# What is a Data Librarian?: A Content Analysis of Job Advertisements for Data Librarians in the United States Academic Libraries

**Hammad Rauf Khan**
Information Science, University of North Texas, Denton, U.S.A.
E-mail address: hammadkhan@my.unt.edu

**Yunfei Du**
Information Science, University of North Texas, Denton, U.S.A.
E-mail address: yunfei.du@unt.edu

**Abstract:**

*Society has begun to transition from the information age into the era of big data. The growth in the volume of data and statistics we have is larger than ever before. Researchers interest for big data and e-research has resulted in librarians taking on new job titles such as data librarian. Data librarian was once essentially a buzzword that has now become an actual job title and position in academic libraries, yet the American Library Association(ALA) has not developed a framework of core competencies for this emerging profession, neither has the Library and Information Science(LIS) curriculum begin to address a specialization for this rapidly growing and expanding field. This research attempts to understand what is required to be a data librarian from viewing job advertisements by hiring professionals in the United States academic libraries and applying a content analysis approach. This will allow us to understand the amplitude and meaning of what a data librarian is and how it is defined by the hiring managers through the description and qualifications they state and list in their job advertisements. By understanding what employers seek in data librarians, LIS curriculum developers can better implement coursework and pedagogy that will help address employer needs as well as developing successful candidates for these new job titles in academic libraries.*

**Keywords:** Data Librarian, Data Services, Big Data, Academic Library, LIS Curriculum

**Introduction**

One of the major changes affecting Library and Information Science(LIS) jobs is the phenomenon of Big Data. Katal, Wazid, and Goudar (2013) described big data as having the four Vs: Volume, Velocity, Variety, and Value. If we look at this definition of big data, then this concept is not new to libraries, as librarians have been dealing with big data since the ancient Library of Alexandria. Librarians in the past dealt with big data in the forms of scrolls and printed books, but librarians never termed these collections as "Big Data".

The emergence of "Big Data" in society has resulted in the increase in the number of datasets available to researchers at academic institutions. Government agencies, health organizations, and university libraries now routinely provide open access to many kinds of data. The rapidly growing information communication technology(ICT) has resulted in the emergence of e-science and e-research, which has created new job opportunities for academic librarians with the advent of the data librarian.

What is a data librarian? This is a question that has been asked before by Liscouski (1997) and his answer was that a data librarian "will provide a means of storing, retrieving, searching and recording access to laboratory data" (p. 199). This definition might hold true in the boundary of a laboratory, but is this how we define a professional title that is now becoming more common in academic libraries? The main issues underlying the need to define and answer this question as part of our research is that academic library directors have begun hiring for "data librarian" without there being any clear definition of what it means and what data librarians do in an academic library setting. It is even more important that LIS educators and curriculum developers know the core competencies of this profession to better prepare students in meeting employer demands. The LIS curriculum today is in need of bridging the gap to meet professional demands of big data skills in academic libraries. The American Library Association(ALA) offers specializations and competency statements for law librarianship, medical librarianship, school librarianship, youth librarianship, and music librarianship. Why not data librarianship?

This research explores data librarian job advertisements found through multiple online job boards in the United States academic libraries. The purpose of doing this exploration is to move towards discovering core competencies for data librarianship and finding emerging trends in skill requirements, job responsibilities, and qualifications needed to be a data librarian. The results of this preliminary study should be a starting point in aligning LIS curricula to meet the needs of prospective employers, along with highlighting the expanding boundaries of the LIS profession.

**Background**

"Academic librarians have a history of facing changes in technology that vastly reshape their work. The digital age has brought incredible changes in the way information and data are produced, consumed, adapted, and shared, requiring a transformation of resources and services" (Frank and Pharo, 2016, p.536). The traditional library was easy to define, it was the bricks and mortar structure with a clear and controlled entry point that contained and protected the selected physical resources over which the library asserted control and curatorial responsibility (Lagoze, 2014). The role and designation of librarian has been clearly defined and stated by ALA, but as society transitions from the information age into the era of big data, the change in academic libraries is noticeable with new job titles

emerging, such as data librarian, data services librarian, data curation librarian, and research data management librarian. Data librarian is not a new concept, but the current state of the data deluge and the increase in digital data requires more professional librarians to manage data at academic institutions, which has resulted in more positions being created for data librarians. Kim (2013) states that "Academic libraries should be challenged to develop a new professional strand of practice in the form of data librarianship" (p.502).

