

Competencies and Responsibilities of Social Science Data Librarians: An Analysis of Job Descriptions

Jingfeng Xia and Minglu Wang

This study examines job announcements for social science data librarians and professionals to identify trends in the profession. A collection of 167 job postings in 2005-2012 from the International Association for Social Science Information Services & Technology website was analyzed on the frequencies of term occurrence and co-occurrence in job qualifications and responsibilities. The study verifies that employers valued non-technical skills as heavily as technical skills, and detects dissimilar emphases of data activities for data librarians and non-librarian professionals: the former on data discovery and collection, and the latter on data analysis and preservation. An increasing requirement of data management planning was also found for data librarians.



Academic libraries have long had individuals responsible for social science data services, providing access to government and other types of data. Traditionally, these individuals included the social science subject librarian or the government document librarian.¹ However, recently it has become more common for the social science data librarian position to be independent, probably as the result of emerging digital technologies and the Internet since the 1990s. In the beginning, many social science data librarians were “accidentally” transferring to the new capacities, normally from a reference or government librarian position.² The past several years witnessed a steady increase of demands for professionals who have

data management skills. These individuals must possess necessary attitudes toward data and be able to handle the intricate process of digital scientific data, along with the increasingly hot discussion of a big-data challenge both within and outside of academic settings. Not only has social science data librarianship evolved from traditional reference positions into dedicated roles and responsibilities, but some other special types of data librarianship, such as those focused on geospatial data and bioinformatics, have also come into play.³ Data librarianship in general, which is sometimes also called e-science librarianship, has now become a significant part of the academic library workforce.

The data librarian has experienced a change in job requirements from the early

Jingfeng Xia is Associate Professor in Department of Library and Information Science at Indiana University; e-mail: xiaji@iupui.edu. Minglu Wang is Data Services Librarian in John Cotton Data Library at Rutgers, the State University of New Jersey; e-mail: minglu@rutgers.edu. © 2014 Jingfeng Xia and Minglu Wang, Attribution-NonCommercial (<http://creativecommons.org/licenses/by-nc/3.0/>) CC BY-NC

days of identity ambiguity, such as an “accidental” data librarian, to today’s determined practices, such as a unanimously expected involvement in data management plans. Yet, the change occurred over such a short time that questions are still left and people keep wondering how to define data librarianship, including ways to echo the major big-data advances in job statements, along with ways to reflect the perceptions and adaptations that libraries have about scholarly communication. On one hand, libraries need to understand the changes so they can make the best use of their existing staff by adjusting and reallocating relevant positions. On the other hand, they want to be able to identify new requirements of necessary skills and knowledge for new hires. Studies focusing on data librarianship in general, and social science data librarianship specifically, are in demand, which can help highlight a unique new library service area that transforms the reference service heritage and entails a new digital curation perspective.

This study was designed to examine job announcements for social science data librarians with the purpose of reaching a better understanding of qualifications sought for relevant positions in the academic setting. We did not perceive a clear separation between geospatial data and social science data because geography is traditionally considered as social science, at least for the human aspects of geography.⁴ We also did not exclude job advertisements for social science data professionals in nonlibrary academic organizations because a comparison of data librarianship and other data positions will shed light on library data services’ unique features in the context of various roles and activities that data professionals have played in the entire community. Job postings in 2005 and afterward were collected from the website of the International Association for Social Science Information Services & Technology (IASSIST), with a total of 167 entries. An analysis of the frequencies of term

occurrence and co-occurrence was taken to evaluate the current condition of competencies and responsibilities required of social science data librarians. Research attentions were also paid to a chronological change and geospatial distribution of the requirements for job qualifications and responsibilities to identify possible trajectories of professional development in social science data management, service, and analysis.

Literature Review

Data Librarianship

Although libraries have a long tradition of providing access to and stewardship of text documents, and social science data services within academic libraries have already been assisting users in identifying and getting access to digital research data, it was not until the mid-2000s when data-sharing policies were adopted by various grant-funding agencies that researchers started examining librarians’ new roles in data management and data services. A series of discussions at several national and international conferences represent the concern of the library community about how people prepare for and enter the profession of data librarianship and what the critical roles are for librarians to undertake requiring the development of new skill set and probably novel career paths within the library workforce.⁵ Subsequently published reports provided a more extensive description of survey results on the career development of data professionals and the associated supply of specialist data curation skill set to the research society.⁶ At the same time, formal journal publications were found that focused on the same topic and were unanimously optimistic about new opportunities for librarians in the area of scholarly communication. Each study took a unique angle to evaluate the differences between a data librarian and a regular librarian including the future integration of data librarians into established library services and perspectives.

Jacobs was among the earliest who tried to theorize various levels of data services in an academic library setting. In addition to the traditional reference services, three new levels of data services were recognized by him in the early 1990s when research data started being acquired, stored, and presented in digital format. His proposed levels for data services include general data services, computing services, and library data services.⁷ Later, in the mid-2000s, Jacobs, along with Humphrey, who is among the founding figures of IASSIST envisioned for data services librarians to participate in the early stages of scholarly activities by helping researchers with the documentation process and ensuring that data discovered will be preservable, usable, and reusable for the long term.⁸ Similar to these scholars, Reznik-Zellen et al. identified three tiers of research data support services, emphasizing the importance of education, consultation, and building infrastructure to support library data services and goals while meeting the needs of research communities in a manner consistent with institutional missions and environments.⁹

In 2007, Gold published two articles, using for the first time the parallel of a research cycle and data life cycle and a downstream/upstream analogy to describe how data librarians could play a role in selecting, acquiring, and licensing scientific data, creating metadata for data discovery and description, organizing documentation for digital curation, and offering data preservation support.¹⁰ Lankes et al. even envisioned a new function of the general information professional for supporting the emerging e-research practices. They called these new roles "cyber-infrastructure facilitators" that will work with scholars more closely during a research process and within the context to discover extant tools, data sets, and other resources that can be integrated into the process.¹¹ Similarly, Creamer et al. perceived data curation and management competencies as part of the health science

librarians' as well as science and engineering librarians' professional development for research in these areas. In addition to technology competencies that are necessary to perform data-related tasks, the librarians that they interviewed also listed nontechnical competencies as job requirements, such as conducting data interviews with researchers to provide better and more accurate services.¹²

Job analyses of data librarianship were conducted by several researchers recently. Alvaro et al. examined a small sample of job advertisements of e-science librarians for the skills and requirements and concluded that "e-science librarianship is at present not a defined field and that the role of librarians in e-science is nebulous."¹³ Stanton et al. carried out a job analysis by interviewing controlled groups and observing students' summer internship in several data centers. They concluded with some suggestions that "the emerging eScience profession comprises a promising educational and research focus for information and library science in the coming decade and that science and R&D labs are an underappreciated setting for productive librarianship."¹⁴

Very few studies have been undertaken to analyze social science data librarianship. Gold's articles discuss data librarianship in general but include social sciences as part of the discussion.¹⁵ Another research project by Pryor explores, but only partially, the competencies and responsibilities of social science data librarians in a more systematic way under the context of current and future trends of e-science and e-social science. Pryor focuses on the effective collaboration with various partners to develop feasible workflows and create usable data collections. The data librarian is expected to possess the following unique skills: data appraisal and retention, advocacy, promotion, marketing, raising awareness, coordination of practices across unit and institution, negotiation skills, and complaints and expectation management skills. The data

librarian will also share the same skills with data manager for preservation and evaluation of data's economic values. Additionally, the data librarian will share the same skills of standards development with data scientists and will possess facilitation and communication skills.¹⁶

However, the above-mentioned job analyses did not devote their full length to run an exclusive examination of the emerging roles and responsibilities of social science data librarianship. Nor did they discuss in detail the changing nature of social science data librarianship in its short history.

