5th International Digital Curation Conference
December 2009

Analyzing Data Curation Job Descriptions

Melissa H. Cragin, Carole L. Palmer, Virgil E. Varvel Jr., Aaron Collie, Molly A. Dolan

Center for Informatics Research in Science and Scholarship

Graduate School of Library and Information Science

University of Illinois at Urbana-Champaign, USA

Introduction

There is rising interest on the part of academic and research institutions to implement strategies to preserve and provide access to their digital assets. At the individual level, scientists and scholars are increasingly engaged in sharing the products of their work through administered repositories or directly via the Web. Meeting the technical and organizational requirements ensuing from of these trends will require a workforce skilled in data management, organization and representation, access and preservation, technology implementation, project management, and often domain expertise. In light of this, recent reports have identified a need for trained data librarians and data scientists (e.g. Swan & Brown, 2008), or called for new, organized programs to train LIS professionals and others to meet the anticipated demand for skilled professionals (e.g. Association of Research Libraries, 2006). However, defined career paths for library and information science professionals interested in data curation or data science are not yet firmly established (Interagency Working Group on Digital Data, 2009). As yet, studies of the emerging job market are rare (for an exception, see Lee, 2008), but these are needed for the design of educational programs intended to meet the needs of an emergent labor force, as well as to support the flow of newly-trained professionals into the market . This poster will present analysis of an initial set of data curation-oriented job postings, and characterize the DC employment landscape.

In developing the Data Curation Education Program (DCEP) at the Graduate School of Library and Information Science at the University of Illinois at Urbana-Champaign, we have been engaged in a continuous, multi-faceted needs assessment of the range of data service needs for scientists and scholars across domains to inform curriculum development. The DCEP is training LIS professionals for employment across a range of information-oriented institutions, including data centers, libraries and institutional repositories, museums, archives, and private industry. As the data curation field matures, it is essential to assess required skill sets in relation to current jobs, as well as emerging or anticipated roles. This study uses current job postings as a source for analyzing the types of jobs emerging, the kinds of institutions looking for professionals with data curation skills, education requirements for current positions, and expectations for different categories of skill sets.

Methods

Seventy-five (75) online job postings within the Sciences and Social Sciences

were harvested from 12 online sites from 8/27/08 - 5/01/09 and pooled with pertinent job postings received by DCEP staff. Job postings were retrieved using keyword queries using the following terms: data; digital; research; librarian; archivist; data scientist; data curator; digital curator; data librarian; digital librarian; data archivist; digital archivist. These terms were selected in order to return broadly relevant job descriptions (i.e. data; digital; research, etc) and narrowly relevant job descriptions (i.e. digital curator; data librarian, etc). Sites without search functionality were both manually browsed and searched via Google site searches. Online sites were chosen to retrieve job announcements across several organization types, including library and information science sites (JobLIST, ARL, LISjobs, iSchool), science and technology sites (National e-Science Center, Society of Systematic Biologists, SAS Google Group, Science-Jobs, Science Careers, SDLP), and a sample of general job hosting sites (e.g. Simply Hired). Inclusion of postings was selective, and relevance was determined by manual examination of the posted job description.

Findings and Discussion

Job postings clustered in a few states, primarily California, New York, Texas and Massachusetts. There is great variation in job duties, and the average years experience required is 4.4 years. More than half of the jobs required at least a master's degree, and 36 (48%) requiring a master's in library science. With respect to general job functions, non-library jobs were more likely to be administrative, and those posted by corporate or research organizations were more likely to require domain expertise, most often in science. It is notable that only 17% of the jobs require some computer programming experience (see Table 1). Several additional characteristics will be reported in the poster, including patterns across employer type (i.e. organization), distribution of required skills across job type, and required computer programming languages.

[Caption] Table 1 Number of job posts requiring computer programming experience organized by employer type.

Employer Type	Number of Jobs	Number requiring computer programming experience
Archives	2	0
Academic Departments	3	0
Corporations	6	3
Humanities Computing Center	1	1
Laboratories	4	1
Libraries	Academic = 29 Public = 1 Special = 4	Academic = 8 Public = 1 Special = 1
Museums	1	1
Research or Data Center	24	6
Total	75	17 (23%)

The analysis of jobs data presented in this poster covers the first stage of this DCEP needs assessment activity. As more LIS professionals prepare to enter the data curation field, it is incumbent on educators to understand and monitor employer needs. It is clear that this initial sample of job postings has some limitations; there is, for

example, bias toward library-based jobs because of the relatively narrow scope of job sites searched. While this constrains some parameters of analysis at this point (such as the range of employer type analysis), this sample is suggestive of trends to be assessed in a larger job post data set. Our current approaches to data collection and analysis are being refined in preparation for an extended, comparative analysis of a much larger pool of curation-oriented jobs being drawn from across the science, social science, and humanities domains.

Acknowledgements

References

[report] Association of Research Libraries (2006). *To Stand the Test of Time: Long-term Stewardship of Digital Data Sets in Science and Engineering.* A Report to the National Science Foundation from the ARL Workshop on New Collaborative Relationships: The Role of Academic Libraries in the Digital Data Universe. Washington, DC: ARL. . Retrieved August 6, 2009, from http://www.arl.org/info/events/digdatarpt.pdf.

[report] Interagency Working Group on Digital Data. (2009). *Harnessing the power of digital data for science and society.* Report of the Interagency Working Group on Digital Data to the Committee on Science of the National Science and Technology Council, January, 2009. Available at http://www.nitrd.gov/about/Harnessing_Power_Web.pdf. Accessed June 25, 2009.

[Conference Presentation] Lee. C. (2008). What do Job Postings Indicate about Digital Curation Competencies? Society of American Archivists Research Forum Research Forum, August 26, 2008, San Francisco, CA. Retrieved August 7, 2009, from http://www.archivists.org/publications/proceedings/researchforum/2008/CalLee-SAA-ResearchForum2008.pdf.

[report] Swan, A., & Brown, S. (2008). Skills, Role & Career Structure of Data Scientists & Curators: Assessment of Current Practice & Future Needs. Retrieved August 6, 2009, from http://www.jisc.ac.uk/media/documents/programmes/digitalrepositories/dataskillscareersfinalreport.pdf.