# Who is Teaching Data: Meeting the Demand for Data Professionals

**Jeonghyun Kim**
*College of Information, University of North Texas. Email: Jeonghyun.Kim@unt.edu*

As data has become critical to our everyday lives, a growing concern with the skills gap required to exploit the data surfeit has arisen; library and information science practitioners and educators have recognized this concern. This paper is intended to identify current trends in library and information science education in response to the rising demand for data professionals. To provide a detailed map of the content of the current curriculum, academic programs and courses that support a data-driven workforce offered by library schools in North America were reviewed. The results of this analysis indicates that various topics are being offered to address skills gaps for data professionals, but there are still insufficient opportunities for students to develop the depth and breadth of knowledge and skills needed to be highly capable data professionals. It is suggested that cross-disciplinary and/or cross-institutional collaboration may be an efficient way to enhance and develop educational and training opportunities for data professionals.

**Keywords:** big data, data professionals, LIS education, curriculum analysis, academic libraries, research skills

## Introduction

We live in an era of big data. Big data is a catchphrase used to characterize massive and complex data sets largely generated from recent and unprecedented advancements in information technology and approach. The ever-increasing growth of such data sets has impacted every aspect of modern society, including industry, government agencies, health care, academic institutions, and research in almost every discipline. It has also prompted us to direct our attention to the question: *How to harness the power of big data*?

With the emergence of this phenomenon, there is a constant call for the ability to work with data. There is a need to discover, structure, manipulate, analyze, visualize, manage, and preserve data in order to harness its power for the greater good. Although the need for big data skills has grown exponentially, one key challenge is the limited availability of skilled workers. Gartner, a research consultancy firm providing information technology-related insight, projected a significant shortfall in the big data job market: "By 2015, 4.4 million IT jobs globally will be created to support big data with 1.9 million of those jobs in the United States. . . . However, while the jobs will be created, there is no assurance that there will be employees to fill those positions" (Pettey, 2012). The discussion regarding the increase in, and diversity of, big data management and analysis job opportunities is not limited to the United States. According to research conducted by e-skills UK, predictions for the United Kingdom point to a 160% increase in labor market demand for big data skills between 2013 and 2020. However, the research also indicates that there is already a shortage of analytical and managerial skills necessary to make the most of big data, with 77% of big data roles being already considered "hard to fill" (McNulty, 2014). In the library and information science profession, this prediction has become a reality. It has been suggested that


ISSN: 0748-5786  
doi:10.12783/issn.2328-2967/57/2/8

"data is an area that has a need for a larger workforce equipped with the specialized skills to manage data and support data analytics activities" (Allard, 2015).

It has become evident that librarians and information professionals must take a leading role in working with big data. Gordon-Murnane (2012) asserted that this is because LIS professionals already have the skills, knowledge, and services to help their communities capitalize on all that big data has to offer. A number of reports produced by professional associations, including the Association of College & Research Libraries Research Planning and Review Committee (2014) and Australian Library and Information Association (2014), anticipate that those working in libraries and information centers will find new roles in big data. In these jobs they will be helping collate, process, and make useful the enormous volume of data that is being generated in all areas of life.

In adopting these roles the LIS profession is being challenged to develop a new professional strand of practice to respond to the growing data needs of their communities. Although there is value in the skills librarians already possess and transfer, there is a need for a new set of skills for the next level of engagement and support for data management and exploitation. The current job market shows that there is a requirement to build capacity and capability for data expertise (Hedstrom, Larsen, & Palmer, 2014). In fact, considerable discussion has been devoted to the question of how libraries and LIS schools can retool to better reflect the requirements and challenges of today's data explosion (e.g., Blake, Stanton, Larson, & Lyon, 2012; Dumbuill, Liddy, Stanton, Mueller, & Farnham, 2013; Lyon, 2012; Lyon & Brenner, 2015). Most discussion has focused on specific fields, such as data management, curation, and preservation, but little has been revealed about the wide range of data management areas that are developing.

How is academia responding to this new professional strand of practice? How well are LIS schools preparing students to be data professionals? The research documented in this paper was conducted in response to the rising demand for data professionals and data expertise in the library workforce by surveying the data-related curriculum of American Library Association (ALA)-accredited library and information schools in North America. Academic programs and courses containing elements of the data profession and practice were reviewed.

## Background

The LIS profession is in a period of considerable change. As data has become a valuable information resource, data librarianship has become part of the profession. This has occurred notwithstanding that data librarianship is still an ill-defined area but one often used to refer to a special set of responsibilities around stewardship of data. While the term has a "new ring" to it, data libraries started back in the 1960s as support services assisting researchers in preserving and distributing machine-readable information when a number of universities and government-supported research institutions established specialized data centers (Martinez-Uribe & Macdonald, 2009). Examples of such data libraries include *Inter-university Consortium for Political and Social Research*, which was established in 1962, and *UK Data Archive*, which was founded in 1967. The *Internet Association for Social Science and Information Service and Technology* was created in 1974 to support a newly emerging profession of social science data archivists and librarians. These information specialists were developing data support services and establishing standards for managing and sharing computer-readable social science data (Adams, 2006).