Methods of Job Description Analyses in LIS

In library and information science (LIS), content analysis of job announcements has been popularly conducted to detect trends in the profession. Although a typical announcement may include such descriptions as position title, required education and training, preferred education and training, institution type, salary level, geographic location, job status (for example: tenure track, temporary), required experience and qualifications, preferred experience and qualifications, and job responsibilities, most content analyses have paid particular attention to job qualifications and responsibilities. A common strategy of conducting content analyses was to create appropriate categories and classify terms and expressions in a job description into these categories.

In a recent study by Choi and Rasmussen on digital librarian positions in academic libraries, the ALA (American Library Association) competency standards were used as the template of categorization.¹⁷ The content of required and preferred knowledge and skills in a group of 87 advertisements was coded into eight areas of job competency. The researchers counted the frequency of the coded content for each type of job competency, upon which further analysis was taken and conclusions were made. This introduction of the ALA standards, however, has its limitations because the ALA professional

competencies were compiled to include every function of librarianship, which may make the analysis of any particular type of librarianship disproportionate.

More efforts on content analyses of job descriptions used in-house-created categorizations. Researchers may have created a category system for coding purposes based on their review of the literature, counts of the frequency of content-bearing terms in job advertisements, and their knowledge of the specific type of librarianship being analyzed. Among others, studies by Hall-Ellis and Park et al. are the representatives that categorized meaningful terms and phrases in required and preferred qualifications as well as job duties to identify the current condition of cataloging librarianship.¹⁸ Nonetheless, a question remains in how to standardize categorization given the diverse uses of words and expressions in individual job descriptions. Not only could the category labels created by a study be arbitrary, but the process of classifying terms and phrases into a category is also vulnerable to personal judgment of the coders. For example, is "project management" a subcategory of "management"?

To make a job analysis more scientifically rigorous, some studies implemented a multiple-methods approach. In a study of e-science professionals, Stanton et al. used a combination of focus groups, interviews of data lab directors and researchers, and observations of summer graduate interns to review the knowledge, skills and capabilities needed for e-science professionals.¹⁹ Following Fine & Cronshaw's job analysis framework, Stanton et al. focused on identifying the dimensions of work characteristics, worker qualifications, and work organizations and exploring the educational implications of these on curriculum and program development in schools of information and library science.²⁰ However, their results, as well as those from other similar studies, are idiosyncratic to the small group of students and professionals involved and need support of supplementary discoveries.²¹

The power law of word distributions in text has been well known in documentation analyses.²² Yet, very few studies have applied such law to the examination of job descriptions. With today's computing, relevant techniques of word distribution analysis (such as text grouping and tag clouding) have been mature enough to process and visualize both structured and unstructured texts.²³ These techniques can rapidly provide readers with an overview of the most salient terms in a large corpus of text. For example, the tag clouding technique, which was originally designed to analyze and visualize labels and keywords for web pages, is able to categorize content and visually emphasize the popularity level of term occurrence and co-occurrence.²⁴ It is particularly useful for deriving inferences from an unstructured data source such as job descriptions that tend to use abbreviations, single words, brief expressions, short sentences, and so on to summarize requirements for a job. Similar strategies have been adopted by Alvaro et al. in a job analysis of e-science librarianship and are the methods of our own analysis of social science librarian positions.²⁵

Methods

IASSIST is an internationally known association of professionals who work in and with information technology and data services in support of research and teaching in the social sciences. It provides a central Web location for employers to market their job openings in the area of social science data. Unlike many other online job portals where posts are periodically removed, IASSIST keeps its postings to as early as 2005. It is an appropriate cutting-off time that coincided with the beginning of big-data popularity in support of scholarly activities in scientific disciplines and with the implementation of data mandatory policies by some institutions and funding agencies such as the open data mandate initiated by National Institute of Health (NIH) in 2003 and the National Science Foundation (NSF)'s data

management plan requirements began in 2011.²⁶ Recognizing the importance of data management in social sciences was grounded on the development of scientific data management and was, therefore, a little behind in time. This availability of job advertisements through IASSIST's website permits an examination of the short history of data librarianship in social sciences from the very beginning in the mid-2000s. The fact that job announcements first appeared on IASSIST's job portal in 2005 may signify the start of the profession as well, which can serve perfectly as the baseline year of our analysis and allow us to explore the chronological changes of the profession with regard to its responsibilities and competency requirements as well as preferences that are stated in the job advertisements.

All jobs, except for duplicates, on the website were entered into our data set in spite of the fact that a certain portion of them do not have the term "librarian" in their title or were not posted for a library. This enables us to analyze the unique aspects of social science librarianship as compared to other data professionals in a nonlibrary environment. At the same time, the international nature of IASSIST makes it possible to compare American jobs with jobs available in other countries, primarily in European countries.

Data were divided into several fields, namely job titles, background information of the institution, professional preparations of the applicant, required and preferred qualifications, and job responsibilities. An examination was first taken for the first three fields of job descriptions to help answer our research questions: How are these newly established social science data services integrated into the existing public or technical services? Whether an MLS is necessary for such positions? Are there additional degree requirements? And what are the duration and types of previous working experience that are required or preferred? For the latter three fields, the combination of an analysis of word occurrences and co-

occurrences in text and a content analysis was carried out to examine the scope and emphasis of social science data librarians' job qualifications and responsibilities. For diversely formatted job advertisements to be handled by text analyzing applications, appropriate word reorganization and adjustment is necessary.

1. *Word cleaning.* Following typical practice in content analysis, meaningless and useless words were removed, such as "the," "of," "at," "on," "and," "for," "which," which provide very little cognitive content. Also removed from the text include such words that are common in the context but are considered nonspecific with limited value of analysis, such as "knowledge" and "applicant," unless these words could be paired with adjacent word(s). In doing so, we followed the quantitative criterion used to measure the level of specificity introduced by Milojević et al.²⁷
2. *Word coupling.* In some cases, a particular relationship between words makes more sense than individual words. Syntactic patterns were detected to define word relations so that "government resources" might be used to represent "government" and "resources" as two separate words. Similarly, "numeric data" is more meaningful while coupled than treating them as two separate terms. This step of data processing required special care for (a) contextual relations and (b) the original frequencies of every term affected.
3. *Word standardization.* Terms were also evaluated and adjusted to minimize unnecessary variations and confusions. For example, one term or co-term might be adopted to replace various similar terms or co-terms that describe the same object/concept/activity, such as

using "statistical packages" to replace all occurrences of the following co-terms: "statistical applications," "statistical software," "statistical tools," and so on. Similarly, the co-term "metadata standards" was used to replace "metadata schemes" wherever the latter was discovered.

Off-the-shelf online text analyzers and tag-clouding generators were applied for the processed texts that are still unstructured and retain their original orders and frequencies. Two different methods of data analysis were applied to the fields of required and preferred qualifications: (1) analysis of word frequencies measured by the occurrence and time of each word in one advertisement, job advertisements of a year or other means of grouping, and (2) analysis of co-term frequencies for the cognitive structure of data librarianship based on the occurrence and time of all co-terms. The reason to include multiple types of analysis is that the job advertisements have presented significant variations in the use of terms, phrases, and even sentence structures. The application of any single type of analysis as mentioned above may not be able to provide more consistent and reliable results than the combination of all types.

