and their needs within and across disciplines will be equipped to enable innovative research in the sciences and humanities with curated data for "experiments in scientific laboratories, the interpretation of texts by scholars in special collections, the development of exhibits in museums, and other purposeful work with data over time."[35]

This conception of curation goes beyond the roles Zorich originally envisioned,[36] attending to data quality, authentication, security, and associated documentation and tools. It is markedly more active and extends beyond assuring reliable data access, to adding value to support and advance research capabilities. The alignment with the research process is the primary feature that distinguishes data curation from data stewardship. Data stewardship is about management of a shared resource, often embodied by one person or a designated group.[37] It is clearly essential to the preservation and persistence of robust data collections, but it is a function of "managing data" that implies a less active, fixed maintenance of data over time.[38] Curation, on the other hand, is concerned with availability *and* future use of data, including the enhancement, extension, and improvement of data products for reuse beyond a single scholarly community.

At the same time, it is essential to recognize that the locally developed processes of data management and stewardship within specific communities are valuable sources of practical knowledge and current practice. The specialization needs to be applicable across domains, and the rate of change in research with digital data is a continual challenge for integrating best practices into the curriculum. Case studies on data management, processing, enhancement, and sharing in scientific

and humanities domains have proven to be some of the best material for teaching applied aspects of data curation. Likewise, for repository-based data curation expertise, we depend on our partnerships with national data centers and data services operations in research libraries for coverage of state-of-the-art practices. Increasingly these partners are also serving as hosts for internships and other field experiences for students to work alongside more experienced curators.

Foundations of Curation

The core of the curriculum is a single foundational course grounded in concepts and functions that underpin the curation of research data in practice. The course draws on principles from both LIS and archival science, as well as current trends and practice within research domains and professional work in data repositories. Of the six areas within the core, three are fairly common to information management systems, especially when dealing with digital content: interoperable technical infrastructure, policy development, and intellectual property. The remaining three areas are more closely aligned with key dimensions of LIS expertise and are central to the "purposeful curation" concept: user communities and their information behavior; collection development, management, and research services; and information organization and representation.

In the data curation curriculum these dimensions are further refined in relation to problems and practices around research data. The user dimension emphasizes *research cultures* that impact practices of data production and sharing within the system of scholarly communication; the *collections* dimension emphasizes heterogeneous, complex data resources and their potential for reuse across the lifecycle of data, informed by key aspects of archival science; and the *representation* dimension emphasizes the formal characteristics of information objects that carry digital data—with a particular focus on issues of identity, ontology development, and provenance. It is important to note that these dimensions cut across some of the basic areas of knowledge and skills for data curation, such as metadata. That is, while metadata activities are generally associated with representation, effective application of metadata will adequately document aspects of research cultures related to data-collection methods. High-functioning metadata will also be guided by collection issues related to granularity of description, properties of aggregate data products, and meaningful groupings of data for discovery and integration.

Research cultures: data production, data sharing, and scholarly communication

The curation of research data needs to be informed by a detailed understanding of how data are produced and used in conducting research and how these processes fit into the larger context of scholarly communication. As is true with information behavior more generally, data practices vary in important ways across disciplines and sub-disciplines. For example, while there are a number of successful, large-scale open data initiatives, with the Protein Data Bank and the Sloan Digital Sky Survey as canonical cases, they are the exception rather than

the rule.

While data curation is largely concerned with how data sharing within and across research communities can advance science and scholarship, most researchers do not make their data publically available and most research communities have yet to develop a culture of data sharing.[39] If researchers are willing to share data, the form in which they are willing to release them may provide an easy preservation target, but it may not have the most value for long-term reuse.[40] Moreover, actual reuse can be highly dependent on rich contextual metadata provided by data producers.[41]

Data curation is part of the more global context of scholarly communication in which research libraries, publishers, and scholars have always been the primary actors and stakeholders. As would be expected in this tradition, much attention has been focused on the data associated with specific journals or published papers. To

date, there have been a few successful approaches in the "journal—underlying data" archiving model.[42] However, some of the most challenging curation problems and, arguably, the greatest opportunities lie in our ability to make more basic research data available and interoperable across fields.[43] Therefore, data curators should be trained and organizationally positioned to assess accurately which data collections are of value, and in which combinations and states other scholars would find them most useful. Data curators should be able to identify data that have high priority for both their immediate service community and broader domains of interest. This is a unique set of roles and responsibilities for the field of LIS, and it requires educated professionals who understand relationships among research areas and the potential affordances of data for new applications.