Since the early 2000s, much discussion has been devoted to the long-term management and preservation of research data (e.g., Beagrie & Pothen, 2001; Lord &

Macdonald, 2003). This culminated with the launch of the UK's *Digital Curation Centre* in 2004. This initiative was intended to provide a national focus for research and development about curation issues, and to promote expertise and good practice for the management of digital research data. The academic library community in various countries, including United States, United Kingdom, and Australia, realized that opportunities to become involved in the curation and management of research data would become a new area of work. Areas of such involvement include, for instance, assisting researchers in designing and implementing data management plans for their projects and providing data repository services for data sets generated through the projects to make them accessible.

Further, data librarianship can be extended to include the concept of data science. Data science as a new profession and academic discipline sits at the intersection of social science, statistics, informatics, and computer science, and recently has been integrated into LIS as a prominent field of practice. As data science techniques and tools for extracting, manipulating, analyzing, and visualizing data are becoming increasingly important to all fields of scholarship, competency in employing such techniques and tools is needed for librarians and information professionals. As such, "there is a pressing need for interdisciplinary professionals who understand software, the Internet, data analytics, data visualization, and data curation. These professionals have their specialties; some are good at working with numbers, others are database experts, still others have expertise in unstructured data (e.g., text), but they also need generalist skills that let them bridge the wide range of tasks and methods needed to manage today's big data problems" (Stanton, 2012, p. 23).

Recently, some discussion has been devoted to the question regarding where librarianship can fit into this new field of data science. The workshop, "Filling the workforce gap in data science and data analytics," was held in *iConference 2013* (Blake, Stanton, & Saxenian, 2013), and in the same year, the *International Digital Curation Conference* hosted a symposium, "What is a data scientist?" (Jones, 2013). A number of academic libraries already have accepted the challenge of closing skills gaps to respond to the growing data needs of the community they serve. Examples include *Data Scientist Training for Librarians* (DST4L), an experimental course currently being offered by the Harvard-Smithsonian Center for Astrophysics John G. Wolbach Library and Harvard Library, and Columbia University's *Developing Librarian Project*, which recognizes the need for changes in the library profession to meet the needs of the digital scholarship at all stages. Since the late 2000s, there have been a number of educational initiatives funded by the Institute of Museum and Library Services to support educating LIS professionals to manage and curate research data. Examples include the University of Illinois at Urbana-Champaign's Data Curation Education Program (DCEP), University of North Carolina at Chapel Hill's Data Curation emphasis within the Post-Masters Certificate (PMC) program, and University of North Texas' Digital Curation and Data Management Certificate Program. In recent years, several iSchools, such as University of California at Berkeley and Syracuse University, have incorporated a data science and analytics component into their curriculum.

## Methodology

A total of 59 ALA-accredited Library master's programs in North America listed on the ALA website (www.ala.org/accreditedprograms/directory) in December 2015 were selected. Each institution's course offering documentation on their website, such as current course catalogue and course description database, were reviewed to identify data-related programs and courses.

An academic program was defined as any combination of courses and/or requirements leading to a degree, i.e., Bachelor's degree, Master's degree, and Ph.D. degree and certificate, or to a major, minor, or academic track, specialization, and/or concentration. Only those programs that list a set of recommended courses are included.[1] To identify the programs intended to prepare students for data profession careers, various search terms were used, including data curation, data science, data librarianship, data management, data analytics, and eScience. It should be noted that digital curation programs are included, although some programs focus on curation of digital objects and collections rather than data from scholarship, science, and education.[2] ( The programs identified were first classified based on their program, such as degree with concentration, graduate certificate, and advanced certificate. They were then classified by their academic level, i.e., graduate level, undergraduate level, and cross-level.

Courses were included if the course description indicated a data focus by using terminology such as data, research data, digital data, and big data. These courses were classified based on their academic level. Additionally, the courses were classified by whether prerequisites are required and whether the course is a regular or special topic course. To identify a taxonomy containing core topics for data-related curriculum, automated content analysis of course titles and descriptions was conducted. Course titles and descriptions were selected as they include descriptive keywords that represent the topics for the course content and provide an "at a glance" summary of the course by conveying the primary focus or purpose of the course. This automated content analysis technique, which assumes the application of the computational methods grounded in text mining to identify key topics and themes in a specific textual corpus, has been adopted in many bibliometric studies (e.g., Lee & Jeong, 2008; Cheng *et al.*, 2014). The analysis consists of two parts: (1) computer-assisted text analysis of course titles and course descriptions to generate a word list with frequency and collocations to characterize the texts; and (2) co-word analysis based on the co-occurrence of phrases to identify major concepts and themes in data-related course descriptions. Text pre-processing, including stop words filtering and lemmatization, was first performed. The most frequently occurring words and phrases in course titles and descriptions were then identified and tabulated using Provalis Research's WordStat text-mining software. Co-occurrence matrix on the phrases in the course descriptions was constructed; it was then exported for visualization in Gephi, a social network analysis tool by applying Force Atlas layout.

