

# Issues in Humanities Data Curation

**Trevor Munoz and Allen Renear**

Center for Informatics Research in Science and Scholarship (CIRSS)

University of Illinois, Urbana-Champaign

## **Abstract**

As both the materials and the analytical practices of humanities research become increasingly digital, the challenge of sustaining meaningful access to the outputs of humanities research becomes more urgent. Continuing technological change and new institutional pressures require a sustainable commitment to curate humanities data throughout its entire lifecycle from creation to re-use and long-term preservation. Success will depend on coordinated, collaborative efforts and arrangements amongst scholars, librarians, administrators, funders, and their organizations.

This document is intended to provide background and provoke discussions about the skills, professional roles, training, and institutional support needed for curation of humanities research materials. Each section contains questions for further exploration and debate that we hope will provide participants with opportunities to share their own experiences and knowledge.

The final white paper resulting from the Humanities Data Curation Summit sponsored by the Data Curation Education Program at the University of Illinois, Urbana-Champaign, and CenterNet is intended to advance the development of a curation agenda for the digital humanities as a vital piece of conceptual infrastructure [\[1\]](#) [\[#svensson2011\]](#) for the field. Based on a shared understanding of current problems, the final white paper will propose concrete, specific recommendations for scholars, librarians and archivists, professional societies, institutions, and funders.

## **Introduction**

Data curation was originally conceptualized as an e-Science problem precipitated by large amounts of data in digital formats [2] [lord2004]. Curation is an emerging challenge for the humanities as well. 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. In this way data curation is distinct from digital curation or digital preservation (though preservation represents an important part of curation).

Curation encompasses gathering material, making it discoverable by describing and organizing it, placing it in a context of related information, supporting its use for diverse intellectual purposes, and ensuring its long-term survival [3] [reside2011]. Curatorial practices form part of many humanities research practices, and the digital humanities community, in particular, already possesses sophisticated experience with preserving access to digital scholarship. Given the scope of activities required, curation may be performed by tenure-track scholars, alternative academics, librarians, software developers, students both undergraduate and graduate, as well as interested members of the public.

The growing amount of data (both digitized and born-digital) is only one motivation for addressing humanities curation needs at this time. The unique features of humanities data (constituted as it often is from social practices, multiple, perhaps conflicting layers of interpretation, and affective responses) may not yet be adequately accommodated by data curation practices adopted from the natural sciences [4] [renew2009]. This can have important implications for the design of curation systems and curation education programs [5] [renew2010]. For example, the challenge of exchanging and "re-using" richly-interpretive datasets (complex markup, visualizations, games) remains an area where further research is needed to ensure effective curation of significant humanities materials [6] [mcdonough2010] [7] [flanders2009]. Most importantly, digital humanities scholarship represents a unique and compelling opportunity for the humanities to demonstrate their relevance and value to the broader society. Sustaining such scholarship---demonstrating good stewardship of the investment its creation represents, and more fully integrating digital materials and methodologies into teaching and research---is an intellectual as well as a practical task, one that is critical during times of financial and organizational pressure and change.

To promote more effective curation of humanities data---that is, better discovery,

retrieval, exchange, re-use, and preservation of humanities data---several areas need to be addressed including skills, professional roles, education, communication, training, and institutional support.

## Skill Sets

We hope to refine our understanding of a "curation skill set," drawn from a number of disciplines and professions---the original humanities subject areas, librarianship, information technology, computer science, archives, and information science [8] [#palmer2008]. This skill set will certainly not be found uniformly represented in a single role, (see Roles [#roles] below), but nevertheless we believe it represents a coherent body of knowledge and practice. Understanding the core skills for humanities data curation may be especially important for structuring partnerships in an environment where tasks and institutional responsibilities are distributed

One of the most important tasks facing humanities data curators is to ensure that the digital representations of objects of study in the humanities function effectively *as data*---that they are processable by machines and processable across systems and collections while still retaining provenance and complex layers of meaning. Thus, curation overlaps and extends the practice of digital humanities research as such research is represented in projects, conference proceedings, and journal articles [9] [#bradley2009] [10] [#flanders2009b] [11] [#mccarty2008]. Further growth of the digital humanities may see differentiation of activities that are curatorial from those that are part of disciplinary research and this change in skill sets may be reflected in changes to staff roles.

Preliminary research conducted by the Data Curation Education Program for the Humanities (DCEP-H) project on data curation needs in digital humanities centers did not clearly identify agreement on the skill sets necessary for data curation in the humanities but our early results are suggestive. The study interviews and survey suggest that the nascent core of a humanities data curation skill set might comprise knowledge of interoperability and standards, metadata, markup, database design, and project management [12] [#munoz2011]. When combined with the results from a Data Curation Education Program (DCEP) study led by Melissa Cragin of job postings for curation-related positions, the emphasis on skills in markup and database design becomes more pronounced. Some of these skills (metadata, markup, databases) seem directly curatorial. Others seem to reflect more the particular

organizational arrangements governing digital humanities research at the current time---heavily project-based while seeking opportunities to standardize datasets and methods.