Data analysis of the job responsibilities was conducted by applying similar term occurrence and co-occurrence approaches as well as a content analysis based on categorization of responsibility descriptions for expressions that are data related. In other words, instead of evaluating all responsibility content, we focused on technical segments of the descriptions. A categorization dictionary was created by consulting various sources including (a) ICPSER's data life cycle model, (b) the Data Documentation Initiative, (c) levels of data services proposed by some researchers for the social science data librarianship, (d) appearance of data-related terms and expressions in our data set, and (e) our understanding of the trends of the data management

profession.²⁸ Particularly helpful are the first two models that articulate a typical research process involving social science data for “key considerations germane to archiving at each step in the data creation process ... that come into play across all stages of the data life cycle.”²⁹

Findings

Employer Types, Job Titles, and Background Requirements

There are various types of employers that posted these 167 job announcements. Table 1 presents the frequencies and percentages of each category where over half (50.30%) of the employers are college and university libraries. Some other service units (7.78%), such as IT support departments, within colleges and universities were also hiring dedicated professionals to provide data-related services. Around 10 percent of the employers are data centers affiliated with a university or university consortium, of which the especially known data center is the Interuniversity Consortium for Political & Social Research (ICPSR) based at the University of Michigan. In Europe, national level data archives or centers (for example, the UK Data Archive) are taking the responsibility of curating data and providing services on research data sets. At the same time, a significant number (13.77%) of research institutes or research projects, either operated independently or

within a university, was hiring their own in-house data professionals. The remaining employers are U.S. Federal Reserve and the Library of Congress (3.59%).

A survey of the ARL membership in 2009 found that about 37 percent of responding ARL libraries in the United States and Canada implemented data support services, and more than 40 percent of them were in the planning stages.³⁰ Although a majority of the libraries have been “reassigning existing staff or providing training to existing staff as part of an overall strategy to incorporate e-science responsibilities into their current portfolios,” the rest created new positions to hire staff specifically to provide e-science services as part of their overall strategy.³¹ The investment of resources in data services even at difficult times of budgeting suggests a strong priority among academic libraries and their host institutions, which explains the large number of newly created data services librarian positions in the past years.

Conversely, there are very few other organizations (2.99%) and private companies (only one post) in the job list. Yet we believe the demands for social science data management and services in both the public and private sectors outside academia are largely disproportionately represented. Whereas “there is great potential for employment of eScience professionals on the front lines of com-

**TABLE 1
Employer Categories**

	Frequency	Percentage
College/University Library	84	50.3%
Research Institute/University Research Institute/Project	23	13.8%
University Consortium Data Center	16	9.6%
College/University IT Support	13	7.8%
National Archive/Data Center	19	11.4%
U.S. Federal Reserve/Congressional Research Service	6	3.6%
Organization	5	3.0%
Company	1	0.6%
Total	167	100.0%

	Frequency	Percentage
General Public Service	37	44.0%
Subject Library	22	26.2%
Data Center/Data Service	15	17.9%
Digital Curation/ Repository Service	7	8.3%
Technical Department	3	3.6%
Total	84	100.0%

mercial activities, particularly those in the pharmaceutical, medical, and biosciences sectors," the public segment, including various types of research institutes, associations and not-for-profit organizations, has heavily relied on social science data to carry out analyses for economic modeling, political forecasts, criminal preventions, transportation planning, and the like.³² IASSIST seems to be the wrong platform to look for nonacademic job postings.

Whenever college and university libraries were facing a changing scholarly environment, they tended to work on initiating new services to meet the expressed or potential needs of their local research communities. One may wonder how academic libraries have developed this new social science data librarianship as a position within their existing organizational system. Our summary of the job advertisements (see table 2) shows that this new position is more often integrated into existing library structure than separated into an independent data center or data services department. Of the 84 college/university libraries that were hiring, 44.05 percent of the data librarians reported to a general public services department, such as a library reference or research and instruction department. Around 26.19 percent of the data librarians were

located within a subject library such as a social science library, map collection or government information section, while only about 17.86 percent of them were hired for a data center or data services unit within the college or university library. When an academic library started taking on more responsibilities for curating research data, the social science data librarian position was also allocated to its digital curation or repository services department (8.33%), and even its technical services department (3.57%).

Because of the variety of employer types and library departments hosting these social science data services professionals, we expect the job titles to be very diverse. Table 3 shows the frequencies and percentages of various groups of jobs appearing in IASSIST's job list. Not surprisingly, librarian is ranked first among all posts (44.12%). The second largest group is data specialist/consultant (12.94%). In Europe, especially among national data archives or centers, data professionals often bear the title of officers (6.47%). There are considerable needs for project/program

	Frequency	Percentage
Librarian	75	44.1%
Specialist/Consultant	22	12.9%
Project Manager/Team Leader/Coordinator	15	8.8%
Head/Director	14	8.2%
Officer/Senior Officer	11	6.5%
Researchers	9	5.3%
Developer/Technologist/ System Analyst	9	5.3%
Archivist/Curator	7	4.1%
Data Analyst	5	2.9%
Data/Content Manager	3	1.8%
Total	170	100.0%

manager or team leader or coordinator (8.82%), which demonstrates the collaborative aspect of social science data services. There is a relatively similar amount of posts for administrative head/director (8.24%) for various data units. In comparison, a small number of data professionals were hired to take research responsibilities and were thus called researchers (5.29%). At the same time, the remaining job titles represent

the heavy technical functions of some of the data professionals including developer/technologist/system analyst (5.29%), archivist/curator for different levels of data repositories (4.12%), data analyst (2.94%), and data/content manager (1.76%).

If one assumes that an MLS degree was required by jobs mostly in academic libraries, the following findings do support the assumption (see table 4). Master of Library Science (MLS) from an ALA-accredited library program is the most often mentioned degree requirement: 31.74 percent of the job advertisements mentioned at least an MLS, and 6.59 percent of them accepted an MLS or another master’s degree, while around 4.19 percent required an MLS in addition to another graduate subject degree. As many as 11 job announcements (6.59%) asked for a PhD degree, which fell mostly under the employer category of research institutes. The remaining 45 jobs (26.95%) did not specify a degree requirement. This large percentage of requirements for an MLS degree indicates potentials for LIS programs to expand their focuses to include training of qualified data services professionals. Many LIS programs have indeed provided corresponding instructions, particularly in the area of data curation. However, very few of them have paid special attention to social science data services, and there are no established standards to regulate the curriculum.

	Frequency	Percentage
Not Specified/No Information	45	26.9%
Bachelor	17	10.2%
Master	23	13.8%
MLS or Master	11	6.6%
MLS	53	31.7%
MLS + Graduate Subject Degree	7	4.2%
PhD	11	6.6%
Total	167	100.0%

Some employers asked for an advanced degree in a particular discipline, rather than training in library science, pertaining to the job responsibilities so that the data services professionals will have necessary domain knowledge and are able to work with researchers on designated research projects (13.8%). This will bring up challenges for providing continuing education in data handling, which has actually been very common in the early days of data librarianship when people without data background or the MLS entered the profession—so called the “accidental” data librarian.³³ Professional associations such as the IASSIST have taken the responsibility of offering workshops and other types of short trainings during conferences as well as online. It is skeptical whether these brief instructions are thorough and systematic enough to provide necessary preparations given the depth of data-related activities; for instance, quantitative data analysis requires a whole different set of technical skills from digital data preservation.