## Results

### *Academic Programs*

Out of a total of 59 ALA-accredited LIS schools, slightly more than one-quarter of the institutions (18) are offering academic programs preparing data professionals. Among those schools that provide data-related programs, more than three-quarters (13) are iSchools. Appendix I table summarizes various programs for data professionals and the institutions in which those programs are housed. Most programs are housed in the department that offers an ALA-accredited Master's degree in library and/or information science. Notable exceptions include University of Illinois's Master of Science in Bioinformatics and University of North Carolina at Chapel Hill's Graduate Certificate in Digital Humanities.

[1]Note that the Directory of Institutions Offering ALA-Accredited Master's Programs in Library and Information Studies lists each institution's areas of concentration or career pathway. However, such concentrations or career pathways do not always have a set of courses as defined by the institution.

[2]Digital curation has become a term and field that better accommodates a broader range of digital materials, which includes digital research data and other digital materials (Palmer, Weber, Muñoz, & Renear, 2013).

Table 1. Academic Programs by Program Type.

| Program | iSchools | Non-iSchools | Total |
|---|---|---|---|
| Bachelor's degree | 2 | 0 | **2** |
| Master's degree | 15 | 5 | **20** |
| Doctoral degree | 1 | 0 | **1** |
| Graduate certificate | 9 | 5 | **14** |
| **Total** | **27** | **10** | **37** |

As presented in Table 1, a total of 37 programs with data coursework were identified (see Appendix I for a full list of programs). Out of 37, approximately 70% of the programs (23) came from iSchools. It was found that 13 programs are being offered as a concentration, specialization, or career pathway in their degree program; many of those programs are often served as a guideline for students wishing to pursue specialized coursework rather than as a formal major or minor. It should be noted that two institutions, Drexel and Rutgers, are offering the program as part of their Bachelor's degree, and one institution, Indiana, is offering the program as part of its Ph.D. degree. Out of 37 programs, 14 programs are being offered as a certificate program, which is a series of courses providing in-depth study for those who want to excel in their chosen field or transition to a new career. Among those programs, only 4 programs are an advanced level for those who already hold their Master's degree.

The scope of programs varies among institutions as dictated by their focus, objectives, and course requirements. The subject areas of the program can be grouped into six areas:

1. Data curation promoting knowledge and skills in the management of scientific or research data generated in academic institutions, data centers, and libraries;
2. Digital curation encompassing the planning and management of digital assets and resources in museums, libraries, and archives;
3. Digital humanities emphasizing digital tools and techniques in high demand in humanities, such as digitization of cultural heritage materials, applied programming for analysis and visualization, and interface design and user experience;
4. Data science covering specific focus areas of statistical analysis, data mining, and data visualization;
5. Knowledge management, which is an extended format of a traditional knowledge management program by combining a field of business analytics; and
6. Informatics promoting an understanding toward the significant technical challenges created by large data environments.

Some exceptions are noted. Rutgers's Bachelor's degree in Information Technology and Informatics Major—Specialization in Data Science, Curation, and Management and Syracuse's Certificate of Advanced Study in Data Science are interdisciplinary in nature to provide an enriching training in science, statistics, research, and information technology by combining the areas of data curation and data analytics.

Typically, the programs list a few required courses but allow opportunity for elective course selections. Where elective selection was possible, it was guided through a list of approved courses, which are often but not limited to courses offered within the department.

## Courses

The total number of data-related courses identified in this study is 418. Of 51 institutions identified as offering those courses, 43 were in the United States and 8 were in Canada. Out of 418 courses, ap-

Table 2. Courses by Academic Levels.

| Level | iSchools | Non-iSchools | Total |
|---|---|---|---|
| Bachelor's | 52 | 15 | **67 (16%)** |
| Master's | 198 | 101 | **299 (72%)** |
| Doctoral | 20 | 5 | **25 (6%)** |
| Cross-level: Bachelor's/Master's | 5 | 5 | **10 (2%)** |
| Cross-level: Master's/Doctoral | 12 | 0 | **12 (3%)** |
| Cross-level: Bachelor's/Master's/Doctoral | 5 | 0 | **5 (1%)** |
| **Total** | **292** | **126** | **(100%)** |

proximately 70% of the courses (292) are being offered by iSchools; University of Illinois at Urbana-Champaign offers the highest number courses (33), followed by University of Pittsburgh (30), and University of Washington (26).

It should be noted that these courses are being taught at different levels. As shown in Table 2, more than three-quarters of the courses (326) are at the Master's level. The University of Illinois at Urbana-Champaign also offers the highest number of Master's-level courses (23), followed by University of Pittsburgh (25) and Indiana University (21). It is also important to note that more courses are at the Bachelor's level (92) than Doctoral level (37). Drexel University offers the highest number of undergraduate-level courses (10), followed by University of Washington (8) and University of Arizona (8).

Out of 418 courses, 83% of the courses (349) are regularly offered courses, while only 17% (69) is special topic courses, which cover topics in-depth in any of the department's regularly listed offerings. Forty percent of the courses (166) are upper-level courses that have prerequisites. Course prerequisites vary depending on the topic, from introductory core courses required for graduation to advanced technology-oriented courses.