Given the need for curatorial work throughout the data lifecycle from planning to creation to publication, reuse and preservation, scholars will need sufficient skills to perform at least some curation of their own data [13] [#kirschenbaum2011]. Librarians and archivists will need technical and research skills to be effective partners in curation.

- Are knowledge of interoperability and standards, metadata, markup, database design, and project management in fact the most important skills for humanities data curators at present?
- What skill sets will be needed in the future?
- What is expected to drive changes in skill sets?
- The DCEP-H study focused on managers of well-established digital humanities centers, how do skill sets vary with type of organizational unit or role, or discipline/field, method or data?
- Can we identify skill sets that cut across disciplinary lines or that apply to different kinds of data?
- Curatorial activities are already part of many humanities research practices (e.g., scholarly editing, creation of thematic/personal research collections). How can traditional research practices, principles, norms, and theories inform humanities data curation?

## **Roles**

The data curation research community has often advocated for the placement of curators "upstream" in labs and research centers, that is, in roles working with scholars during research development and data creation phases to support curatorial lifecycle planning and execution [14] [#swan2008]. However, when asked by DCEP-H researchers about plans for new staff and new roles, managers of digital humanities centers expressed skepticism. While this resistance on the part of managers must be taken seriously since it likely represents relevant experience and judgement, the proper balance must be struck between distributing basic data curation among multiple staff or project researchers and employing trained professionals in dedicated roles. There is a limit beyond which distributing curation

responsibilities among existing staff or outsourcing functions of data management may not be effective. Research on the role of information work in other disciplines dealing with increasing amounts of digital data suggests that data curation as an "add-on" to disciplinary doctoral or post-doctoral training may have limited success or may result in siloed data. Neither is curation a task which can be given to information technology staff alone. Interventions by data curators may be most effective at project stages when information work is most routine or when it is highly speculative, as is often the case with new interdisciplinary research questions or in emerging collaborations [\[15\]](#) [\[#palmer2006\]](#) [\[16\]](#) [\[#palmer2007\]](#).

Libraries and data centers may have staff with both the necessary content knowledge and technical skills to be effective data curators but they must ensure that roles are defined in ways that allow and encourage staff to engage in research design and consultation [\[17\]](#) [\[#walters2011\]](#). The role of curator is not only a service role. Librarians and archivists acting as humanities data curators must embrace roles as researchers in the manner that programmers and software engineers in the digital humanities have come to do. Humanities data curators will have their own research agenda related to description, discovery, retrieval, contextualization and preservation.

The establishment of new roles will need to be accompanied by a continuing re-examination of the distribution of labor, recognition, and reward in academic work [\[18\]](#) [\[#clement2011\]](#). Prevailing practices, divisions between tenure-track scholars, alternative-academics, technologists, and other professionals, are too unequal to support the effective team-based work conducive to curation.

- What would it mean to understand curators' contributions in a similar way to that in which the contributions of technologists and programmers are recognized and valued in the digital humanities?
- What activities are best located in what sorts of staff roles and relationships?
- Do we need new named roles? If so what are they and where should they be located (projects, centers, libraries?). Or do we need new responsibilities for roles that already exist in organizational units? Or both?
- How do we bring about these new responsibilities within existing institutional expectations?

## **Education, training, communication, and sharing**

A summer institute in humanities data curation run by the DCEP-H project received surprisingly strong interest from senior-level staff. While the program was initially designed for lower-level staff who would have direct day-to-day responsibility for curation work, division managers and director-level staff were heavily represented in the applicant pool [19] [#renew2010b]. This encouraging but surprising result encapsulates an important point about the challenge of providing education and training for a humanities data curation workforce. The high proportion of senior staff looking to gain familiarity with curation issues raises the question of whether there are currently enough administrators equipped to effectively recruit staff, and supervise and coordinate curation programs focused on stewardship of humanities data. A focus on training new graduates and lower-level staff may not yield the most effective curation programs if education and training is not also directed toward creating a group of higher-level managers and advocates conversant with data curation issues.

- What kinds of changes in staffing and training are needed?
- How can we develop professionals with these skills?
- Library education programs have been criticized for failing to prepare graduates for technically-oriented roles. How can library schools and perhaps particularly iSchools better prepare students for careers curating humanities data? Basic data curation should also be incorporated into disciplinary post-graduate programs in the humanities---what form might this training take?