Semeler, Pinto and Rozados (2017) state that data librarianship has its origin in the social sciences and they go on to assess the concepts used in data science to serve as a basis for data librarianship. Semeler et al. (2017) comment that data librarians "focus on disseminating the important findings relating to their research in the form of relevant information by gathering data from various sources and organizing and cleaning them. The role of data librarians is to act as facilitators in all stages of scientific research, contributing with potential services that might be useful for data management and data curation" (p.2). Semeler et al. state that data librarians should have the skills that are taught in data science, however Kellam (2011) states that data librarians "come from a variety of disciplines and by diverse routes(p.151). While Kellam's observation is not in disagreement to Semeler et al. view on preparing data librarians through data science, it does bring up the question on how to prepare students to fill these new roles as data librarians? Kellam and Thompson (2016) have addressed the concerns of data librarianship in their book titled Databraianship: The Academic Data Librarian in Theory and Practice. In the book they have focused on how data is viewed and used across disciplines, its role in scholarly communication, and specific case studies of data support services in academic libraries. They have also included a proposal on teaching data librarianship to LIS students. Kellam and Thompson (2016) state that academic data librarianship "is not a single specialty but rather a varied collection of overlapping but distinct roles that center on providing access to, documenting, and preserving data, much as traditional librarianship has done for print sources" (p.5-6).

Federer (2018) looked at defining data librarianship through a survey of information professionals who spent a significant portion of their work providing data services. She created a survey in SurveyMonkey which gauged employment and educational background, skills and expertise, and data-related skills. The survey was distributed to the e-mail list for the Medical Library Association(MLA) and the Association of Academic Health Sciences Libraries. 90 total participants responded, and she concluded that data librarians are "a heterogeneous community of information professionals from varied educational and professional background, conducting many different types of work" (Federer, 2018, p.299). The main limitation of this study is that the respondents only belonged to the biomedical and health science field. Federer (2018) recognized that data librarianship jobs are becoming more common, but the field is still nascent.

Since the field of data librarianship has not been fully defined, this research intends to build upon previous work and adopt a content analysis approach to job advertisements for data librarian positions in academic libraries. Conducting a job analysis of library positions is common in the discipline of librarianship. The majority of the studies (Clyde, 2002; Deeken & Thomas, 2006) focus on analyzing trends over time or reviewing the job market in terms of specific positions or skills. Kim, Warga, & Moen (2013) conducted an analysis of job advertisements for the field of digital curation. Zhu (2008) sought to understand employers expectations for head of technical services positions by conducting a content analysis for the positions responsibilities. Zhu (2008) noted that job ads might not express all the employers'

expectations, but she chose to analyze job ads since they were the most public expressions of those expectations that were readily available for analysis.

**Research Design**

Online sources of job ads were preferred since it is in this venue that many employers advertise vacant positions. Job advertisements were gathered from ALA JobLIST, Indeed, Glassdoor, and the International Association for Social Science Information Services & Technology(IASSIST) job site. A coding scheme was created to organize the different components present in the job advertisements. A total of 50 random job advertisements for data librarians were gathered through NCapture and uploaded into NVivo software for coding. NCapture is a web browser extension, developed by QSR, to capture online material for import into NVivo. NVivo is a qualitative data analysis software application produced by QSR International for performing qualitative analysis on text-based and multimedia information. As the researchers collected the data, the data was organized into different nodes in the NVivo software and then analyzed.

**Assumptions and Limitations**

An assumption of this study is that the content analysis of job ads is a valid representation of the workforce demands of library directors. Content analysis is sometimes criticized as only interpreting the frequency of selected words and phrases, however it a technique which content can be examined in a systematic way. The researchers acknowledge that some ads are likely to be more accurate in their specification than others. No study could hope to identify all available jobs at any point in time as jobs are advertised in a wide range of sources and some are filled without public advertisement.

For this research, we are interested in understanding what is a data librarian through review of recent job advertisements in United States academic libraries. The researchers gathered job titles that have a variation of the words "Data Librarian" in them. In order to be included in our preliminary sample, the library must be an academic library located in the United States and should have the minimum qualification to include an ALA-Accredited Masters degree in Library and Information Science. The position must clearly be a professional position rather than a para-professional position. Part-time positions and appointments were excluded from this study.