Only less than half of the advertisements specified a required number of years of experience (see table 5), of which 1–3 years of work experience are the most popular requirements (26.35%). The remaining employers hoped to see more years of experience, which are, however, mainly for managerial and other senior positions. This relatively large number of

	Frequency	Percentage
Not Specified/ No Information	107	64.1%
1 Year	8	4.8%
2 Years	20	12.0%
3 Years	16	9.6%
4 Years	4	2.4%
5 Years	6	3.6%
7 Years	1	0.6%
8 Years	3	1.8%
10 Years	1	0.6%
Several	1	0.6%
Total	167	100.0%

not-specified may reflect the concurrent condition of social science data librarianship, namely that the field is too new to accrue experience for job candidates.

Although most employers did not quantify the experience requirements, many specified the particular types of experiences that they were looking for. By examining these experience requirements and identifying some common themes, we coded the various requirements into the following groups as shown in table 6, with the frequencies of the occurrence of each category. Experience with either GIS or statistical software was mentioned most frequently with a total of 41 jobs (17.6%). Yet, this does not count job requirements for experience working with either spatial or numeric data (11.6%), which are alternative descriptions of the GIS and statistical work. Previous services experience in an academic library (12.9%) and teaching and training experience in an academic environment (16.8%) were also frequently described as the requirements. Managerial, supervisory, and leadership

	Frequency	Percentage
GIS and/or Statistical Software	41	17.6%
Academic Library Services	30	12.9%
Spatial and/or Numeric Data/Government Information	27	11.6%
Research/Academic Environment	27	11.6%
Management/Supervising/Leadership	20	8.6%
IT (Web, Programming)	17	7.3%
Data Services	13	5.6%
Teaching/Training	12	5.2%
Data Archive/Preservation/Repository	12	5.2%
Data Documentation/Metadata	10	4.3%
Data Management	7	3.0%
Grant	4	1.7%
Databases	4	1.7%
Data Analysis	4	1.7%
Data Creation/Collection	3	1.3%
Data Purchasing Contract	2	0.9%
Total	233	100.0%

background are counted 20 times (8.6%), especially for head/director or senior officers' positions. In addition, grant application and implementation experience was slightly desired (1.7%).

IT background, including web and programming experiences, was asked in 17 posts (7.3%). Other requirements found in the job advertisements include experience in data services (5.6%), in specific data archive and preservation or working with a digital repository (5.2%), data documentation/metadata experiences (4.3%), and data management (3%). A few specifically identified skills were also visible, including requirements for relational databases (1.7%), data analysis (1.7%), data creation/collection (1.3%), and data purchasing/contracting (0.9%).

In regard to geographic distributions of the employers, the United States is far out front, contributing 64.1 percent of all job posts, followed by the United Kingdom (16.2%) and Canada (8.4%). Some other countries in Europe and Australia also made the list. Outside these regions, Qatar, Singapore, and the United Arab Emirates have each posted one job on the website (see table 7). We anticipate

	Frequency	Percentage
Australia	6	3.6%
Canada	14	8.4%
Germany	5	3.0%
Ireland	1	0.6%
Netherlands	2	1.2%
New Zealand	1	0.6%
Qatar	1	0.6%
Singapore	1	0.6%
Sweden	1	0.6%
U.K.	27	16.2%
United Arab Emirates	1	0.6%
U.S.	107	64.1%
Total	167	100.0%

a gradual increase of countries from Africa, Asia, and Americas in the future, much like the diffusion patterns of digital projects and the open access movement from the West to the rest of the world as observed in the past.³⁴

Required and Preferred Qualifications

Word frequency analysis weighs all terms entered in the text analysis application and yields a long list of ranked term occurrences and frequencies. Table 8 lists the top ten terms used in required qualifications. It is not surprising that "data" as an individual term is the most frequently used word in the text of required qualifications, while this calculation does not consider any contextual relationship between "data" and other words. Interpersonal skills are heavily valued so that "communication" is ranked second and "interpersonal" is eighth in the list. The third most popular word "skills" delivers little meaning if it is not coupled with other words. When "management" is not distinguished by its application in project manage-

Word	Frequency	Rank
Data	7.1%	1
Communication	3.0%	2
Skills	2.9%	3
Management	2.6%	4
Analysis	2.5%	5
Statistical	2.1%	6
Service	2.0%	7
Interpersonal	1.7%	8
Packages	1.6%	9
Project	1.5%	10

TABLE 9
Top Ten Terms in Preferred Qualifications

Word	Frequency	Rank
Data	6.6%	1
Management	1.7%	2
GIS	1.5%	3
Metadata	1.4%	4
Technologies	1.3%	5
Library	1.3%	6
Statistical	1.3%	7
Resources	1.3%	8
XML	1.2%	9
Web	1.2%	10

ment, data management, personal management, or other types of management, it is the fourth most used word. The rest of the top ten words are rather technical.

When the frequency of word distributions in preferred qualifications is examined, a different pattern from that in required qualifications is detected (see table 9). Although “data” and “management” are still among the most found ones, specific technological terms have become dominant in the text. Here, the terms “GIS,” “metadata,” and “statistical” occur more frequently than terms representing personal characters, communication, and management skills; and even narrower technical terms, such as “XML,” have made the top ten terms.

Individual terms may not provide more cognitive values than co-terms due to the importance of contextual structure in any job descriptions. “Data” as an individual term in such descriptions has little meaning until it is joined by another term such as “data access,” “data analysis,” and “data preservation.” Therefore, we ran a co-term analysis for both required and preferred

qualifications. In the resulting top ten list for required job qualifications (see table 10), “communication skills” as a co-term jumps to the number one position, and its occurrence rate to other top ten terms is indistinguishable from that of the single term analysis. This popularly highlighted skill set has also been found in job announcements of many other types of librarianship in previous studies, and is corresponding to what the ALA has compiled in its professional competency statement within which eight categories are included, namely, professional ethics, resource building, knowledge organization, technological knowledge, knowledge dissemination (service), knowledge accumulation (education & lifelong learning), knowledge inquiry (research), and institution management.³⁵ Communication skills are the important component of both technological knowledge and institution management categories in the ALA statement and of data librarianship requirements in our data set. The next frequently appearing co-term occurrences are “statistical packages,” “project management,” and “metadata standards,” which reflect employers’ inclination of technical abilities and management characteristics. The same technical and management requirements are also found in all popularly jointed terms except

TABLE 10
Top Ten Co-terms in Required Qualifications

Expressions	Percentage	Prominence
Communication Skills	2.0%	61.5
Statistical Packages	1.0%	44.3
Project Management	1.0%	73.5
Metadata Standards	0.9%	46.5
Changing Environment	0.8%	42.4
Geospatial Data	0.8%	51
Numeric Data	0.7%	47.2
Data Analysis	0.6%	40.9
Problem Solving	0.5%	58.7
Data Management	0.5%	62.8

the fifth ranked “changing environment” and the ninth ranked “problem solving.”

When co-term distributions in preferred job qualifications are examined, it becomes apparent that all top ranked co-terms are technologically related, except one term: “academic library” (see table 11). This finding is consistent with the analysis of single word distributions mentioned above and indicates that employers value nontechnical skills as high as technical skills and pay special attention to the maturation and independence level of job candidates by requesting excellence in communicating to all constituencies, collaborating with various partners, and managing assigned tasks. It is common in the job descriptions that applicants are required to possess the ability to communicate with researchers to identify their data needs and help them locate appropriate data sources, the ability to work well in a team environment to deliver capable library services, the ability to collaborate with data providers (individual, local, state, national, and international) to build data collections, and the ability to quickly learn new material to develop standards for data literacy.