## **Institutional Support**

The identity and culture of organizational units tasked with leading curation efforts will affect the way that curation is envisioned, scoped, and carried out. Research centers, libraries, campus IT, and disciplinary and institutional repositories all have interests in providing curation services to scholars. There are significant cultural differences between these units.

The current landscape features a number of stable and well-funded digital humanities centers with new centers being founded. At the present time, then, there might be an argument for finding institutional homes for curation functions in centers. However, this set of institutional arrangements is a recent development. In the past, often as digital humanities centers grew, tensions developed with libraries

and campus IT. Many digital humanities units were moved around, subsumed by other divisions, responsibilities, and priorities. Many curation activities are not a natural match for centers' research missions. Funding for centers, which often includes substantial "soft money," is also problematic from a curation perspective. Data is particularly vulnerable to loss when funding is disrupted [20] [#blueribbon2010]. The career of institutional repositories might likewise make them uncertain homes for curated data [21] [#salo2008]. The university library is viewed by many faculty members as a purchasing agent and a warehouse for printed material rather than a dynamic home for new research services [22] [#schonfeld2009]. The development of systems to allow the library to take stewardship of digital research materials is still, in many places, a work in progress.

While all of the potential homes for curation services have drawbacks, the library is most likely to have the longevity and impartiality to take the lead on humanities data curation and should embrace this role. This may require diverting resources from development and management of print collections or other budgetary changes. Connecting students and faculty and their research materials with analysis and presentation tools, datasets, and identity and other curation services should be part of the mission of the academic research library even if this means considering additional fees or increases in indirect cost structures.

At the same time, the priorities of funders and the availability of new services have encouraged more cross-institutional partnerships. While there are still few models of disciplinary repositories in the humanities, the funding environment and the expensive and specialized requirements of long-term digital preservation, makes the prospect of any single institution being able to curate data effectively seem remote. Conversations between funders, professional societies, libraries, and archives should aim to converge on technically and financially sustainable requirements for digital scholarship [23] [#maron2011].

- How can the relative roles and positions of libraries and digital humanities centers be reimagined?
- Do we need new organizational units? Is change internal to organizational units (e.g., projects, centers, departments, libraries, central IT, universities) the most important priority? Or, is restructuring collaborations and relationships a more important priority?
- What role should/can professional societies and funding agencies play in

encouraging best practice? How have such top-down changes to practice worked in the past?

- Should there be requirements for data management plans in the humanities, similar to those taking hold in the sciences? Who are the stakeholders in developing such plans and who should develop them---funders? societies? institutions? centers? research groups?
- How can the humanities participate in larger efforts/institutions/initiatives /infrastructures related to the curation of research data, such as those underway in the scientific data curation community?

## Acknowledgements

This work was funded by a grant from the Institute of Museum and Library Services (RE-05-08-0062-08). We have benefited from the suggestions and comments of Melissa Cragin and Carole Palmer. Kay Walter and Julia Flanders read an early draft of this document and provided valuable suggestions for its improvement.

This preliminary document was created to provide background and promote discussion. Please do not cite directly. The final version of the white paper from the Humanities Data Curation Summit will be available after the meeting. If you have any questions, please **contact Trevor Munoz [mailto:trevor.munoz@gmail.com]** directly.

Last updated: 2011-06-15

## References

[1] Svensson, Patrik. 2011. "From Optical Fiber To Conceptual Cyberinfrastructure," *Digital Humanities Quarterly* 5(1). Accessed June 15, 2011. <http://www.digitalhumanities.org/dhq/vol/5/1/000090/000090.html>

[2] Lord, Philip, Alison McDonald, Liz Lyon and David Giaretta. "From Data Deluge to Data Curation," Paper presented at the eScience All Hands Meeting, Nottingham, UK, September 2004. <http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/150.pdf>

[3] Reside, Doug. "What is a Digital Curator?" Accessed June 15, 2011. <http://www.nypl.org/blog/2011/04/04/what-digital-curator>

[4] Renear, Allen H., Molly Dolan, Kevin Trainor and Melissa H. Cragin. "Towards a Cross-Disciplinary Notion of Data Level in Data Curation," Paper presented at the 72nd ASIS&T



Annual Meeting, Vancouver, B.C., November 8-11, 2009. <http://hdl.handle.net/2142/14547>

[5] Renear, Allen H., Trevor Munoz and Kevin Trainor. "Data Curation Education for the Humanities: Principles and Challenges," Poster presented at the Fifth Annual Chicago Colloquium on Digital Humanities and Computer Science, Evanston, IL, November 21-22, 2010. <http://hdl.handle.net/2142/17421>

[6] McDonough, Jerome P., Robert Olendorf, Matthew Kirschenbaum, Kari Kraus, Doug Reside, Rachel Donahue, Andrew Phelps, Christopher Egert, Henry Lowood and Susan Rojo. *Preserving Virtual Worlds Final Report* (2010) <http://hdl.handle.net/2142/17097>

[7] Flanders, Julia. "Dissent and Collaboration," Paper presented at Digital Humanities 2009, College Park, MD, June 22-25, 2009.

