

Data Jobs: A Place for Science and Social Science PhDs in the Libraries?

Jodi Reeves Eyre

Introduction

This white paper is a snapshot of trends in library jobs related to data management, curation, and services (*data jobs*) current in the first four years of the program. The paper focuses on how responsibilities and qualifications listed in data job postings from 2013–2017 relate to qualifications and experience held by recent PhDs in the Sciences and Social Sciences.

For this study, Jodi Reeves Eyre, then of Eyre & Israel, LLC, and now program officer for CLIR, conducted an analysis of more than 161 job postings from 2013–2017 to identify trends in desired qualifications, requirements, and the numbers of data jobs advertised, with focus on the suitability of PhDs in the Sciences and Social Sciences to fill these positions as advertised.

The analysis shows the following findings:

- Libraries seek candidates who can provide services related to data management, curation, instruction, and visualization, and related scholarly communication activities¹.
- Ideally, candidates have a knowledge of metadata, qualitative and/or quantitative research experience, and some library service experience.
- Hard barriers that keep recent PhDs from these positions include requirements for a library degree, library experience, or both.
- Soft barriers include the emphasis on service/consultation and lack of emphasis on active research opportunities in job postings.
- There were no appreciable increases or decreases in the total number of data-related position postings analyzed between 2013 and 2017 when compared to trends in the number of overall postings.

Further research should be done into whether recent PhDs are interested in entering service/consultation fields and whether recent graduates have the skills required for this type of work. CLIR's early experiences with the Fellowships in Data Curation for the Sciences and Social Sciences suggest that data service and consultation work can be

¹ These terms mean different things to different people depending on their professional or disciplinary contexts. They may also refer to similar sets of activities. This is illustrated by the common co-occurrence of terms, as discussed in the Results section of this paper. CLIR's *The Open Data Imperative* has useful definitions of data curation, digital curation, and data management (Allard et al. 2016, 9).

attractive to some recent PhDs, and that these candidates bring and can quickly develop skills that make them valuable library colleagues.

Methodology

Data Collection

The majority of data for 2013 through most of 2016 came from the [ARL Position Description Bank](#), a service that the Association of Research Libraries (ARL) provided for its members. ARL staff shared data on thousands of jobs with CLIR. The ARL Position Description Bank was curated by ARL staff with human resources officers at ARL member institutions submitting the job listings for distribution within the ARL community. CLIR received the data from ARL on October 20, 2016.² The data set included 2,773 job postings. Reeves Eyre identified duplicates and unposted listings during the initial review of the ARL data but determined that these listings did not significantly affect the analysis needed for this study. These duplicates and unposted listings were left in the original ARL data set for quantitative analysis.

Additional job postings were collected for October 2016 to December 2017 from the Digital Library Federation (DLF) Job Board, a job posting service for DLF member institutions.³⁴ The final DLF data set included 275 job postings with no duplicates. Combined with the ARL data, the entire sample included a total of 3,048 job postings. Not all postings contained complete job descriptions. Some described the positions, but did not list all requirements or compensation, while others included basic information and a now defunct link to the complete posting.

Creating the Data Jobs Data Set

Job postings that contained the words *data*, *digital curation*, or *curator* in the position title were selected for quantitative and qualitative analysis. Some job postings advertised two positions in one posting. These postings were split in two with each position analyzed as if it were a separate posting. This resulted in a subset of 130 data job postings.

For comparison, 31 additional job postings that did not contain the words *data*, *digital curation*, or *digital curator* in the position title were included in quantitative and qualitative analysis. For this subset, one posting meeting the criteria out of every 100 consecutive postings for each year was arbitrarily chosen. This guaranteed at least one

² Thank you to Sue Baughman and Gary Roebuck of ARL for their help in making the ARL Position Description Bank data accessible.

³ The DLF is a program of CLIR.

⁴ Thank you to James DeVos, Arizona State University, for his help in collecting these job postings.

selection from each year under study and a selection representing approximately one percent of the original 3,048 jobs postings.