To review course-specific details, two-word phrases used in the course titles and descriptions were identified and tabulated. Table 3 presents the top 25 core phrases that were used in the course titles and course descriptions with the number of cases, which represents the number of courses whose title or description includes the phrase. For instance, there are a total of 16 courses being offered simply using the title "database management."

Excluding some general descriptors for the intended audience, such as "information science" and "information professional," phrases used in the courses imply that data is being studied in various topic areas. Popular phrases, such as "data mining," "information visualization," "data analytics," and "data science," imply that topics for a broader field of data science[3] are prevalent across the courses. Other popular phrases, like "digital curation," "data curation," and "data management," indicate that management of data assets and data resources is certainly one core area where data is being taught. The phrases, including "data model," "data modelling," "database design," and "database management," present the topic of data administration, which deals with database implementations. Data also seems to be a core topic of study for methodology courses; this is supported by the phrases "data analysis," "data collection," and "research method." It should be noted that the term "big data" in the course title appeared with reference to various applied areas, such as "curation," "management," and "analytics"; this implies that acquiring and curat-

[3]Data science is often used as an overarching umbrella term for the field encompassing analytics, analysis, and mining of data.

ing big data as well as performing large-scale analytics are a core topic for big data.

Phrases highlighting skills for tools and techniques were often found in the course descriptions. This indicates that a majority of these courses are mainly engaged in practical application rather than theory-based learning; they include laboratory hands-on exercises and activities relevant to the topic designed to build conceptual knowledge and application. Large-scale datasets, real world problems/scenarios, and/or case studies are employed to support such exercises and activities.

To identify the inter-relationship of major themes adopted in data-related courses, the co-occurrences of phrases used in the course description were calculated and exported into Gephi for visualization. It should be noted that descriptors for intended audiences and instructional methods were excluded to only present topical themes. The map displayed in Figure 1 depicts the relationships among the phrases co-occurring in the course description. In this map, nodes (the circles in the image) represent the words or phrases, and edges (the lines connecting the nodes) represent the co-occurrence of two phrases; that is, if two phrases appeared in the same article abstract together, they were connected by an edge. It should be noted that the node size for each word/phrase is determined by its degree, which is the total number of other words/phrases with which it co-occurs. Additionally, concept communities (clusters) are distinctly presented in blue, yellow, red, green, and pink; these com-

Table 3. Frequently Occurring Phrases in Course Titles and Descriptions.

| Rank | Phrase in Title | Case | % | Phrase in Description | Case | % |
|---|---|---|---|---|---|---|
| 1 | Information Science | 19 | 4.55% | Data Analysis | 47 | 11.24% |
| 2 | Information System | 17 | 4.07% | Data Collection | 41 | 9.81% |
| 3 | Database Management | 16 | 3.83% | Information System | 37 | 8.85% |
| 4 | Research Method | 16 | 3.83% | Data Mining | 36 | 8.61% |
| 5 | Data Mining | 13 | 3.11% | Data Management | 30 | 7.18% |
| 6 | Information Visualization | 13 | 3.11% | Information Science | 27 | 6.46% |
| 7 | Big Data | 11 | 2.63% | Database Management | 23 | 5.50% |
| 8 | Data Analysis | 11 | 2.63% | Information Technology | 23 | 5.50% |
| 9 | Data Analytics | 10 | 2.39% | Big Data | 21 | 5.02% |
| 10 | Data Science | 10 | 2.39% | Data Structure | 21 | 5.02% |
| 11 | Digital Curation | 10 | 2.39% | Data Modeling | 21 | 5.02% |
| 12 | Information Professional | 10 | 2.39% | Real World | 20 | 4.78% |
| 13 | Data Management | 9 | 2.15% | Relational Database | 20 | 4.78% |
| 14 | Information Technology | 9 | 2.15% | Database Design | 17 | 4.07% |
| 15 | System Analysis | 9 | 2.12% | Information Retrieval | 17 | 4.07% |
| 16 | Data Curation | 8 | 1.91% | Information Professional | 16 | 3.83% |
| 17 | Database Design | 8 | 1.91% | Research Method | 15 | 3.59% |
| 18 | Information Management | 7 | 1.67% | Data Model | 15 | 3.59% |
| 19 | Information Study | 7 | 1.67% | Data Analytics | 14 | 3.35% |
| 20 | Management System | 7 | 1.67% | Data Visualization | 14 | 3.35% |
| 21 | Health Informatics | 6 | 1.44% | Large Scale | 14 | 3.35% |
| 22 | Geographic Information | 5 | 1.20% | Social Science | 13 | 3.11% |
| 23 | Health Informatics | 5 | 1.20% | Data Curation | 13 | 3.11% |
| 24 | Information Organization | 5 | 1.20% | Case Study | 13 | 3.11% |
| 25 | Information Retrieval | 5 | 1.20% | Information Visualization | 12 | 2.87% |

munities represent a group of courses on similar themes. Although a total of 12 communities were identified in this study, the following 6 communities are represented.