These findings have an educational implication. For a long time, LIS programs have set their goals to prepare socially responsible graduates for “fulfilling careers characterized by ethical practice, professional values, analytical skill, leadership, and lifelong learning.”³⁶ While it is important for future librarians to become experts in their unique areas of librarianship, training in leadership and interpersonal skills has always been a high priority. Data librarians, in addition to being competent for providing data services, need to be collaboratively working with all constituencies in an efficient and effective manner. LIS instructors ought to

Expressions	Frequency	Prominence
Metadata Standards	0.8%	43.8
Data Resources	0.6%	64.8
HTML & XML	0.5%	65.4
Data Management	0.5%	73.2
Data Analysis	0.5%	50.8
Academic Library	0.5%	55
Statistical Data	0.5%	56.3
Information Systems	0.4%	72.4
Collection Development	0.4%	29.5
Spatial Data	0.4%	43

regenerate pedagogical strengths in the offering of various management-related courses and use the real-world cases to train students in adapting to an ever-changing academic environment.

Fine and Cronshaw’s job analysis framework set worker characteristics into three types: knowledge areas, skills, and abilities.³⁷ According to their framework, knowledge denotes a body of information being memorized or mastered such as knowledge of databases. Skills represent learned competencies depending on education, training, and improvement with practice such as database management skills, while abilities signifies one’s potentialities in a specified area such as the ability to quickly learn new material. We feel it unnecessary to organize the skill set in the job descriptions of social science data librarians into these divisions because we believe competencies are trained but not born. The significance of professional training can never be overestimated.

To trace chronological changes, if any, of required job qualifications, a tag clouding analysis was applied to measure co-term occurrences. Tag clouding (also known as word clouding or word crowding) provides a proper means of data visualization for weighted terms or co-terms in free text. In the visualization, tags are displayed with varied font sizes or colors

to specify their frequencies and importance so that viewers can instantly verify their relative prominence in the text. We believe that tag clouds, by showing an aggregate of tag-usage statistics, are capable of enhancing our understanding of the real requirements that employers wanted to highlight in their job descriptions. When all co-terms are processed through an appropriate algorithm, biases caused by variations in personal judgment of any artificial categorization can be reduced to certain extents.

In the following tag-clouding analysis, co-terms are presented in an arbitrary mode, while the calculation of tag frequencies is similar to other established tag-clouding algorithms.³⁸ Each graph in

figures 1–8 contains all co-term distributions in the required qualifications of one year’s job descriptions. With eight graphs in a chronological order from 2005 to 2012, we can easily detect the consistency of employers’ desire for personal characters, communication and management skills (that is, “communication skills” and “project management” abilities) of job candidates as were discovered in the analysis of co-term frequencies in required job qualifications (see table 11). This consistency is observable throughout time. Also as mentioned above, the visual results by no means devalue employers’ emphasis on the requirements of technical competencies, which are relatively underrepresented due to the diverse

FIGURE 1
A Chronological Change of Co-Term Occurrences in Required Qualifications (2005)



FIGURE 2
A Chronological Change of Co-Term Occurrences in Required Qualifications (2006)



FIGURE 3
A Chronological Change of Co-Term Occurrences in Required Qualifications (2007)



FIGURE 4
A Chronological Change of Co-Term Occurrences in Required Qualifications (2008)



FIGURE 5
A Chronological Change of Co-Term Occurrences in Required Qualifications (2009)



FIGURE 6
A Chronological Change of Co-Term Occurrences in Required
Qualifications (2010)



FIGURE 7
A Chronological Change of Co-Term Occurrences in Required
Qualifications (2011)



FIGURE 8
A Chronological Change of Co-Term Occurrences in Required
Qualifications (2012)



descriptions of data-related activities. In fact, various technical co-terms are also visible in all tag clouds.

Responsibilities

The analyses of term and co-term occurrences in the descriptions of social science data librarian responsibilities were taken among two different groups of posted positions: that is, jobs available in academic libraries and jobs in a nonlibrary setting including university-affiliated centers and institutes and individually operated organizations such as the U.K. Data Archive. This division will allow us to observe any differences in the requirements of job responsibilities between two dissimilar groups of data professionals. Tables 12 and 13 display the results of single-term frequency analysis for the two groups side by side for a quick comparison. Although most of the top-ranked terms are found in both tables, library jobs set more requirements for responsibilities in providing reference services, instructional programs, and collection development, whereas nonlibrary jobs require more support to and involvement in research activities. “Services” is a common word that appeared in the responsibility list of both groups; yet,

Word	Frequency	Percentage
Data	628	15.6%
Services	169	4.2%
Resources	86	2.1%
Research	86	2.1%
Reference	81	2.0%
Management	76	1.9%
Information	75	1.9%
Instruction	63	1.6%
Collections	50	1.2%
GIS	50	1.2%

nonlibrary positions tend to highlight “collaboration” as a general requirement and “statistics” as a specific requirement.

Similar distinctions can also be discovered in co-term frequency analysis as shown in tables 14 and 15, where library jobs require planning, implementing, and engaging in reference services and collection development activities. Unlike nonlibrary positions that emphasize the technical components of data management (such as “data preservation” and “data curation”), librarians are required to provide quality services for “data discovery” and “data access.” Please note that, in the term co-occurrence analyses, “social science” as a co-term was not included, while “data” as a term was counted if it has a contextual relation to another word.

A chronological analysis of job responsibilities was also conducted to examine any possible changes in the job requirements throughout time. Once again, the comparison is separated by year, and no group separation has been taken due to the limited numbers of job advertisements. Each tag clouding image in figures 9–16 represents job posts of a year from 2005 to 2012. In contrast to the tag clouding visualizations for required job competencies that

Word	Frequency	Percentage
Data	608	17.2%
Management	102	2.9%
Services	81	2.3%
Research	78	2.2%
Support	54	1.5%
Statistical	51	1.4%
Analysis	36	1.0%
Collaboration	36	1.0%
Access	34	1.0%
Information	34	1.0%

TABLE 14
Top Ten Co-terms in Job Responsibilities for Library Positions

Expression	Frequency	Percentage	Prominence
Data Management	58	1.4%	36.9
Data Services	35	0.9%	57.1
Numeric Data	35	0.9%	63.3
Data Use	26	0.6%	46.2
Data Resources	25	0.6%	61.3
Spatial Data	23	0.6%	67.9
Reference Services	20	0.5%	57.6
Data Access	19	0.5%	41.6
Collection Development	19	0.5%	48.7
Data Discovery	18	0.4%	30.5