Collections: heterogeneous data, lifecycles, and archives

Collections have always been central to librarianship and archival science, and their importance will continue as the stores of data grow and become increasingly searchable and browsable. They will be essential units for planning and implementing curation objectives and resources, for supporting research communities, and supporting the reuse and repurposing of data. Variations on the traditional processes of selection, appraisal, and collection management are particularly relevant.[44] At various scales, from big data-driven research to smaller data-intensive research, collections have significant properties that provide intellectual, social, and organizational structures for meaningful work with data.[45]

Collections have life cycles of usefulness to scholarship that need to be regulated through curation, including maintenance, migration, and preservation, over time.

Moreover, "collection development and collection description are formative curation activities that add value for scholarly inquiry at both the collection and [larger] aggregation levels."[46] As the data accessible in digital environments continue to increase at a rapid pace, sound collection development and description will be essential to presenting data for exploration and discovery. In conjunction with searching capabilities, researchers will greatly benefit from the ability to browse dense and cohesive layers of data sources that not only anticipate researchers' known needs, but also allow them to effectively navigate through and interpret extensive bodies of openly accessible data.

Collections will increasingly need to be organized within local institutional repositories and disciplinary data repositories as researchers submit data to comply with funding-agency requirements for data-management planning and data sharing. Curators will be involved as advisors and collaborators in data-management planning, for which knowledge of collections will be key. They will need to identify which versions to retain and share; align with institutional aims and capacity for preservation and access options; assess the fitness of data for deposit in domain or consortia repositories; and advise on intellectual property issues that researchers will face in attempting to comply with these mandates.

More generally, collections will be vital to providing reliable access and preservation to data products. Over the long term, we risk ending up with data dumps rather than functional libraries of relevant and usable data assets if we ignore the art of building and organizing data collections for research communities.

Representation: identity, ontologies, and provenance

Rigorous conceptual analysis of core curatorial notions such as representation, identity, authenticity, reference, and provenance is a distinctive aspect of the DCEP approach to data curation education. Although stewardship requires verifying bit sequences and ensuring that files have not been corrupted, many curation problems turn on much more challenging issues, such as whether two files contain the same data even though they use different representation languages, encodings, or formats. A comparison of bit sequences will provide no assistance here, and so is inadequate for curation that goes beyond stewardship. What is needed is a terminological framework that clarifies these issues, as well as techniques for documentation and confirmation. Getting full value from cyberinfrastructure and computational systems requires that curators be able to refine formal terminology and document the "interpretive frames" that connect reasoning across levels of abstraction.[47] Similarly, precisely documenting the derivation of one dataset from another, critical for both scientific and humanities scholarship, also requires sophisticated conceptual frameworks, tools, and practices. To support curatorial education we are therefore developing general frameworks of concepts and terminology that can be used across disciplines.[48]

A solid familiarity with semantic technologies (such as the Resource Description

Framework [RDF] and the Web Ontology Language [OWL]) is increasingly important as structured terminologies and ontologies for representing scientific and cultural information are now widely used. These data organization strategies are fundamental to data curation: they help ensure that data remains meaningful; they improve software interoperability and data integration; and they support unanticipated use, as well as the mobilization of interdisciplinary approaches. Finally, even the distinctions between publication and data, and reading and analysis, are blurring; new tools are both responding to and driving changes in scholarly publishing and how researchers search, filter, scan, link, annotate, and analyze fragments of content from the literature.[49] For data curators there is no escape from engaging these fundamental issues of representation, identity, ontologies, and provenance.

Curation Professionals in the Workforce

A recent analysis of job placements for students completing the Specialization of Data Curation in the GSLIS master's program at Illinois has been instructive for understanding the types of organizations and positions that need data professionals. The discourse in the field has emphasized that data curation expertise will need to permeate the research process at large, with curation beginning at the initial planning stages of a research project, followed by long-term data management and providing open access data for public use.[50] Interestingly, our placement data show that curation professionals are filling positions in many types of organizations in roles that will, in fact, have an impact throughout the life cycle of data in the research enterprise.