If we are to move towards an endeavor in identifying core competencies for data librarians, then we must understand what is meant by the term competence. Ewens (1977) states that the term competence is used in the following ways in English and American dictionaries: adequate supply or sufficiency; a capacity to deal adequately with a subject; a quality or state of being functionally adequate or of having sufficient knowledge, judgment, skill or strength. The defining feature of competencies is that they are linked to positions, roles, or responsibilities (Gale & Pole, 1975). A competency is the skills, knowledge, and attitudes one should possess in order to perform adequately in a particular job. There are related concepts found in the literature such as job, function, task, subtask, and role. A job description is a task statement that defines the work tasks to be done by someone in a particular position (Stewart & Brown, 2014). Not all employers make distinctions between different types of competencies when listing them in their job postings. For this particular research we focus on required skills, jobs tasks(responsibilities), qualifications, and preferred skills. We sought to use content analysis as the method for this preliminary study to gauge the

level of information, skills, tasks, and qualification requirements for a data librarian position in academic libraries. Future research will build upon the results from this study and incorporate a job analysis methodology.

**Results**

For the job searcher the job title is the preliminary identifier as to whether a position may be suitable, hence some investigation of titles seemed a necessary step in this project. Figure 1. shows the different job titles which were included in our sample.

**Figure 1. Job Titles**
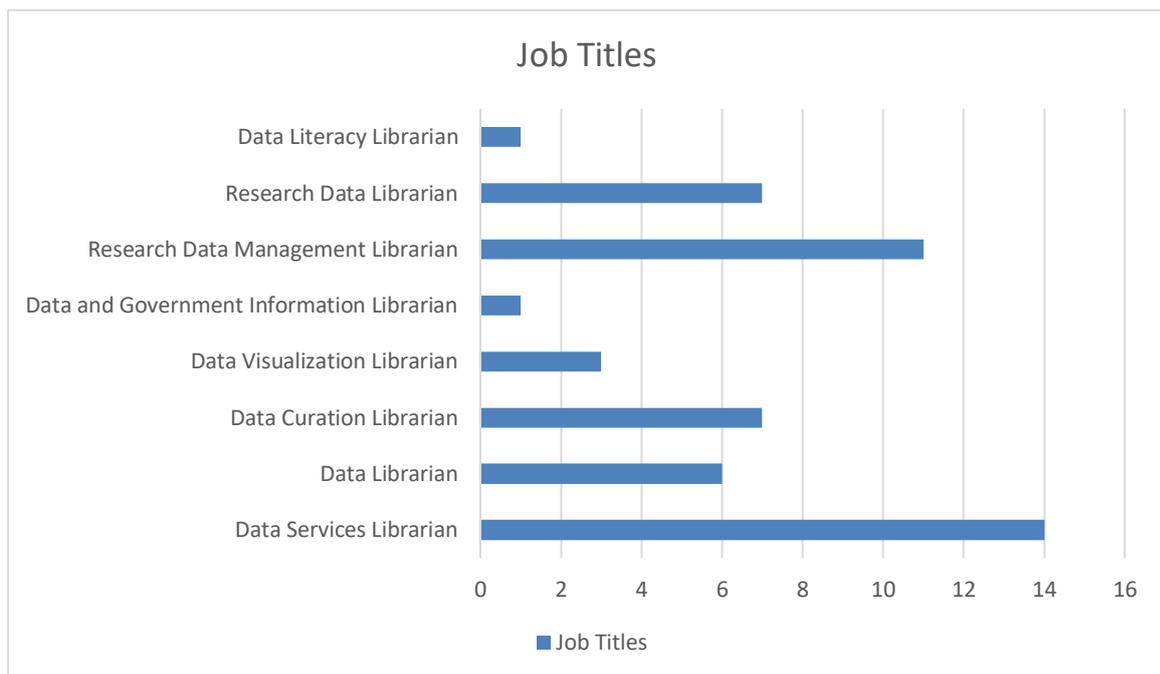


Data Services librarian job title was advertised the most with 14 included in this sample, followed by Research Data Management Librarian with 11. Data Literacy Librarian and Data and Government Information Librarian had only one job advertisement supporting these titles. These job advertisements all required an MLS from an ALA accredited program. All job titles had a variation of the keyword "Data Librarian" in the title and were used in the study to further understand the skills, qualifications, and responsibilities needed for data librarians in academic libraries.