The resulting subset of 161 job postings was analyzed to identify general trends regarding job responsibilities and required qualifications using the quantitative and qualitative methodology outlined below.

Analysis Methodology

A mixed methods approach was used to determine any commonalities among the requirements and preferences employers sought in the data-related job postings, in comparison with the sampling of other jobs from the same time period, as well as whether these requirements and preferences shifted over the four-year time span covered by the postings. Quantitative data collected included the date submitted (ARL postings) or posted (DLF postings), degree required or preferred (if any mentioned), tenure-track status (if mentioned), years of experience required or preferred (if mentioned), and whether the position was a liaison role for one or more academic departments (if mentioned). Job descriptions were analyzed using the qualitative analysis software package, Dedoose. The descriptions were coded, a process in which excerpts from the position descriptions were selected and assigned a set of tags. These tags, or codes, were then used to identify trends in the postings.

Coding was based initially on the terms used in the postings themselves (data management, service, and teaching are some examples). As coding proceeded, however, some terms were grouped under more general codes, such as *specific tools* as a code for any reference to a specific software package, digital tool, or programming language, etc. that was mentioned. Teaching, workshops, and training were grouped together under *instruction*.

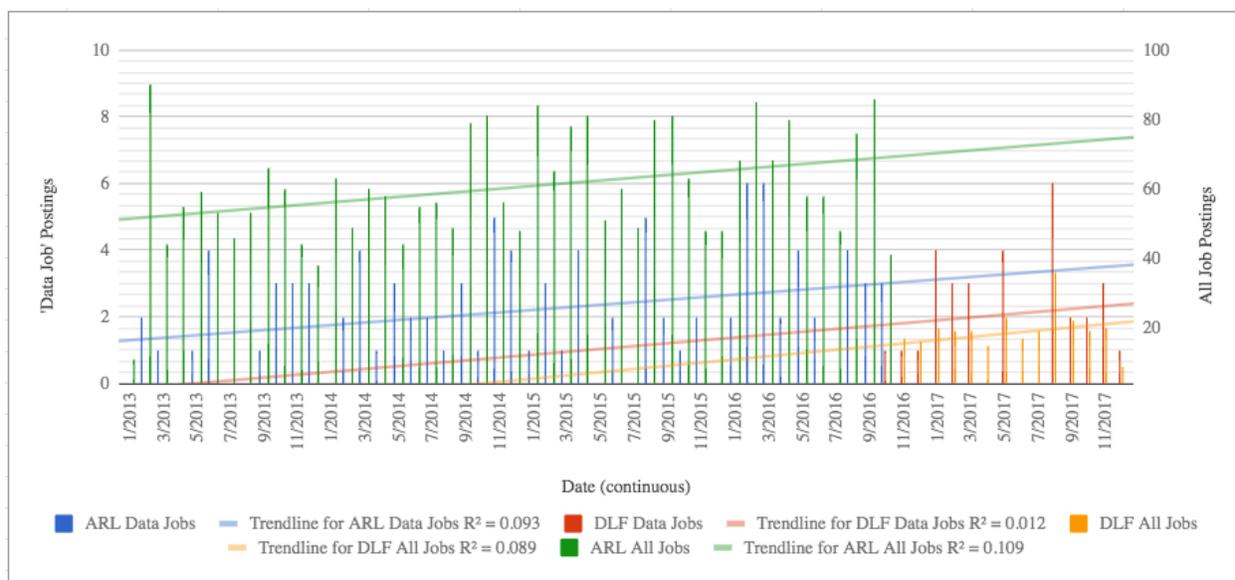
Results

Quantitative Trends in the Numbers of Data Jobs Advertised

The frequency of data job postings matching the selection criteria was compared to the overall frequency of job postings between 2013 and 2017. There was no appreciable increase or decrease in the number of data postings analyzed over this period when compared to trends in the number of overall postings (figure 1).