Of the fifty-five graduates to date, we have been able to track placement of forty-nine individuals. Thirty-three percent have taken positions in research libraries and museums. A good number of these graduates have gone to academic libraries, but some are working in national libraries, special libraries, large art museums, and other cultural heritage institutions. Twenty percent have gone to research centers, including national data centers and digital humanities and scientific research institutes, where they are more directly involved with the research applications and curation of data. Another twenty percent are in the corporate sector, which needs high- functioning data to gain competitive market advantage, but in these settings curation skills are also being

blended with statistical competencies that have more traditionally been labeled as data science. The nearly fifty position titles also show how curation positions are being formalized; they range from data curator, data- management consultant, research data librarian, and digital preservation librarian to data analyst, digital asset manager, and information architect.

Through our experience with the DCEP program, the benefits of internships and practicum in data centers became evident, and field experiences are now a priority in our next phases of program development. The Data Curation Education in Research Centers (DCERC) initiative, funded by IMLS, has been a

major step in building a model for integrating field experiences into MA and PhD programs through a partnership with the National Center for Atmospheric Research (NCAR), a long-standing international leader in scientific data infrastructure and services. The DCERC collaboration includes the University of Tennessee, School of Information Sciences, whose participating masters students complete the Foundations of Data Curation course at Illinois.

The NCAR internships, piloted with Illinois doctoral students and Tennessee masters students, have been explicitly designed to mentor students in both data management and the scientific research contexts. It is our belief that apprenticeships need to involve students directly in best practices at mature data centers and expose them to the actual data problems that active researchers experience; this will prepare them adequately to excel at curation in data-intensive research environments.

Conclusion

This paper provided an overview of our approach to data curation education. It is important to note that this approach has been greatly influenced, both formally and informally, by our many collaborators and colleagues working to improve data access and research capabilities for scholarship in both the sciences and humanities. Our "purposeful curation" perspective is not meant to be comparative to other curation programs or an exhaustive, fully articulated pedagogy. However, we do feel that our student placement data, especially the high number of graduates placed across different types of research organizations, is solid evidence for the success of this particular approach.

The historical overview presented at the beginning of the paper gave an account of the concepts and definitions related to data curation based on the literature encountered through our work on the DCEP and DCERC initiatives, as both seasoned and new LIS educators and researchers. So, while we have identified a number of landmarks and trends in the conception of data curation to date, much of the story is still to be traced, interpreted, and retold. We hope that this initial effort will encourage additions, corrections, and amendments to the history of the field, and in turn lead to better sharing of both seasoned and new pedagogical approaches to what may prove to be an important turning point for the information professions.

FOOTNOTES

1. CarolePalmer,AllenRenear,andMelissaCragin,"PurposefulCuration:ResearchandEducatio nforaFuturewithWorkingData,"Proceedings of the 4th International Digital Curation Conference, 2008. [↵]

2. WilliamG.Fry,"MethodsinTaxonomy,"Nature207(1965):246,doi:10.1038/207245a0.[↵]

3. Ibid.[↵]

4. DianeM.Zorich,"Datamanagement:Managingelectronicinformation:Datacurationinmuseu ms,"MuseumManagementandCuratorship14, no. 4 (1995): 431. [↵]

5. DavidBearman,"WhatAre/IsInformatics?"ArchivalInformaticsNewsletter(1987):8.[↵]

6. SusanHockey,"DevelopingTextStandards,"SelectedpapersfromthesecondInternationalCo nferenceonHypermediaandInteractivityin Museums, Cambridge, England (1993): 255, http://www.archimuse.com/publishing/ichim93/hockey.pdf (http://www.archimuse.com/publishing/ichim93/hockey.pdf). [↵]

7. Jisc,http://www.jisc.ac.uk/aboutus.aspx.[↵(#identifier_6_4819)]

8. "InterviewwithSheilaAndersonandHamishJames,AHDS,"DigitalCurationCenter,accessed May28,2013, http://www.dcc.ac.uk/community/interviews/sheila-anderson-and-hamish-james. [↵]

9. DanielGreenstein,"ServingtheArtsandHumanities"Ariadne4(1996),http://www.ariadne.ac .uk/issue4/ahds.[↵(#identifier_8_4819)]

10. J. Unsworth, "A Master's Degree in Digital Humanities: Part of the Media Studies Program at the University of Virginia," Congress of the Social Sciences and Humanities, Université Laval, Québec, Canada, 2001, http://people.lis.illinois.edu/~unsworth/laval.html (http://people.lis.illinois.edu/~unsworth/laval.html); M. G. Kirschenbaum, "What Is Digital humanities and What's It Doing in English Departments?" ADE Bulletin 150 (2010): 1-7. [↵]