The largest community (red) is comprised of 17.31% of the total nodes and contains the key phrases "information system," "information retrieval," "data structure," and "data model." The community (light blue, 9.09%) adjacent to the red community includes the key phrases "data modeling," "relational database," "database management," and "data warehousing." These two communities represent courses on information systems, which typically consists of a database together with programs that capture, store, manipulate, and retrieve data. Some examples of courses include *Information System Design; Database Technologies; Database Management Systems*; and *Data Administration Concepts and Database Management*. Fundamental knowledge on data structure and algorithms is essential in designing and implementing information systems. Additionally, databases are an integral part of any information system; some fundamental concepts of databases covered in these courses include database modeling and design, relational databases, structured query language, database system architectures, and data warehousing techniques.

The second largest community (blue, 17.21%) is the cluster around "data management," "data curation," "open access," "research data," and "data archive." The courses in this community examine principles, practices, trends, and challenges in the curation and management of scientific research data. Most courses are intended to provide a foundation in data services, policy, and planning for information professionals in academic institutions involved with data-intensive research and scholarship. Specific topics for study include data selection and appraisal, data representation and organization, practices of data sharing and reuse, intellectual property issues, and institutional challenges in stewardship of research data.

The third community (green, 13.64%), which includes the phrases "data collection," "data analysis," "research question," "research design," and "data visualization," constitutes courses on research methods. These courses provide students with a comprehensive understanding of research methods with an emphasis on linking theory to practice. They examine connections among research questions, design, methods of data collection, and analysis. Further, they stress qualitative and quantitative data analysis skills using descriptive and inferential statistics. The titles of the courses include *Research Methods; Research, Assessment, and Design; Statistics and Data Analysis*; and *Research Data Analysis and Management*.

The community (yellow, 11.69%) adjacent to the green community is the cluster around "data mining," "big data," "machine learning," "data analytics," and "text mining." The study of prediction from data is the central topic of machine learning and statistics, and more generally, data mining. These courses emphasize various aspects of statistical data mining, including statistical data analysis as well as classic machine learning and data mining algorithms. Some of these courses introduce practical skills for applying data mining techniques using R as a primary analysis platform. The phrase "social network" occurred in the course descriptions as some courses focus on social media mining, with a particular emphasis on techniques for collecting and analyzing social media. These courses are titled *Data Mining with Machine Learning; Applications of Data Mining*; and *Exploratory Data Analysis*.

The last community (pink, 9.09%) encompasses the phrases "digital curation," "digital preservation," "digital object," "born digital," and "digital repository." These courses provide theoretical and practical perspectives on digital curation; they cover strategies, techniques, and standards related to preserving digitized and born-digital materials in archives, libraries, museums, and other cultural heritage