The period of time covered by the data could account for the lack of significant trends. In 2013, the US government issued a mandate requiring federal agencies with annual research and development expenditures of more than \$100 million to create plans for

Fig. 1. Number of jobs and data jobs submitted to ARL or posted to DLF. Data jobs follow the scale along the right vertical axis. All jobs follow the scale along the left vertical axis.



public access to agency data (Holdren 2013). Including jobs listed prior to the mandate in the analysis may have yielded an increase in the frequency of postings for data jobs starting in 2013. Limiting analysis to postings with data or digital curation in the job title may also have affected the ability to detect time-based trends. For example, analyzing a larger group of postings including all jobs mentioning any responsibilities related to data may have yielded different results. Table 1 shows seven postings from the 32 non-data job comparison sample that do mention data-oriented responsibilities or qualifications (table 1).

Table 1. Non-data job postings that include data-oriented responsibilities or qualifications

Job Posting	Year Posted/Created	Data-oriented Keywords
Digital Archivist	2013	Data Management, Metadata
Metadata Librarian	2014	Metadata
Digital Services Specialist	2014	Metadata
Social Sciences and Documents Librarian	2015	Data Management, Data Curation, Other Data Services/Support
Authorities/Identities Metadata and Cataloging Librarian	2016	Metadata
Black Studies Librarian	2016	Data Services, Data-intensive Research Support
Digital Scholarship Repository Specialist	2017	Metadata, Data Curation, Linked Data

Responsibilities

Serving Research Needs: The analysis showed that data jobs posted during this period were service- or consultancy-oriented as opposed to partnership-oriented. This was determined through qualitative and quantitative analysis of all the data job postings and by analyzing a subset of the postings. The subset of 13 data job postings selected for in-depth analysis represented approximately 10% of the total data job postings analyzed. For this subset, one out of every 10 consecutive postings was selected for analysis, guaranteeing at least one selection from each year under study.

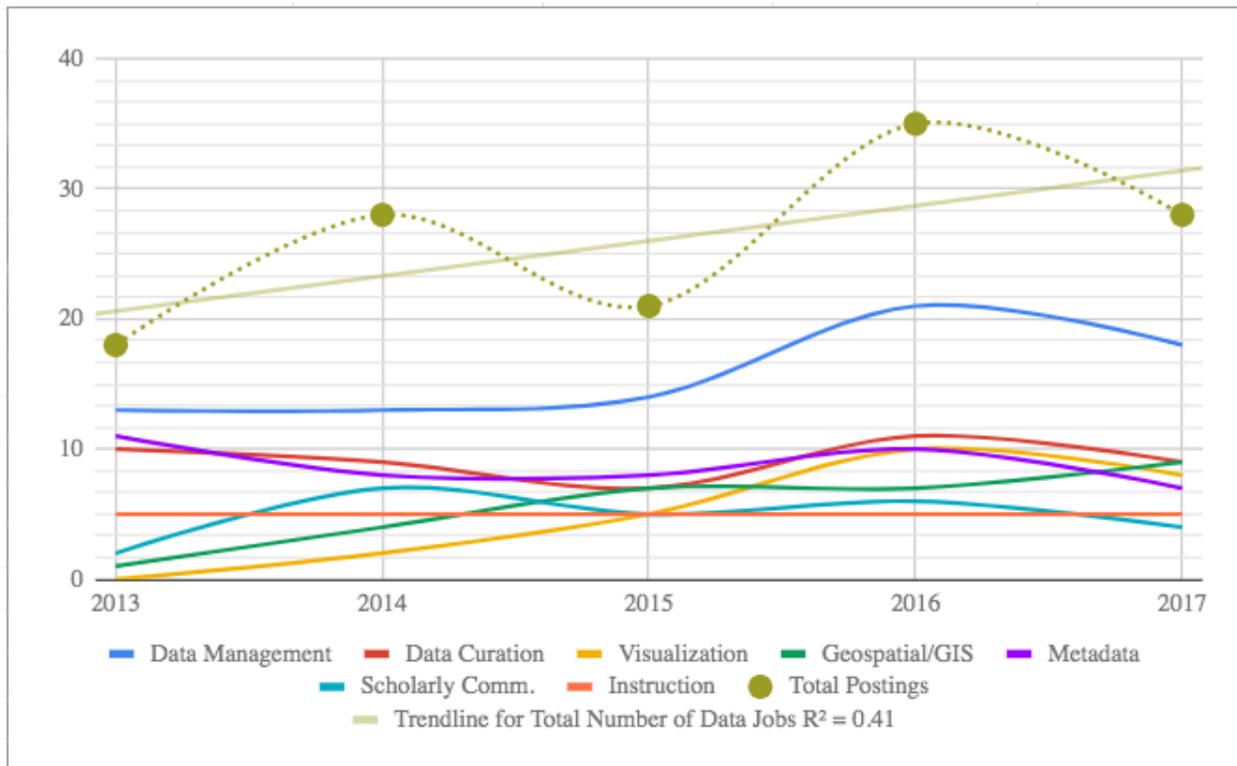
Five of the 13 postings selected for in-depth analysis were for candidates that would be based in the library and expected to work within internal library research support groups or centers; five were for liaison roles where candidates would be assigned to work with specific academic departments; and two were for candidates who would collaborate with or serve external research centers. None of the postings were for positions embedded or engaged with a specific faculty or departmental research project. This was a departure from the model for CLIR postdoctoral fellowship positions; hosts for these positions are encouraged to arrange dual appointments between the library and an academic department. Among the 13 positions selected for detailed analysis, there was only one dual appointment identified: a physics and research data librarian position at the University of Toronto that would be shared by the Department of Physics and the Science Libraries.

