Analyzing Data Curation Job Descriptions

Center for Informatics Research in Science and Scholarship

Melissa H. Cragin — cragin@illinois.edu
Carole L. Palmer — clpalmer@illinois.edu
Virgil E. Varvel Jr. — vvarvel@illinois.edu
Aaron Collie — collie1@illinois.edu
Molly A. Dolan — madolan@illinois.edu

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 research through administered repositories or directly via the Web. Meeting the technical and organizational requirements ensuing from these trends will require a new, skilled workforce. Recent reports have identified a need for new programs to meet the demand (ARL, 2006) for skilled professionals trained as data librarians and data scientists (Swan & Brown, 2008). However, defined career paths for library and information science professionals interested in data curation or data science are not yet firmly established (IWGDD, 2009). As yet, studies of the emerging job market are rare (for an exception, see Lee, 2008), but these studies are needed to support the design of educational programs intended to prepare this emergent labor force, and to support the flow of newly-trained professionals into the market.

Context

The Data Curation Education Program (DCEP) at the Graduate School of Library and Information Science at the University of Illinois, has been engaged in a continuous, multi-faceted assessment of data management and curation needs by scientists and scholars. In order to inform data curation curriculum, 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 emerging careers, 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

One-hundred and twenty (120) online job postings within the Sciences and Social Sciences were harvested from 13 online sites from 8/27/08 - 5/01/09 and pooled with pertinent job postings identified by DCEP staff. Harvested job postings were retrieved with keyword queries using these terms: data; digital; research; librarian; archivist; data scientist; data curator; digital curator; data librarian; digital librarian; data archivist; digital archivist. For sites without an internal search function, job titles were browsed and job descriptions mined using Google site search. The resulting pool of job posts were reviewed individually for inclusion in the pool.

Findings

- **Location:** Job postings clustered in California, New York, Texas and Massachusetts.
- **By Domain:** 56.67% of jobs (68) fell directly into the Science domain, 11.67% of jobs (14) were from the Social Science domain, and 31.67% of jobs (38) captured did not specify the domain or spanned multiple domains.*
- **Largest Employers:** Academic Libraries made up 55.83% of postings (67) and Research or Data Centers made up 28.3% (34) of the postings.
- **Minimum Degree Level:** 55% of jobs (66) required a Master's Degree as the minimum education level.

* These numbers do not include jobs from the humanities domain; a subsequent paper will present those findings.

Harvesting Locations

Source	Count
ALA Joblist	12
ARL	13
HERC	4
I-ASSIST	13
IIA	6
Internal Res	16
LIS Jobs	5
SAS Google G	1
SDL Project	4
Science-Jobs	3
Simply Hired	23
Texas iSchool	19

Job announcements were chosen to represent several organization types:

Library and Information Science Sites
- e.g. ARL, LISjobs, Texas iSchool

Science and Technology sites
- e.g. National e-Science Center, Society of Systematic Biologists, Science Careers

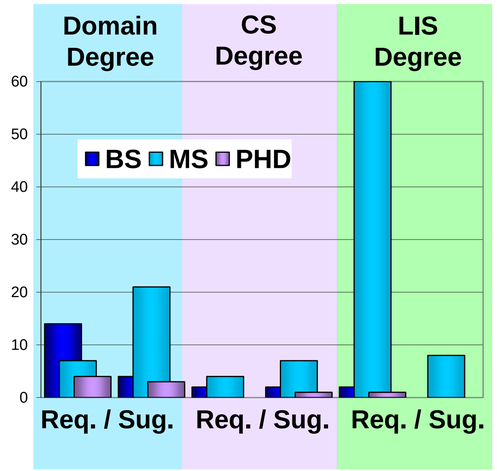
General Job Hosting sites
- e.g. Simply Hired, HERC

Overview & Skills

- **Average Salary**: $ 60,779.18
- **Average Required Experience:** 4.22 years
- **Geography:** 86.6% of jobs (104) were in the US vs. 13.3% (16) outside the US
- **Setting:** 58.33 % of jobs (70) are in a Library vs. 41.67 % of jobs (50) are outside the Library

Employer Type	Number of Jobs	Python / Perl / PHP / Java	DC / METS / MODS
Archives	5	0	0
Corporations	3	0	0
Computing Center	1	0	0
Laboratories	3	0	0
Libraries	Academic = 67 Public = 1 Special = 5	Academic = 9 Public = 1 Special = 1	Academic = 14 Public = 0 Special = 1
Museums	1	1	
Research or Data Center	34	1	2
Total	**120**	**11 (9.2%)**	**15 (12.5%)**

Education by Degree and Field



- **Domain:** 11.67% of jobs (14) require a bachelor level degree while 16.67% (20) suggest a master's degree in a domain
- **CS:** 95% of jobs (114) did not require a degree in CS
- **LIS:** 50% of jobs (60) require a Master's level degree in LIS

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. Retrieved June 25, 2009, from http://www.nitrd.gov/about/Harnessing_Power_Web.pdf.
- [Conference Presentation] Lee. C. (2008). *What do Job Postings Indicate about Digital Curation Competencies? Society of American Archivists 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.

Acknowledgements
We thank Xinrong Lei for her help in data processing and analysis. This work is part of the Needs Analysis activities funded as part of the IMLS grant award # RE 05-06-0036-06.

INSTITUTE of Museum and Library SERVICES