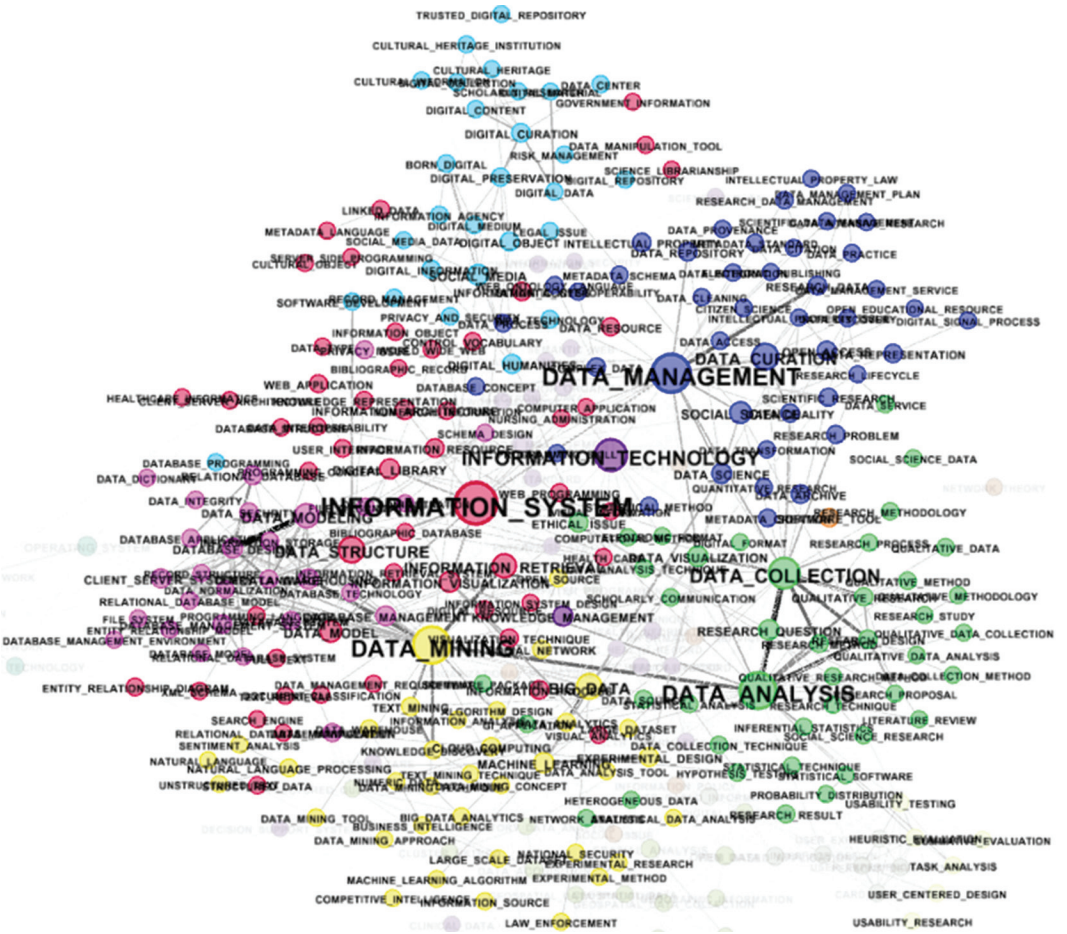


**Figure 1.** Visualization of the network of key phrases co-occurrences.

institutions. Several institutions are offering digital humanities elements within digital curation with a specialized pedagogic focus, including tools and techniques used by digital humanists, scholarly communication issues impacted in the field of digital humanities, and evaluation of digital humanities projects.

## Discussion

The results presented in the previous section provide some useful insights into the current state and future direction of LIS education.

First, a number of institutions are responding to the need for data skills in the marketplace by launching new academic programs aimed at boosting the number of qualified data professionals, but the content and focus of such programs varies widely. In the past decade, digital/data curation programs have been embedded in LIS education, yet the educational options appeared to be uneven, with limited opportunities for intensive preparation, as noted in the report recently published by the National Academies Press (2015). As such, a number of formal data science/analytics programs have begun to emerge as a new academic entity. The increasing development of interdisciplinary programs embracing the multidisciplinary nature of the subject studies within the larger units is also noted; such programs are not housed in a single department, which claims an advantage in being able to contract with experts from disparate disciplines.

Second, the number of courses varies considerably from institution to institution, as does the content of individual courses. New topics regarding big data, which have not been a major component of LIS education, have been incorporated. Additionally, a wide range of technologies, tools, and techniques needed to work with data has been presented in those courses. However, there is still an insufficient number of courses that support the depth and breadth of knowledge and skills needed to be a highly capable data professional. To fill this gap, some institutions recommend courses from other departments as electives for their programs.

Third, iSchools, which "serve as a naturally occurring experiment of the creation of interdisciplinary academic units" (Wiggins & Sawyer, 2012), have a strong track record in education for data professionals; this is evidenced by the finding that the number of academic programs and courses of the iSchools is significantly larger than that of non-iSchools. There might be a number of reasons for this. One factor attributing to the wide range of curricular offerings at iSchools may be that they have faculty from a wide variety of subject disciplines. Another factor may be that many iSchools are home to academics from multiple disciplinary departments, including informatics, information system, or computer science departments. Certainly more input from those departments within their larger unit enable the iSchools to support extended curricular offerings.

One remaining question is what gaps remain in current education and training programs to produce a workforce of data professionals. To address this question, we first need to define data professional roles and responsibilities, then identify workforce needs for data professionals. In fact, there have been some efforts to disambiguate various data roles, including data curator, data scientist, data analyst, data manager, and data librarian (e.g., Lyon & Takeda, 2012; National Science Board, 2005; Swan & Brown, 2008) under the umbrella term of "data professional." Despite such efforts, different data roles have been often conflated as further roles and responsibilities have evolved over the years. For instance, the term "data scientist" has been used loosely for several years, leading to a general sense of confusion over the role and its duties. It is still fairly unclear what exactly the domain of data science is and what career paths are available for data scientists. Further, little insight exists on what skill sets should acquired to become a data scientist. Accordingly, the academic programs for data science have many different interpretations of their focus and learning outcomes depending on where it is used; programs from computer science departments highlight programming skills required to acquire, store, and process data, whereas programs from statistics departments and business schools focus on utilizing rigorous statistical methods to analyze and interpret the data. As such, there is a call for reaching an agreement on definition and clarification of different roles in the data workforce. Responding to such a call is critical for strengthening the identity of academic program courses to support the current and future data workforce.

## Conclusion

This paper provides a snapshot of a key facet of education for data professionals within ALA-accredited LIS schools. It should be noted that given the rate of increase of new programs, new programs were being created even as we conducted the study. As such, our list could not be exhaustive; rather, it is representative of the frequency and relative visibility of various programs and courses offered. Implications from this study are relevant to several areas that impact LIS education. These areas include professional standards for accreditation, program curriculum offerings, and the relevance of research course objectives and content as revealed by the language used in course titles and descriptions.