11. Susan Hockey, "The history of humanities computing," A companion to digital humanities (2004): 3-19. [↵]

12. Roberto Busa, "The annals of humanities computing: The index thomisticus," Computers and the Humanities 14, no. 2 (1980): 83. [↵]

13. Cited in Hockey, "The History of Humanities Computing." [↵]

14. Harry M. Logan, "Electronic lexicography," Computers and the Humanities 25, no. 6 (1991): 351-61 [↵]

15.   Hockey, Susan. 1992. Humanist 5.0632 CETH Inaugural Seminar 1/17. Retrieved from http://dhhumanist.org/Archives/Virginia/v05/0621.html (http://dhhumanist.org/Archives/Virginia/v05/0621.html). [↵]

16.   For an interesting post hoc alignment, see Allen H. Renear, Molly Dolan, Kevin Trainor, and Melissa H. Cragin, "Towards a cross-disciplinary notion of data level in data curation," Proceedings of the American Society for Information Science and Technology 46, no. 1 (2009): 1-8. [↵]

17.   William Thomas, "Computing and the historical imagination," A companion to digital humanities (2004): 56-68. [↵]

18.   Tobias Blanke, Mark Hedges, and Stuart Dunn, "Arts and humanities e-science—Current practices and future challenges," Future Generation Computer Systems 25, no. 4 (2009): 474-80. [↵]

19.   Tony Hey and Anne Trefethen, "e-Science and its implications," Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences 361, no. 1809 (2003): 1809-1825. [↵]

20.   Jim Gray, Alexander S. Szalay, Ani R. Thakar, Christopher Stoughton, and Jan Vandenberg, "Online scientific data curation, publication, and archiving," Proceedings of SPIE (2002): 103-107; Jim Gray, David T. Liu, Maria Nieto-Santisteban, Alex Szalay, David J. DeWitt, and Gerd Heber, "Scientific data management in the coming decade," ACM SIGMOD Record 34, no. 4 (2005) 34-41; Tony Hey, Stewart Tansley, and Kristin Michele Tolle, eds. The fourth paradigm: data-intensive scientific discovery (Redmond, WA: Microsoft Research, 2009). [↵]

21.   Gray et al., "Online scientific data curation, publication, and archiving." [↵]

22.   Paulien Hogeweg, "The roots of bioinformatics in theoretical biology," PLoS computational biology 7, no. 3 (2011): e1002021. [↵]

23.   Philip Lord and Alison Macdonald, e-Science Curation Report: Data curation for e-Science in the UK: an audit to establish requirements for future curation and provision (Digital Archiving Consultancy Limited, 2003). [↵]

24.   Ibid., 12. [↵]

25.   Neil Beagrie, "Digital curation for science, digital libraries, and individuals," International Journal of Digital Curation 1 (2006), 3-16. [↵]

26.   Neil Beagrie, "The Digital Curation Centre" Learned Publishing 17, no. 1 (2004): 7-9. [↵]

27.   Philip Lord, Alison Macdonald, Liz Lyon, and David Giaretta, "From data deluge to data curation," Proc 3th UK e-Science All Hands Meeting (2004): 371-75. [↵]

28.   David Giaretta, "DCC approach to digital curation, version 1.23," (white paper,2005), http://dev.dcc.rl.ac.uk/twiki/bin/view/Main/DCCApproachToCuration. [↵ (#identifier_27_4819)]

29. Elizabeth Yakel, "Digital curation," OCLC Systems & Services 23, no. 4 (2007): 335-340. [↵]

30. Melissa H. Cragin, P. Bryan Heidorn, Carole L. Palmer, and Linda C. Smith, "An educational program on data curation," American Library Association Conference, Science and Technology Section, Washington, D.C., June 25, 2007. http://hdl.handle.net/2142/3493 (http://hdl.handle.net/2142/3493) [↵]

31. Trevor Muñoz and Allen Renear, "Issues in Digital Humanities Data Curation," (white paper, 2011), http://hdl.handle.net/2142/30852). [↵ (#identifier_30_4819)]

32. Cragin et al., "An educational program on data curation." [↵]

33. Seamus Ross, "Digital preservation, archival science and methodological foundations for digital libraries," New Review of Information Networking 17, no. 1 (2012): 43-68. [↵]

34. Palmer, Renear, and Cragin, "Purposeful Curation." [↵]

35. Ibid. [↵]

36. Zorich, "Data management." [↵]

37. Karen S. Baker and Lynn Yarmey, "Data stewardship: Environmental data curation and a web-of-repositories," International Journal of Digital Curation 4, no. 2 (2009): 12-27. [↵]