TABLE 15
Top Ten Co-terms in Job Responsibilities for Nonlibrary Positions

Expression	Frequency	Percentage	Prominence
Data Management	75	2.1%	40.1
Data Access	21	0.6%	41.9
Data Service	18	0.5%	50.1
Data Use	18	0.5%	50.8
Data Archive	17	0.5%	61.6
Data Analysis	15	0.4%	49.7
Data Preservation	13	0.4%	37.9
Geospatial Data	12	0.3%	61.2
Statistical Data	12	0.3%	71
Data Curation	11	0.3%	46.3

do not reveal obvious changes over the time, this analysis instead shows some shifting patterns of job responsibilities. In the early years of this time sequence, the responsibility requirements focused on various forms of instructions, from delivering formal classroom teaching to arranging specialized workshops, and on different types of library services, from providing general references to offering specific data services. An effort to increase data-awareness among researchers and to provide data support was prioritized. This instruction- and service-orientation in the job responsibilities has changed to an increasing focus on data management

from 2008 on. Relevant to the shift is a later appearance of the “data management plans,” which may well reflect an increasing requirement for the planning and implementation of data access mandate policies by many grant funders, government agencies, publishing entities, and scholarly organizations at the concurrent time.³⁹

The first open access mandate policy for scientific data sharing was adopted by the National Institute of Health (NIH) upon the Congress’ demand in 2004.⁴⁰ Initially, it was a relatively weaker policy that only *requested* grantees to self-archive raw data of any NIH-sponsored projects in a recog-

FIGURE 15
A Chronological Change of Co-Term Occurrences in Job Responsibilities
(2011)



FIGURE 16
A Chronological Change of Co-Term Occurrences in Job Responsibilities
(2012)



nized data repository. A few years later, the policy was strengthened to *require* data sharing; and in fall of 2007, both the House of Representatives and the Senate adopted an appropriations bill demanding an OA publications mandate at the NIH. Corresponding to the publications mandate implementation was the requirement for a data management plan as an integral part of any grant proposals to NIH. At almost the same time, the National Science Foundation (NSF) was suggested by its board to consider a similar data management plan requirement to be part of the NSF grant process in 2005, which was implemented several years later.⁴¹

The NIH data policy was designed for data sharing primarily in the areas

of biomedical sciences, but was also relevant to many closely related social science fields such as psychology. On the other hand, the NSF public access mandate has specific requirements for social, behavioral, and economic sciences that explicitly define social science data as the recorded factual material commonly accepted in the scientific community as necessary to validate research findings, including original data and metadata.⁴² Its data management plan requirements serve as a driver for change in descriptions of social science data librarianship positions. It is, therefore, not surprising to observe the addition of an assignment in data management plans to the job responsibilities of data librarians who

now have been actively involved in assisting in the development of NIH and NSF grant proposals at many academic libraries.⁴³

The tag clouding visualizations for job responsibilities in figures 9–16 have also indicated an increasing importance of requirements for collaborative efforts (its alternative for library jobs is liaison responsibilities) through time. As a service provider, the social science librarian needs to work closely with different units of a library or data center, various departments of an institution, and diverse groups of organizations at local, national, and international levels. S/he will cooperate with scholars to support their research and instructional endeavors; with data providers, internal as well as external, of an institution, to acquire and prepare data sets; with technology personnel to convert, preserve, and curate digital data; with team members to ensure high standards of raw data discovery, access, and retrieval; and with experts to assist in scientific data manipulation, analysis, and visualization. This finding helps confirm the importance of communication and interpersonal skills in required competencies of a job candidate.

The ARL survey in 2009 found that data services provided by academic libraries were primarily through a team effort or some combination of individuals, units, and teams working together. Collaboration has been the major way of effort for data support and “will continue to be an important method to address the enormity of the challenges posted by e-science,” because data sets generated by modern scientific devices are often substantial in volume and require extensive resources to manage. An example is the necessity and benefit of the involvement of subject liaisons who traditionally offered a range of services for social science data. The state of liaison roles and responsibilities will allow data services librarians to work together with liaisons to provide data support in such areas as

analysis of data set deposit requirements, development of data management plans, instruction for data practices to researchers, collection and dissemination of social science data sets, and design of data preservation standards.⁴⁴

For the content analysis, job responsibilities not closely associated with data activities were excluded to provide a focused justification. The concept of data life cycle is becoming more and more important for social science data professionals when we realized that the implementation of research data documentation and the consideration of data preservation cannot simply be wrapped up at the very end of a research project; rather, they are ongoing processes that need to be started even before research data are collected and should be incorporated into every stage of the research cycle. We synthesized the data life cycle model for DDI’s metadata schema and the data life cycle model developed by ICPSR to advise on the best practices of social science research data preparation and archiving. “Data Management Plan” is considered to be the first stage of data life cycle based on ICPSR’s recognition of its increasing importance and the reality of the funding agencies’ requirement for a grant proposal. “Data Discovery” is listed as the second stage, consistent with DDI’s data life cycle model and the long tradition of library services related to information retrieval. “Data Collection,” “Data Analysis,” “Data Sharing,” and “Data Preservation” are defined stages in both DDI and ICPSR models. All data-related job responsibilities are coded accordingly, and the frequencies of each stage are summarized to show the different concentrations of these jobs with the ongoing data life cycle (see table 16).

It becomes apparent that social sciences data professionals are still performing traditional primary services in the stages of data discovery, data collection, and data analysis.⁴⁵ At the same time, support for data preservation as a relatively new

TABLE 16
Categories of Data Activities

Data Life Cycle Stage	Sample Terms and Expressions	Frequency/ Percentage (Library Jobs N=84)	Frequency/ Percentage (Nonlib Jobs N=83)
Data Management Plan	Data Management, Data Management Plan	23 / 27.38%	20 / 24.09%
Data Discovery	Data Access, Data Identification, Data Discovery, Data Consultation, Data Reference	48 / 57.14%	19 / 22.89%
Data Collection	Data Entry, Data Collection, Local Collections, Data Acquisitions, Databases	37 / 44.04%	19 / 22.89%
Data Analysis	Data Visualization, Statistical Analysis, Numeric Analysis, Spatial Data Analysis	26 / 30.95%	25 / 30.12%
Data Sharing	Data Dissemination, Data Use, Data Sharing	20 / 23.80%	18 / 21.69%
Data Preservation	Data Repository, Data Archive, Data Curation, Ddi, Data Storage	27 / 32.14%	24 / 28.92%

task of research data services has already been taken up by the profession quickly, probably due to the professional spirit of performing data stewardship. Although data management planning has not been the most needed support area, in comparison with other data life stages, we have observed an increasing emphasis of it in the recent years. The reason that data sharing is the least mentioned support area could be because it is actually implied in data preservation and archive services and thus not counted separately.

Table 16 compares the same categories of job responsibilities between library jobs and nonlibrary jobs. The same as the co-term analysis, social science data services in libraries are assuming more duties related to data discoveries and data collections. Particularly, 57.14 percent of library jobs mentioned data discovery service, while only 22.89 percent of nonlibrary jobs required it. Likewise, 44.04 percent of library positions asked for data collection development, but only 22.89 percent of nonlibrary positions listed this responsibility. On the

other hand, nonlibrary jobs have had a strong accent on data analysis functions (30.12%), which is its largest percentage. In regard to the task of data preservation, both types of data professionals gave it a higher mark.

With the purpose to provide empirical data on qualifications and skills wanted for social science data librarians, this study reveals changing practices in the recent years of library services. The study clearly shows that social science data services are in high demand and are rapidly evolving. It is, therefore, critical to raise understanding among data professionals, LIS educators, and students of the importance of job competencies so that they can be informed of the current trends of data discovery and analytics. They need to be acquainted with professional standards and be prepared to participate in collaborative projects at various levels to promote awareness of data services for guiding national and international practices on one hand and for meeting the research needs of local communities on the other hand.