Qualifications for the different job titles included ALA accredited MLS degree, knowledge of data cleaning, proficiency with statistical packages, familiarity with institutional and subject data repositories. Demonstrated experience in quantitative and qualitative research. One job advertisement mentioned "domain expertise", but it does not state what that is? Only three job ads out of the fifty in this sample stated years of experience required for the position. 2-3 years experience was the qualification requirement for these three job ads.

The researchers found many job advertisements listing required and preferred skills separately, so the researchers chose to separate these in the coding scheme. Table 1 represents the skills required to fill these positions. Table 2 represents preferred skills found in the job advertisements.

**Table 1. Skills Required**

| Skills Required | Percentage of Job Ads Reporting |
|---|---|
| Data information literacy competencies | 6% |
| Knowledge of U.S. federal and proprietary data sources | 82% |
| Research Assistance | 100% |
| Outreach | 61% |
| Teaching | 14% |
| Proprietary or Open Source statistical software packages (R, SPSS, Stata, SAS, Python, NVivo) | 70% |
| Visualization | 25% |
| Knowledge of standards-based metadata schema | 8% |
| Quantitative and Qualitative Data Analysis | 20% |
| Critical Thinking/Problem Solving | 89% |

Table 1 highlights research assistance as the most mentioned skill in all of the job advertisements, followed by knowledge of U.S. federal and proprietary data sources. Proprietary and Open Source statistical software was found as a skill requirement for 70% of the job ads. It is also interesting to note that these statistical software packages offer visualization tools and analysis, but some job advertisements did not mention these statistical software packages but asked for visualization and analysis as required skills. Only 20% made any mention of quantitative and qualitative data analysis skills. Only 6% asked for digital information literacy competency skills, while 14% asked for teaching skills. Outreach was listed under a required skill in 61% of the job advertisements, which comes in much higher than knowledge of standards-based metadata schema which only has 8% representation in the job advertisements.

**Table 2. Skills Preferred**

| Preferred Skills | Percentage of Job Ads Reporting |
|---|---|
| Geospatial Data & Technologies | 20% |
| Programming Languages | 60% |
| Statistical Packages (R, SPSS, Stata, SAS, Python) | 24% |
| Data Management & Curation | 30% |
| Grant Writing | 4% |
| Project Management | 14% |
| Critical Thinking/Problem Solving | 28% |
| Data packaging, data re-use, data encoding | 7% |
| Institutional repositories & metadata | 26% |
| Current technologies | 45% |

Skills preferred had 45% of employers listing current technologies as a preferred skill, without highlighting what are "current technologies"? Only 24% mentioned statistical packages as preferred skill, but we notice that majority (70%) of job advertisements have these as a required skill. Would the statistical packages be considered "current technologies"?

Data management & Curation was shown as a preferred skill of 30% versus the 14% for Project Management. Grant writing only appeared as a preferred skill in 4% of the job advertisements.

The job responsibilities and tasks were found in the job ads through word frequency count and are summarized here. 90% wanted instructional support, this includes instructing research faculty and graduate students with data management practices. 89% mentioned one-on-one consultations, this could also include research support consultations. 86% required preservation and access to research data, in which we include discover & use of data, management of locally created and externally available statistical data, and support for data sets. 61% would describe campus and funding agency data policies. 56 % will maintain data services web pages. 49% will create data guidelines & best practices. 32% will market data services offerings, this might be related to the required skill outreach. 11% will conduct data workshops. 3% will do publication & presentations.

**Discussion**

One job advertisement mentioned the following statement under responsibilities, the candidate must be "adventurous enough to expand or shift their range of responsibilities to meet evolving campus needs". Adventurous could be a skill set but is not found in the top word frequencies in the data collected through content analysis, but we wanted to highlight a unique statement found in a job advertisement for a research data management librarian position. Most of the job responsibilities are related to data, but we can see that marketing data service offerings, workshops, instructional activities, creating guidelines and policies, is no different from other types of jobs in academic libraries that call for developing and promoting library services. What is interesting however, and can be addressed in the future, is how these responsibilities fit into the overall organization of the academic library and how these data professionals work with subject, reference, and technical service librarians.