Based on the analysis of academic programs and curriculum preparing data professionals, we suggest that LIS educators engage in dialog in an attempt to model curricula to meet the needs of today's data environment and to address the direction needed to design continuing education programs. The LIS profession is in a position to advocate for the changes required to increase the flow in the data professional pipeline. LIS professionals have core skills in collecting, organizing, managing, and preserving data. Further, some have begun to advocate for a new role in manipulating and analyzing data using computational and statistical methods. However, such advocacy will require LIS educators and professionals to step outside their comfortable disciplinary silos and reach out to other disciplines to understand how data can be contextualized by the profession and integrated into their curricula.

As early as 1996, Van House and Sutton asserted that LIS schools should expand their focus at the institutional level and focus on specialization and hybridization. This assertion is still true today. LIS schools are being given opportunities to broaden and expand academic programs for data professionals. As Wallace (2009) argued, such opportunities are decidedly more beneficial than harmful.

## Acknowledgments

The author wishes to thank JELIS editor Dr. Peta Wellstead for her helpful comments and feedback on an earlier version of this paper.

## Appendix I: A List of Academic Programs for Data Professionals

*Arizona, University of*
School of Information
- Master of Science in Information—Emphasis Area: Data Science
- Digital Information Graduate Certificate

*California—Los Angeles, University of*
Department of Information Studies
- Master of Library & Information Science—Specialization: Informatics

*Dominican University*
Graduate School of Library and Information Science
- Certificate in Data and Knowledge Management
- Certificate in Digital Curation

*Drexel University*
College of Computing and Informatics
- Bachelor of Science in Data Science (Coming Fall 2016)
- Master of Science in Library and Information Science—Concentration: Digital Curation

*Illinois at Urbana Champaign, University of*
Graduate School of Library and Information Science
- Master of Science—Specialization: Data Curation
- Master of Science—Specialization: Socio-technical Data Analytics
- Master of Science in Bioinformatics

*Indiana University*
Department of Information & Library Science, School of Informatics and Computing
- Master of Library Science—Specialization: Data Science
- Master of Information Science—Specialization: Data Science
- Certificate in Data Science
- Ph. D. in Data Science Minor

*Maryland, University of*
College of Information Studies
- Master of Library Science—Specialization: Archives and Digital Curation
- Master of Library Science—Specialization: Community Analytics and Policy
- Master of Information Management—Specialization: Archives and Digital Curation
- Master of Information Management—Specialization: Data Analytics
- Curation and Management of Digital Assets Certificate

*North Carolina—Chapel Hill, University of*
School of Information and Library Science
- Master of Science in Information Science-Specialization: Digital Humanities
- Graduate Certificate in Digital Humanities
- Graduate Certificate in Digital Curation
- Post-Masters Certificate Data Curation

*North Texas, University of*
Department of Library and Information Sciences, College of Information
- Digital Curation and Data Management Graduate Academic Certificate

*Pittsburgh, University of*
School of Information Sciences
- Master of Science in Information Science—Specialization: Big Data Analytics
- Certificate of Advanced Study—Big Data Analytics

*Pratt Institute*
School of Information
- Master of Science in Library and Information Science—Concentration: Conservation and Digital Curation
- Master of Science in Library and Information Science—Concentration: Digital Humanities
- Master of Science in Library and Information Science—Concentration: Data Analytics, Research, and Assessment

*Rutgers University*
School of Communication and Information
- Bachelor's Degree in Information Technology and Informatics—Specialization: Data Science, Curation, and Management

*San Jose State University*
School of Information
- Post-Master's Certificate in Digital Curation
- Advanced Certificate—Pathway: Data Analytics and Data Driven Decision Making

*Simmons College*
School of Library and Information Science
- Digital Stewardship Certificate

*Syracuse University*
School of Information Studies
- Certificate of Advanced Study in Data Science

*Toronto, University of*
Faculty of Information
- Master of Information—Concentration Pathway: Knowledge Management & Information Management

*Washington, University of*
The Information School
- Master of Science in Information Management—Specialization: Data Science & Analytics

*Western Ontario, University of*
Faculty of Information & Media Studies
- Master of Library and Information Science—Program Content Areas: Information Organization, Curation, and Access

## References

Adams, M. O. N. (2006). The origins and early years of IASSIST. *IASSIST Quarterly, 30*(3), 5-15. Retrieved from http://www.iassistdata.org/sites/default/files/iq/iqvol303adams.pdf

Allard, S. (2015). Placements & salaries 2015: Titles & tasks from core to cutting edge. *Library Journal*. Retrieved from http://lj.libraryjournal.com/2015/10/placements-and-salaries/2015-survey/titles-tasks-from-core-to-cutting-edge/

Association of College & Research Libraries Research Planning and Review Committee (2014). Top trends in academic libraries: A review of the trends and issues affecting academic libraries in higher education. *College & Research Libraries News, 75*(6), 294–302. Retrieved from http://crln.acrl.org/content/75/6/294

Australian Library and Information Association (2014). *Future of the library and information science profession: Library and information professionals*. Retrieved from https://www.alia.org.au/sites/default/files/ALIA-Future-of-the-LIS-Profession-03-Professionals_0.pdf

Beagrie, N., & Pothen, P. (2001). Digital curation: Digital archives, libraries and eScience seminar. *Ariadne, 30*. Retrieved from http://www.ariadne.ac.uk/issue30/digital-curation