While data job candidates were not typically expected to lead or collaborate on specific research projects, the larger data set shows that prior research experience (18 of 130) and grant writing experience (9) were occasionally required or desired. Candidates with the skill set and experience that would enable them to lead or collaborate on a research project outside of a service role may have been attractive for those positions, even in cases where these leadership skills were not explicitly mentioned. Additional analysis of responsibilities and required or requested experience showed that strong research, data management, analysis, and visualization skills were valued in potential candidates for the 130 data job postings. These skills are notably similar to those valued in among research collaborators working outside library organizations.

Data Responsibilities and Experience: Most of the data-oriented positions analyzed for this study focused on providing data services and consultation. Data management services were the most dominant in the sample (mentioned in 79 of the 130 data jobs), followed by data curation support and services (44). One third of the postings also emphasized responsibilities and experience related to metadata (46). Instruction (34), data visualization (25), scholarly communication services (24), and geospatial services

(28) were also in high demand, relative to other types of responsibilities and experience mentioned. Following these areas of emphasis, project management was the next most frequently mentioned responsibility, but it appeared in only 16 job postings. There was no discernable increase or decrease over time in terms of the types of data-related responsibility and experience included in the postings (figure 2).

Fig. 2. Number of data job postings that mention major data responsibilities and experience over time



Most postings required a range of experience and responsibilities related to data services. Each of the seven most frequently mentioned areas of responsibility and experience co-occurred with the others, with the highest instances of co-occurrence for data management and data curation, data management and metadata, and data curation and metadata (table 2). This may indicate that these three areas of expertise are seen as integral to one another; for example, a professional with solid data management and metadata expertise would have good qualifications for providing data curation services.

Table 2. Number of data job postings where co-occurrence of primary data responsibilities and experience is present

	Data Mgmt.	Scholarly Comm.	Geospatial	Instruction	Visualization	Data Curation	Metadata
Data Mgmt.		20	9	24	15	37	34
Scholarly Comm.			10	10	6	11	6
Geospatial				10	15	4	11
Instruction					10	13	14
Visualization						2	6
Data Curation							26
Metadata							