Future studies need to address the data librarian work environment. This might be done through a job analysis approach. Where do data librarians fit in the academic library ecosystem? What parts of the data curation and data lifecycle are responsibilities of a data librarian in an academic library? Distinguishing characteristics of metadata librarian, data services librarian, and geospatial data librarian are necessary in forming a more comprehensive understanding of their roles and overall place in the academic library ecosystem.

If we go back and revisit Liscouski's (1997) definition of data librarian, we can now see with the results found in this preliminary research that in an academic library, data librarian is an umbrella term with many different job titles, skills, and professional responsibilities associated with it. A data librarian as it is understood today represents Liscouski's definition, but much more. Data librarians in academic libraries provide guidance on data to their campus through collaborative, instructional, consultative, and technical support. In order for the library profession to move towards a common understanding and definition of data librarianship in academic libraries, more effort needs to be given in understanding the core competencies of data librarians.

The sample we selected were all job advertisements that required an ALA accredited MLS degree, but a few job ads under job qualifications and preferred skills stated that an advanced degree or certificate in data science should accompany the MLS degree. This preliminary

research provides LIS curriculum developers with an overview of data librarian jobs in academic libraries, which could help better prepare students who are interested in pursuing data librarian positions.

**Conclusion**

Academic libraries are undergoing a period of intense change with new roles and services being created to meet the big data challenge. Faced with increasing demands from funding agencies, faculty, students, and researchers, academic librarians are in need of creating and supporting data intensive research. The LIS curricula is in need of addressing employer demands for data librarians. This research, albeit with its limitations, is a starting point towards understanding what is needed in academic libraries in order to better prepare students for these new data librarian positions.

**References**

1. Clyde, L. A. (2002). An instructional role for librarians: an overview and content analysis of job advertisements. *Australian academic & research libraries*, *33*(3), 150-167.
2. Deeken, J., & Thomas, D. (2006). Technical services job ads: Changes since 1995. *College & research libraries*, *67*(2), 136-145.
3. Ewens, T. (1977). *Think Piece on CBE and Liberal Education*(No. 1). CUE Project, Bowling Green State University.
4. Federer, L. (2018). Defining data librarianship: a survey of competencies, skills, and training. *Journal of the Medical Library Association: JMLA*, *106*(3), 294.
5. Frank, E. P., & Pharo, N. (2016). Academic librarians in data information literacy instruction: a case study in meteorology.
6. Gale, L. E., & Pol, G. (1975). Competence: A definition and conceptual scheme. *Educational Technology*, *15*(6), 19-25.
7. Katal, A., Wazid, M., & Goudar, R. H. (2013). Big data: issues, challenges, tools and good practices. In Contemporary Computing (IC3), 2013 Sixth International Conference on (pp. 404-409). IEEE.
8. Kellam, L., & Peter, K. (2011). Numeric data services and sources for the general reference librarian. Elsevier.
9. Kellam, L. M., & Thompson, K. (2016). Databrarianship: The academic data librarian in theory and practice. Association of College and Research Libraries, a division of the American Library Association.
10. Kim, J. (2013). Data sharing and its implications for academic libraries. *New Library World*, *114*(11/12), 494-506.
11. Kim, J., Warga E., & Moen (2013). Competencies required for digital curation: An analysis of job advertisements.
12. Lagoze, C. (2014). Big Data, data integrity, and the fracturing of the control zone. *Big Data & Society*, *1*(2), 2053951714558281.

13. Liscouski, J. (1997). 2. The Data Librarian: introducing the Data Librarian. *Journal of Analytical Methods in Chemistry*, *19*(6), 199-204.

14. Semeler, A. R., Pinto, A. L., & Rozados, H. B. F. (2017). Data science in data librarianship: Core competencies of a data librarian. *Journal of Librarianship and Information Science*, 0961000617742465.

15. Stewart, G.L., & Brown, K.G. (2014). Human resource management: linking strategy to practice. Hoboken, NJ: Wiley.

16. Zhu, L. (2008). Head of cataloging positions in academic libraries: An analysis of job advertisements. Technical Services Quarterly, 25(4), 49-70.