Research Limitations

It is necessary to mention that the size of our data set is relatively small. When all jobs were broken into year or other groupings, the representatives of job descriptions became limited. In a report on skills, role, and career structure of data scientists and curators, Swan and Brown distinguished data-related roles into four groups: data creator, data scientist, data manager, and data librarian.⁴⁶ In our analysis, jobs are only separated for those posted by libraries and those posted by nonlibrary organizations because of our narrower focus on social science data professionals and a smaller number of available data.

Conclusion

This study has adopted alternative strategies to analyze job descriptions of social science data librarians with a focus on the descriptions of job competencies and responsibilities. Instead of following the common approaches of content analysis for job ads that categorize and classify content items, we applied text analysis and tag clouding techniques to measure term occurrence and co-occurrence in the body of unstructured text. Only a small portion of our analysis was designed to code content into categorization; and our coding effort was restrained to technical aspects of the job responsibilities. We believe this combination of various analytical techniques will be able to reduce uncertainties created in any coding process owing to differences in the use of advertisement languages and personal judgment of coders.

The study verifies that scientific data has brought new opportunities to library services. With the new elements of data support, librarians and other data services professionals have been able to get directly involved in research enterprise.

The research cycle signifies the data life cycle within which data librarians found their particular niche and assumed novel responsibilities. Following ICPSR's data life cycle diagram, one will find how data librarianship has presented value in every phase of the data creation, use, and curation process, beginning from helping researchers to develop data management plans, discovering appropriate raw data for research projects, conducting data analyses, preparing data for free sharing, and depositing data to digital repositories to preserve data for accessibility, integrity, and longevity.⁴⁷ These responsibilities apply to the job requirements for both librarians and nonlibrarians, although each group has its own focuses: the former with more responsibilities in data discovery and collection and the latter with more responsibilities in data analysis and preservation.

Results of the job announcements analysis also reveal that employers have valued nontechnical skills as heavily as technical skills, if not more. Measuring the frequency of term and co-term distributions yielded a strong focus of job requirements on the interpersonal skills, management skills, and problem-solving skills of a job candidate. In regard to job responsibilities, assisting in the development of data management plans has become increasingly important for librarians, indicating the efforts of many academic libraries to better serve scholarly activities by taking advantage of library specialties and resources, and the changing environment of scholarly communication that values data sharing and reuse. Further studies may explore how data librarians and professionals have actually participated in various research projects of faculty, and how faculty scholars value their contributions of data-related support.

Notes

1. Stefan Kramer, "Data Librarianship: Past, Present, Future, Challenges, Opportunities," invited presentation to staff of Leibniz Institute for the Social Sciences in Bonn, Germany (Dec.

2010), available online at www.ecommons.cornell.edu/handle/1813/19484 [accessed 12 January 2013].

2. The 2005 IASSIST Conference organized a session entitled "Discovering a Profession: The Accidental Data Librarian" to discuss how people prepared and entered the profession of data librarianship. The concept of "accidental" data librarian was introduced to describe those who had not pursued a career in data librarianship but by happenstance discovered the profession. Relevant discussions can be read on *iBlog*, available online at www.iassistdata.org/blog/discovering-profession-accidental-data-librarian [accessed 12 January 2013].

3. Anna Gold, "Cyberinfrastructure, Data, and Libraries, Part 1: A Cyberinfrastructure Primer for Librarians," *D-Lib Magazine* 13, no. 9/10 (Sept./Oct. 2007); Anna Gold, "Cyberinfrastructure, Data, and Libraries, Part 2: Libraries and the Data Challenge: Roles and Actions for Libraries," *D-Lib Magazine* 13, no. 9/10 (Sept./Oct. 2007).

4. Richard Peet, "The Development of Radical Geography in the United States," *Progress in Human Geography* 1, no. 2 (Jun. 1977): 240–63.

5. See note 2; also, an ARL-NSF workshop was held in October 2006 for participants to discuss various views of librarians' roles in cyberinfrastructure, which can be read in the report available online at www.arl.org/bm~doc/digdatarpt.pdf [accessed 12 January 2013].

6. Alma Swan and Sheridan Brown, *Skills, Role & Career Structure of Data Scientists & Curators: Assessment of Current Practice & Future Needs* (London: JISC Report, Jul. 2008); Catherine Soehner, Catherine Steeves, and Jennifer Ward, *Science and Data Support Services: A Study of ARL Member Institutions* (Washington, D.C.: Association of Research Libraries, Aug. 2010), available online at www.arl.org/bm~doc/escience_report2010.pdf [accessed 12 January 2013].

7. Jim Jacobs, "Providing Data Services for Machine-Readable Information in an Academic Library: Some Levels of Service," *Public-Access Computer Systems Review* 2, no. 1 (1991): 144–60.

8. James A. Jacobs and Charles Humphrey, "Preserving Research Data," *Communications of the ACM* 47, no. 9 (Sept. 2004): 27–29.

9. Rebecca C. Reznik-Zellen, Jessica Adamick, and Stephen McGinty, "Tiers of Research Data Support Services," *Journal of eScience Librarianship* 1, no. 1 (2012): 27–35.

10. Gold, "Cyberinfrastructure, Data, and Libraries, Part 1: A Cyberinfrastructure Primer for Librarians"; Gold, "Cyberinfrastructure, Data, and Libraries, Part 2: Libraries and the Data Challenge: Roles and Actions for Libraries."

11. R. David Lankes, Derrick Cogburn, Megan Oakleaf, and Jeffrey Stanton, "Cyberinfrastructure Facilitators: New Approaches to Information Professionals for E-Research," paper submission to Oxford e-Research Conference, University of Oxford, U.K. (Sept. 2008), available online at <http://quartz.syr.edu/rdlankes/Publications/Proceedings/Oxford-Lankes.pdf> [accessed 12 January 2013].

12. Andrew Creamer, Myrna Morales, Javier Crespo, Donna Kafel, and Elaine Russo Martin, "An Assessment of Needed Competencies to Promote the Data Curation and Management Librarianship of Health Science and Science and Technology Librarians in New England," *Journal of eScience Librarianship* 1, no. 1 (2012): 18–26.

13. Elsa Alvaro, Heather Brooks, Monica Ham, Stephanie Poegel, and Sarah Rosencrans, "E-Science Librarianship: Field Undefined," *Issues in Science and Technology Librarianship*, no. 66 (Summer 2011).

14. Jeffrey M. Stanton, Youngseek Kim, Megan Oakleaf, R. David Lankes, Paul Cogburn, Derrick Cogburn, and Elizabeth D. Liddy, "Education for eScience Professionals: Job Analysis, Curriculum Guidance, and Program Considerations," *Journal of Education for Library and Information Science* 52, no. 2 (Spring 2011), 79.

15. Gold, "Cyberinfrastructure, Data, and Libraries, Part 1: A Cyberinfrastructure Primer for Librarians"; Gold, "Cyberinfrastructure, Data, and Libraries, Part 2: Libraries and the Data Challenge: Roles and Actions for Libraries."

16. Graham Pryor, "Skilling Up to Do Data: Whose Role, Whose Responsibility, Whose Career?" *International Journal of Digital Curation* 4, no. 2 (2009): 158–70.

17. Youngok Choi and Edie Rasmussen, "What Qualifications and Skills Are Important for Digital Librarian Positions in Academic Libraries? A Job Advertisement Analysis," *Journal of Academic Librarianship* 35, no. 5 (Sept. 2009): 457–67.