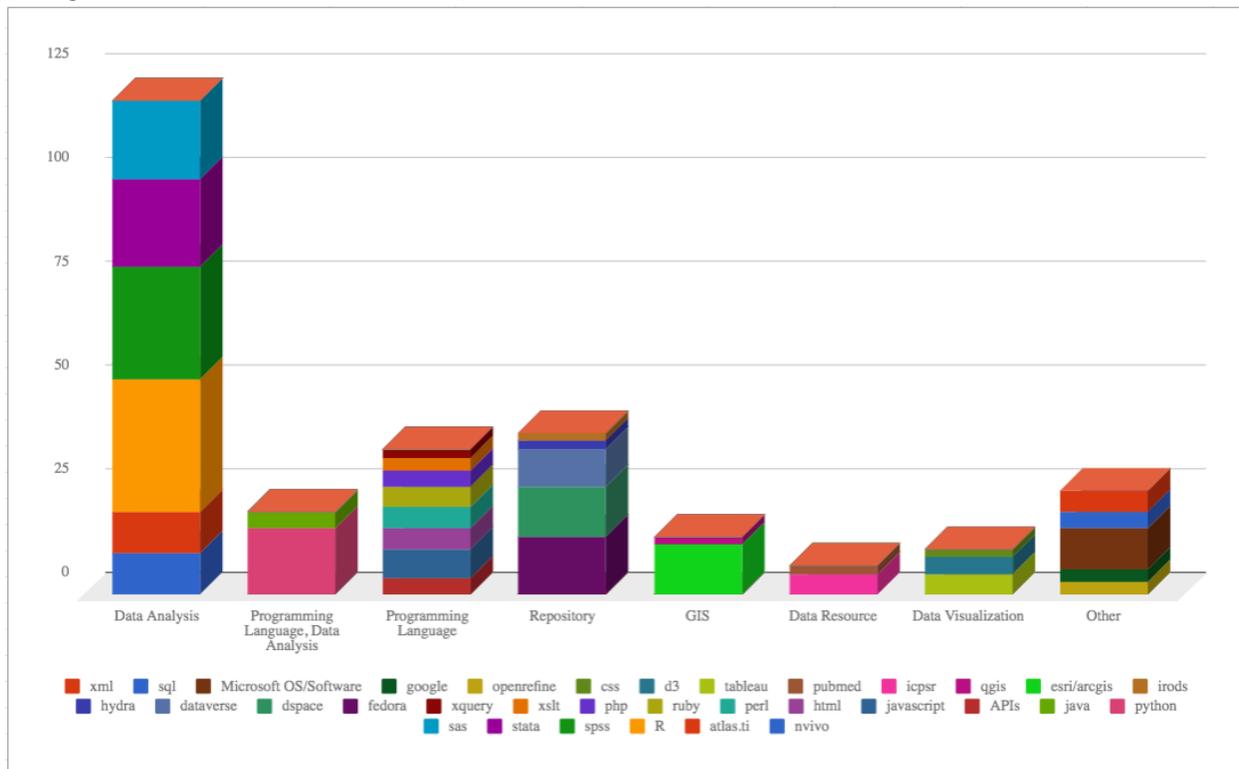
The required level of familiarity and experience with these seven areas varied among the postings. A posting for a digital data repository specialist required a candidate with an “[u]nderstanding of the research process, current issues in scholarly communication, and the role of libraries and librarians in research data curation.” A research data management librarian needed to demonstrate “active engagement in the scholarly discussions and experimentation taking place to address questions about long-term research data curation and preservation for reuse.” Neither asked for a specific number of years’ experience with data curation, unlike the research data librarian posting that required “four to five years of data preservation, curation management including data sets, content description and representation, metadata standards, and relevant workflows.”

Qualifications

Knowledge and Skill Sets: To determine the range of skills and knowledge sets candidates were expected to possess for the data-related positions, the text of all 130 postings was compiled into a single corpus for text mining. *Terms* from Voyant Tools identified the raw frequency of the common terms that had been grouped under *specific tools* through the coding process in Dedoose. This list was limited to tools and software that appeared more than once. The tools and software on this list were manually

grouped into broader categories (figure 3). These categories overlapped in many places; for example, some tools for data analysis were also programming languages. The tools grouped under data analysis encompassed both qualitative and quantitative analysis.

Fig. 3. Frequency of references to tools and software, including tools and software mentioned more than twice. Tools and software have been grouped into broad categories.