[8] Palmer, Carole L., Allen H. Renear and Melissa H. Cragin, "Purposeful Curation: Research and Education for a Future with Working Data," Paper presented at the Fourth International Digital Curation Conference, Edinburgh, Scotland, December 1-3, 2008. <http://hdl.handle.net/2142/9764>

[9] Bradley, John. 2009. "What the Developer Saw: an Outsider's View of Annotation, Interpretation and Scholarship," *Digital Studies* 1(1). Accessed June 15, 2011. [http://www.digitalstudies.org/ojs/index.php/digital\\_studies/article/view/143/202](http://www.digitalstudies.org/ojs/index.php/digital_studies/article/view/143/202)

[10] Flanders, Julia. 2009. "The Productive Unease of 21st-century Digital Scholarship," *Digital Humanities Quarterly* 3(3). Accessed June 15, 2011. <http://www.digitalhumanities.org/dhq/vol/3/3/000055/000055.html>

[11] McCarty, Willard. "Knowing ...: Modeling in Literary Studies," *A Companion to Digital Literary Studies*, ed. Susan Schreibman and Ray Siemens (Oxford:Blackwell), 2008. <http://www.digitalhumanities.org/companionDLS/>

[12] Munoz, Trevor, Virgil Varvel, Allen H. Renear, Kevin Trainor and Molly Dolan. "Tasks vs. Roles: A Center Perspective on Data Curation Needs in the Humanities," Paper presented at Digital Humanities 2011, Palo Alto, CA, June 19-22, 2011. <http://dh2011abstracts.stanford.edu/xtf/view?docId=tei/ab-223.xml;query=Munoz;brand=default>

[13] Kirschenbaum, Matthew, "Digital Humanities Archive Fever" (Institute Lecture, Digital Humanities Summer Institute, Victoria, B.C., June 6, 2011.

[14] Swan, Alma and Sheridan Brown, "Skills, Role & Career Structure of Data Scientists & Curators: Assessment of Current Practice & Future Needs," Report to the JISC, July 2008. <http://www.jisc.ac.uk/publications/reports/2008/dataskillscareersfinalreport.aspx>

[15] Palmer, Carole L. 2006. "Weak Information Work and 'Doable' Problems in Interdisciplinary Science," *Proceedings of the American Society for Information Science and Technology*, 43(1): 1-16. <http://eprints.rclis.org/handle/10760/8636>

[16] Palmer, Carole L., Melissa H. Cragin and Timothy P. Hogan. 2007. "Weak information work in scientific discovery," *Information Processing and Management*, 43: 808-820. <http://dx.doi.org/10.1016/j.ipm.2006.06.003>

[17] Walters, Tyler and Katherine Skinner. *New Roles for New Times: Digital Curation for Preservation*, Report Prepared for the Association of Research Libraries (2011). <http://www.arl.org/rtl/plan/nrnt/>

[18] "Off the Tracks: Laying New Lines for Digital Humanities Scholars," Report of a workshop held at the Maryland Institute for Technology in the Humanities, January 20-21, 2011. <http://mediacommons.futureofthebook.org/mcpress/offthetracks/>

[19] Renear, Allen H., Molly Dolan, Kevin Trainor and Trevor Muñoz, "Extending an LIS Data Curation Curriculum to the Humanities: Selected Activities and Observations," Poster presented at the iConference, Champaign, IL, February 3-6, 2010. <http://hdl.handle.net/2142/15061>

[20] *Sustainable Economics for a Digital Planet: Ensuring Long-Term Access to Digital Information*. Final Report of the Blue Ribbon Task Force on Sustainable Preservation and Access, February 2010. <http://brtf.sdsc.edu/>

[21] Salo, Dorothea. 2008. "Innkeeper at the Roach Motel." *Library Trends* 57(2). <http://minds.wisconsin.edu/handle/1793/22088>

[22] Schonfeld, Roger C., and Ross Housewright. *Faculty Survey 2009: Strategic Insights for Libraries, Publishers, and Societies*. Report by Ithaka S & R (April 7, 2010). <http://www.ithaka.org/ithaka-s-r/research/faculty-surveys-2000-2009/faculty-survey-2009>

[23] Maron, Nancy L., and Matthew Loy. *Funding for Sustainability: How Funders' Practices Influence the Future of Digital Resources*. Report by Ithaka S & R (June 2011).

<http://www.ithaka.org/ithaka-s-r/research/funding-for-sustainability>