Blake, C., Stanton, J., & Larson, R. R. (2013). Filling the workforce gap in data science and data analytics. *iConference 2013 Proceedings* (pp. 1015–1016). doi:10.9776/13424

Cheng, B., Wang, M., Morch, A. I., Chen, N-S., Kinshuk, K., & Spector, J. M. (2014). Research on e-learning in the workplace 2000–2012: A bibliometric analysis of the literature. *Educational Research Review, 11*, 56–72. doi:10.1016/j.edurev.2014.01.001

Committee on Future Career Opportunities and Educational Requirements for Digital Curation. (2015). *Preparing the workforce for digital curation.* Washington, D.C.: National Academies Press. Retrieved from http://www.nap.edu/catalog/18590/preparing-the-workforce-for-digital-curation

Dumbuill, E., Liddy, E. D., Stanton, J., Mueller, K., & Farnham, S. (2013). Educating the next generation of data scientists. *Big Data, 1*(1), 21–27. doi:10.1089/big.2013.1510

Gordon-Murnane, L. (2012). Big data: A big opportunity for librarians. *Online, 5*, 30–34. Retrieved from http://www.infotoday.com/online/sep12/index.shtml

Hedstrom, M., Larsen, R.L., & Palmer, C. (2014). *Preparing the workforce for digital curation: The iSchools perspective*. Panel presented at the 9th International Digital Curation Conference, San Francisco, CA. Retrieved from http://www.dcc.ac.uk/webfm_send/1351

Jones, S. (2013, January 21). *IDCC13: What's in a name? The 'data scientist' symposium*. Retrieved from http://www.dcc.ac.uk/blog/idcc13-data-scientist-symposium

Lee, B., & Jeong, Y-I. (2008). Mapping Korea's national R&D domain of robot technology by using the co-word analysis. *Scientometrics, 77*(1), 3–19. doi:10.1007/s11192-007-1819-14

Lord, P., & Macdonald, A. (2003). *e-Science curation report: Data curation for e-science in the UK—An audit to establish requirements for future curation and provision. Report prepared for the JISC Support of Research Committee (JCSR)*. Retrieved from https://www.cs.york.ac.uk/ftpdir/pub/leo/york-msc-2007/information/vsr-curation/science-dc-report.pdf

Lyon, L. (2012). The informatics transform: Re-engineering libraries for the data decade. *The International Journal of Digital Curation, 7*(1), 126-138. doi:10.2218/ijdc.v7i1.220

Lyon, L., & Brenner, A. (2015). Bridging the data talent gap: Positioning the iSchool as an agent for change. *The International Journal of Digital Curation, 10*(1), 111–122. doi:10.2218/ijdc.v10i1.349

Lyon, L., & Takeda, K. (2012). *What is data scientist? (Data scientists in the wild)*. Presentation at the Microsoft eScience Workshop. Retrieved from http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/liz-lyon-microsoft-escience-chicago-october-2012-final.pdf

Martinez-Uribe, L., & Macdonald, S. (2009). *A new role for academic librarians: Data curation.* Retrieved from http://www.era.lib.ed.ac.uk/handle/1842/3207

McNulty, E. (November 3, 2014). *Big data jobs earn 24% more than other it positions, new study reveals*. Retrieved from http://dataconomy.com/big-data-jobs-earn-24-more-than-other-it-positions-new-study-reveals/

National Science Board. (2005). *Long-lived digital data collections: Enabling research and education in the 21st century*. Retrieved from http://www.nsf.gov/pubs/2005/nsb0540/nsb0540.pdf

Palmer, C. L, Weber, N, M., Muñoz, T., & Renear, A. H. (2013). Foundations of data curation: The pedagogy and practice of "purposeful work" with research data. *Archive Journal, 3*. Retrieved from http://www.archivejournal.net/issue/3/archives-remixed/foundations-of-data-curation-the-pedagogy-and-practice-of-purposeful-work-with-research-data/

Pettey, C. (October, 22, 2012). *Gartner says big data creates big jobs: 4.4 million IT jobs globally to support big data by 2015*. Retrieved from http://www.gartner.com/newsroom/id/2207915

Swan, A., & Brown, S. (2008). *The skills, role, and career structure of data scientists and curators: An assessment of current practice and future needs. Report to the JISC*. Retrieved from http://eprints.soton.ac.uk/266675/

Van House, N. A., & Sutton, S. A. (1996). The panda syndrome: An ecology of LIS education. *Journal of Education for Library and Information Science, 37*(2), 131–147. doi: 10.2307/40324268

Varvel, V. E., Palmer, C. L., Chao, T., & Sacchi, S. (2010). *Report from the Research Data Workforce Summit*. Retrieved from https://www.ideals.illinois.edu/handle/2142/25830

Wiggins, A., & Sawyer, S. (2012). Intellectual diversity and the faculty composition of iSchools. *Journal of the American Society for Information Science and Technology, 63*(1), 8–21. doi:10.1002/asi.21619