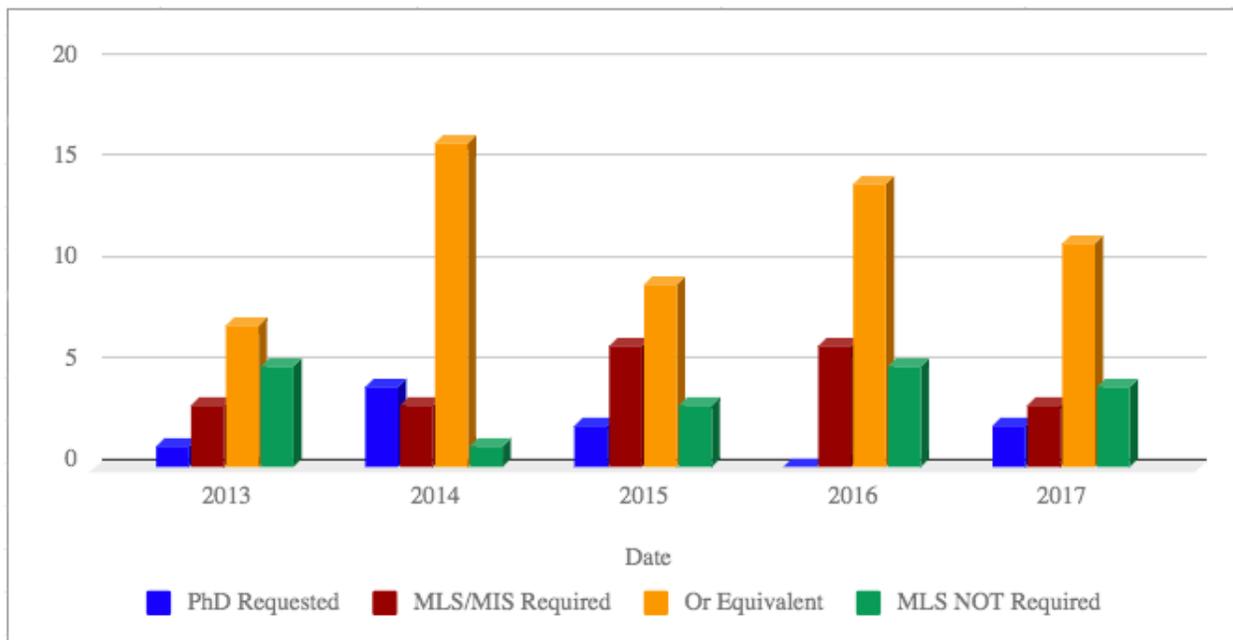


The majority of sought-after skills related directly to both quantitative and/or qualitative data analysis. This may indicate that data analysis experience is seen as an intrinsic part of providing data services, even if the primary focus of a position is on management and curation. For example, a candidate with data management experience and experience using R to analyze and visualize qualitative and quantitative data would likely be well qualified for a number of the library-based data jobs in the sample.

Degree Qualifications: Ninety-seven of the 130 data job postings listed some kind of required or desired degree. Most required or desired a library or information science master’s degree or equivalent degree (also see Figure 4). Specifically:

- Twenty-one required or desired a library or information science master’s degree, usually from an accredited program. This is a barrier for PhD holders aspiring to these positions.
- Fifty-seven required or desired a library or information science master’s degree or an equivalent degree—allowing for some leeway for considering candidates with other advanced degrees and relevant experience.
- Nine data postings requested candidates with PhDs or advanced degrees beyond an MLS/IS. One of the 31 non-data jobs analyzed requested a PhD. One data job mentioned a LIS PhD.⁵

Fig. 4. Number of postings per year that stated whether a library or information science degree was required or not, or whether an equivalent was required or desired, as well as number of postings that stated whether a PhD or terminal degree was required/desired.



There was a substantial increase in positions accepting *equivalent* degrees or experience as opposed to only MLS/MIS degrees from 2013 to 2014. Requests for PhDs also increased at this time. This coincides with the publication of the 2013 OSTP memo mandating public access to data (Holdren 2013), but a causal relationship cannot be established.