38. Cliff Jacobs and Steven J. Worley, "Data Curation in Climate and Weather: Transforming Our Ability to Improve Predictions through Global Knowledge Sharing," The International Journal of Digital Curation (2009). [↵]

39. Christine L. Borgman, "The conundrum of sharing research data." Journal of the American Society for Information Science and Technology 63, no. 6 (2012): 1059-1078; Anne E. Thessen and David J. Patterson, "Data issues in the life sciences," ZooKeys 150 (2011): 15; Carol Tenopir, Suzie Allard, Kimberly L. Douglass, Arsev Umur Aydinoglu, Lei Wu, Eleanor Read, Maribeth Manoff, and Mike Frame, "Data sharing by scientists: practices and perceptions," PLoS One 6, no. 6 (2011). [↵]

40. Melissa H. Cragin, Carole L. Palmer, Jacob R. Carlson, and Michael Witt, "Data sharing, small science and institutionalrepositories," Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 368, no. 1926 (2010): 4023-4038. [↵]

41. Matthew S. Mayernik, Jillian C. Wallis, and Christine L. Borgman, "Unearthing the infrastructure: Humans and sensors in field-based scientific research," Journal of Computer Supported Cooperative Work (2012); Nicholas M. Weber, Karen S. Baker, Andrea K. Thomer, Tiffany C. Chao, and Carole L. Palmer, "Value and context in data use: Domain analysis revisited," Proceedings of the American Society for Information Science and Technology 49, no. 1 (2012): 1-10; Laura A. Wynholds, Christine L. Borgman, Jillian C. Wallis, Ashley Sands, and Sharon Traweek, "Data, Data Use, and Scientific Inquiry: Two Case Studies of Data Practices," ACM Press (2012): 19. [↵]

42. For example, Michael C. Whitlock, Mark A. McPeek, Mark D. Rausher, Loren

Rieseberg, and Allen J. Moore, "Data archiving," The American Naturalist 175, no. 2 (2010): 145-46. [↵]

43. Hanisch, Robert, and Sayeed Choudhury, "The Data Conservancy: Building a Sustainable System for Interdisciplinary Scientific Data Curation and Preservation," Proceedings of the PV 2009 conference, European Space Agency (2009). [↵]

44. Angus Whyte and Andrew Wilson, How to Appraise & Select Research Data for Curation, Digital Curation Centre, 2010, http://www.dcc.ac.uk/resources/how-guides/appraise- select-research-data. [↵ (#identifier_43_4819)]

45. Carole L. Palmer, Oksana L. Zavalina, and Katrina Fenlon, "Beyond size and search: Building contextual mass in digital aggregations for scholarly use," Proceedings of the American Society for Information Science and Technology 47, no. 1 (2010): 1-10. [↵]

46. Katrina Fenlon, Jacob Jett, and Carole L. Palmer, "Digital Collections and Aggregations," DHCuration Guide (2011), http://guide.dhcuration.org/collections/; Carole Palmer, "Scholarly work and the shaping of digital access," Journal of the American Society for Information Science and Technology 56, no. 11 (2005). [↵ (#identifier_45_4819)]

47. David Dubin, Karen Wickett, and Simone Sacchi, "Content, Format, and Interpretation," Proceedings of Balisage: The Markup Conference 2011. Balisage Series on Markup Technologies 7 (2011), doi:10.4242/BalisageVol7.Dubin01. [↵]

48. Renear et al., "Towards a cross-disciplinary notion of data level in data curation"; Simone Sacchi, Karen Wickett, Allen Renear, and David Dubin, "A framework for applying the concept of significant properties to datasets," Proceedings of the American Society for Information Science and Technology 48, no. 1 (2011): 1-10; Karen M. Wickett, Simone Sacchi, David Dubin, and Allen H. Renear, "Identifying Content and Levels of Representation in Scientific Data," Proceedings of the American Society for Information Science and Technology 49, no. 1 (2012): 1-10. [↵]

49. Allen H. Renear and Carole L. Palmer, "Strategic reading, ontologies, and the future of scientific publishing," Science 325, no. 5942 (2009): 828-

50. Nicholas M. Weber, Carole L. Palmer, and Tiffany C. Chao, "Current Trends and Future Directions in Data Curation Research and Education," Journal of Web Librarianship 6, no. 4 (2012): 305-20. [↵]

LICENSE