18. Sylvia D. Hall-Ellis, "Descriptive Impressions of Entry-Level Cataloger Positions as Reflected in American Libraries, AutoCAT, and the Colorado State Library Jobline, 2000–2003," *Cataloging & Classification Quarterly* 40, no. 2 (2005): 33–72; Sylvia D. Hall-Ellis, "Cataloging Electronic Resources and Metadata: Employers' Expectations as Reflected in American Libraries and AutoCAT, 2000–2005," *Journal of Education for Library and Information Science* 47, no. 1 (Winter 2006): 38–51; Sylvia D. Hall-Ellis, "Descriptive Impressions of Managerial and Supervisory Cataloger Positions as Reflected in American Libraries, AutoCAT, and the Colorado State Library Jobline, 2000–2004: A Content Analysis of Education, Competencies, and Experience," *Cataloging*

- & *Classification Quarterly* 42, no. 1 (2006): 55–92; Jung-ran Park, Caimei Lu, and Linda Marion, "Cataloging Professionals in the Digital Environment: A Content Analysis of Job Descriptions," *Journal of the American Society for Information Science and Technology* 60, no. 4 (Apr. 2009): 844–57.
19. Stanton, et al., "Education for eScience Professionals: Job Analysis," 79–94.
20. Sidney A. Fine and Steven F. Cronshaw, *Functional Job Analysis: A Foundation for Human Resources Management* (Mahwah, N.J.: Lawrence Erlbaum Associates, 1999).
21. Youngseek Kim, Benjamin K. Addom, and Jeffrey M. Stanton, "Education for eScience Professionals: Integrating Data Curation and Cyberinfrastructure," *International Journal of Digital Curation* 1, no. 6 (2011): 125–38.
22. S. Naranan and V.K. Balasubrahmanyam, "Models for Power Law Relations in Linguistics and Information Science," *Journal of Quantitative Linguistics* 5, no. 1–2 (1998): 35–61.
23. Sam Murugesan, "Understanding Web 2.0," *IT Professional* 9, no. 4 (July 2007): 34–41.
24. Frank van Ham, Martin Wattenberg, and Fernanda B. Viégas, "Mapping Text with Phrase Nets," *IEEE Transactions on Visualization and Computer Graphics* 15, no. 6 (Nov./Dec. 2009): 1169–76; James Sinclair and Michael Cardew-Hall, "The Folksonomy Tag Cloud: When Is It Useful?" *Journal of Information Science* 34, no. 1 (2008): 15–29.
25. Alvaro et al., "E-Science Librarianship: Field Undefined."
26. National Institutes of Health, *NOT-OD-03-032: Final NIH Statement on Sharing Research Data* (2003); Patricia Hswe and Ann Holt, *Guide for Research Libraries: The NSF Data Sharing Policy* (Chicago: Association of Research Libraries, 2010), available online at www.arl.org/rtl/eresearch/escien/nsf/index.shtml [accessed 12 January 2013].
27. Staša Milojević, Cassidy Sugimoto, Erjia Yan, and Ying Ding, "The Cognitive Structure of Library and Information Science: Analysis of Article Title Words," *Journal of the American Society for Information Science and Technology* 62, no. 10 (July 2011): 1933–53.
28. ICPSR, *Guide to Social Science Data Preparation and Archiving: Introduction*, available online at www.icpsr.umich.edu/icpsrweb/content/deposit/guide/index.html [accessed 12 January 2013]; DCC Curation Lifecycle Model, available online at www.dcc.ac.uk/resources/curation-lifecycle-model [accessed 12 January 2013]; Lynda M. Kellam and Peter Katharin, *Numeric Data Services and Sources for the General Reference Librarian* (Oxford: Chandos Publishing, 2011).
29. Data Sharing for Demographic Research (DSDR), *Guide to Social Science Data Preparation and Archiving—Best Practice Throughout the Data Life Cycle*, 3rd ed. (Ann Arbor, Mich.: ICPSR, University of Michigan, 2005), vii.
30. Soehner, Steeves, and Ward, *Science and Data Support Services*.
31. *Ibid.*, 1.
32. Stanton et al., "Education for eScience Professionals," 91.
33. See note 2.
34. Jingfeng Xia, "Diffusionism and Open Access," *Journal of Documentation* 68, no. 1 (Jan. 2012): 72–99.
35. Abdus Sattar Chaudhry and N.C. Komathi, "Requirements for Cataloging Positions in the Electronic Environment," *Technical Services Quarterly* 19, no. 1 (2002): 1–23; Hanna Kwasiak, "Qualifications for a Serials Librarian in an Electronic Environment," *Serials Review* 28, no. 1 (Spring 2002): 33–37; Jung-ran Park, Caimei Lu, and Linda Marion, "Cataloging Professionals in the Digital Environment: Draft Proposed ALA Core Competencies Compared to ALA-Accredited, Candidate, and Precandidate Program Curricula: A Preliminary Analysis," *Journal of Education for Library and Information Science* 47, no. 1 (Winter 2006): 52–77.
36. See "Mission of the School" on the Website of Indiana University's School of Library and Information Science, available online at <http://slis.iu.edu/welcome.html> [accessed 12 January 2013].
37. Sidney A. Fine and Steven F. Cronshaw, *Functional Job Analysis*.
38. Owen Kaser and Daniel Lemire, "TagCloud Drawing: Algorithms for Cloud Visualization," *Proceedings of the WWW2007 Meeting*, Banff, Canada (May 2007).
39. NIH, Office of Extramural Research, *NIH Data Sharing Policy*, available online at http://grants.nih.gov/grants/policy/data_sharing [accessed 12 January 2013]; NIH, *NOT-OD-03-032: Final NIH Statement on Sharing Research Data* (2003); Hswe and Holt, *Guide for Research Libraries*.
40. Peter Suber, "An Open Access Mandate for the NIH," *SPARC Open Access Newsletter* 117 (Jan. 2008), available online at http://dash.harvard.edu/bitstream/handle/1/4322583/suber_nih-mandate.html?sequence=1 [accessed 12 January 2013].
41. Clifford A. Lynch and Joan K. Lippincott, "Institutional Repository Deployment in the United States as of Early 2005," *D-Lib Magazine* 11, no. 9 (Sept. 2005).
42. Data Management for NSF SBE Directorate Proposals and Awards, available online at www.nsf.gov/sbe/SBE_DataMgmtPlanPolicy.pdf [accessed 12 January 2013].
43. NIH, Office of Extramural Research, *NIH Data Sharing Policy*; NIH, *NOT-OD-03-032: Final NIH Statement on Sharing Research Data* (2003); Hswe and Holt, *Guide for Research Libraries*.

44. Tracy Gabridge, "The Last Mile: Liaison Roles in Curating Science and Engineering Research Data," *Research Library Issues: A Bimonthly Report from ARL, CNI, and SPARC*, no. 265 (Aug. 2009): 15–21, available online at <http://www.arl.org/bm-doc/rli-265-gabridge.pdf> [accessed 12 January 2013].

45. Soehner, Steeves, and Ward, *Science and Data Support Services*.

46. A. Swan and S. Brown, *Skills, Role & Career Structure of Data Scientists & Curators: Assessment of Current Practice & Future Needs* (London: JISC Report, 2008), available online at <http://eprints.soton.ac.uk/266675/> [accessed 12 January 2013].

47. ICPSR, "Guide to Social Science Data Preparation and Archiving: Introduction," Figure 1—The Data Life Cycle, available online at www.icpsr.umich.edu/icpsrweb/content/ICPSR/access/deposit/guide/#cycle [accessed 12 January 2013].