⁵ “Master’s or PhD in Library Science, Information Science, Library and Information Science, or Data Science; strong background in a social science/humanities discipline with a tradition of generating and using qualitative data . . . strongly preferred – OR – Master’s or PhD in a social science with a demonstrably strong background in data management, information science, and/or data science.”

Extent of Experience: Thirty-seven of the data jobs that listed a required degree also listed a minimum number of years of experience (table 3).

Table 3. *Number of job postings with a minimum number of years of experience required categorized by the number of years required*

<i>Min Years' Experience</i>	Number of Data Job Postings
1	12
2	7
3	12
4	6

Suitability for Recent PhDs in the Sciences and Social Sciences

If only research and subject knowledge, experience, and skill requirements are taken into account, PhD holders with the following kinds of experience might reasonably be assumed to be attractive candidates to these positions:

- accessing data from various sources, including repositories and databases common to a field/s
- managing larger amounts of data
- analyzing, visualizing, and making data available using the tools mentioned
- preparing data to a standard for publication or as part of funding responsibilities for internal or external reuse
- being aware of subject-specific descriptive and metadata practices and able to quickly build that skill set
- teaching or developing workshops, dependent on the program, subject, funding, or other factors

However, clear barriers exist for PhD-holding candidates for positions where the specific master's degree is required. Hard barriers included the requirement that candidates hold a master's in library or information science, or that they have library-based experience. Most of the postings accepted equivalent degrees or experience. One area for further study is whether libraries consider research or other experience as part of this requirement or if equivalent experience needs to be based on an academic library or data repository. PhDs interested in library positions need experience in a library or a service or consultation setting to be competitive. This kind of experience can be difficult to obtain for very recent science and social science graduates who are more likely to

have worked in collaborative settings, such as in laboratories, in the field, or in teaching, but not in a setting easily analogous to an academic library.

There is another, albeit soft, barrier to PhDs applying to and obtaining data job positions in libraries: the focus on service. Many positions required service-oriented individuals or individuals with the experience or disposition to work as a service provider to faculty. Several positions mentioned involvement with research centers, or the opportunity to conduct research as part of a tenure-track library position. This may lead recent PhDs to wonder whether data-related positions in libraries may limit the development of their own research interests or diminish their value as researchers. Recent PhD graduates may actually see many of their interests exemplified in these positions: the chance to use and grow current research skill sets, improve access to data, teach, and assist in the research endeavors of others. With some additional emphasis on the value of research and independent professional development and a broad vision for building data curation capacity through the infusion of diverse skills into the provision of data services, academic libraries may find they can attract strong candidates from a variety of disciplinary backgrounds to their teams.

References

Allard, Suzie, Christopher Lee, Nancy Y. McGovern, Alice Bishop. 2016. *The Open Data Imperative: How the Cultural Heritage Community Can Address the Federal Mandate*. Council on Library and Information Resources: Washington, DC. Available at <https://www.clir.org/pubs/reports/pub171/>.

ARL Staff. 2016. Descriptions from job postings from the ARL Position Descriptions Bank. 2016.obs_desc_10-20-16.csv. Data set created October 20, 2016.

ARL Staff. 2016. Metadata for job postings from the ARL Position Descriptions Bank (Jobs_meta_10-20-16.csv). Data set created October 20, 2016.

Atkins, Winston, Carol Kussmann, and Katherine Kim. 2017. "Staffing for Effective Digital Preservation 2017: An NDSA Report." *Open Science Framework*. October 20. doi:10.17605/OSF.IO/3RCQK.DLF Jobs. Available at <https://jobs.diglib.org>. Accessed October 15, 2017 and January 8, 2017.

Holdren, John P. 2013. *Memorandum for the Heads of Executive Departments and Agencies: Increasing Access to the Results of Federally Funded Scientific Research*. Available at https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf.

Sinclair, Stéfan and Geoffrey Rockwell. "Terms." *Voyant Tools*. Web. 20 December 2017. Available at <https://voyant-tools.org>.