Meeting Data Workforce Needs: Indicators Based on Recent Data Curation Placements

Carole L. Palmer[1], Cheryl A. Thompson[1], Karen S. Baker[1] and Megan Senseney[1]

[1] Center for Informatics Research in Science and Scholarship, Graduate School of Library and Information Science, University of Illinois

Abstract

iSchools have been steadily advancing data curation education and practice in response to workforce demands. This paper reports on a formative evaluation of the Specialization in Data Curation at the University of Illinois, aimed at understanding job preparedness and work experiences of graduates and areas for improvement in data curation education. Survey results are complemented by additional graduate placement analysis. Employment and career satisfaction were high. Internships, practicum, and assistantships were considered key employability factors. Duties emphasize liaison and consulting, user instruction, data management, metadata, and policy development. About half of all placements were in academic libraries, with the second largest group in the corporate sector. This study, focused on the earliest formal LIS program in the U.S. dedicated to curating research data, provides important evidence of data curation responsibilities in the workforce and perceived educational gaps that can guide planning, design, and improvement of data curation programs.

1 Introduction

Education for information professionals has been evolving for many years to meet the challenges of digital content and infrastructure growth and complexity. Programs have emphasized different aspects of the profession, including digital librarianship, digital preservation, data stewardship, and digital curation. As documented in Gold (2010), "data curation" education and practice has been steadily advancing in the field of Library and Information Science (LIS) since 2006. However, the emergence of the field was recognized at least a decade earlier by both government agencies and the museum community (Palmer, et al., 2013). It gained momentum as scientists acknowledged the need for curation to sustain contemporary research (Gray, et al., 2002) and organizations emerged to promote best practices (Lord, et al., 2004).

Responding to the expected demand for expertise in the curation of research data, the Graduate School of Library and Information Science (GSLIS) at the University of Illinois began a Specialization in Data Curation in its MSLIS program in 2007. The specialization was created through a 2006 grant from the Institute of Museum and Library Services (IMLS) to develop educational capacity in the field of data curation, with an initial focus on the sciences. It was extended to include the humanities with a second award in 2008. To date, 63 graduates have completed the specialization.

This paper reports on a formative evaluation of the program, primarily a survey of graduates with the specialization, aimed at understanding work experiences, job preparedness, and areas for improvement in the program. Survey results are complemented by analysis of placement patterns of the graduates and the emergence of new kinds of positions for information professionals with responsibility for digital content.

Together, the different evaluation components provide a baseline for longer term tracking of the market and assessment of the specialization. More importantly, the results offer important benchmarks for other iSchools as they plan, develop, or advance educational programs in data curation. As an evaluation of the earliest formal LIS program dedicated to the curation of research data, it provides important evidence of actual data curation responsibilities in the workforce and perceived educational gaps, to enable better recruitment and design of programs to meet new data demands in the information professions.

2   Data Workforce Needs

A well-prepared and trained workforce is the key to managing and preserving data to advance science and scholarship, as asserted in numerous reports issued by federal agencies, including the ACLS (2006) and the NSF Blue-Ribbon Advisory Panel on Cyberinfrastructure (Atkins, 2003). Data curation requires a workforce with specific knowledge and skills to manage and preserve data to be scientifically useful to others (Rusbridge, 2007). The next generation of science needs professionals with expert capabilities to select and store data; support the discovery, access, and use of data; and ensure data integrity over time (Lord & Macdonald, 2003). As essential intermediaries between domain scientists and computer scientists in the system of cyberinfrastructure (Bowker & Star, 2009), data curators will be the experts that ensure that data are available for public access and fit for reuse.

With advances in technology and the changing conduct of science, new professional roles have emerged as expected (National Science Board, 2005; Hey, Tansley, & Tolle, 2009). Positions such as data curator, data archivist, data scientist, and data journalist now exist (Lyon, 2013), and recent growth of data curation positions has been documented (Maatta, 2012; Sierra, 2012). New positions in the area of data science, where discoveries are dependent on curated data (Stanton, et al., 2012), have attracted national attention by being named "the sexiest job" in 2012 (Davenport & Patil, 2012). However, despite calls for a more precise analysis of the data workforce needs and responsibilities (Varvel et al. 2010), little is known about how data curation roles are currently emerging in the workforce.

A number of important workforce studies have been conducted concurrent with data workforce changes, but unfortunately they have not been designed to identify trends specific to data roles for LIS professionals (Marshall et al. 2010; Sivak & De Long, 2009; Griffiths, 2009; Steffen, Lance, Russell & Lietzau, 2004; Walch, 2006). Moreover, they tend to not represent information professionals working outside of the traditional LIS settings.

3   LIS Data Workforce and Education Trends

Positions in data curation have proliferated while education capacity has developed more slowly, despite clear predictions on demand for data curation expertise in LIS:

> "Library educators have an important role to play in planning for and delivering appropriately skilled people to meet the latent demand for data librarians to manage the libraries' potential data curation role. Yet very few library and information science schools currently teach the skills that future data librarians will need." (Swan & Brown, 2008, p. 25)

Influenced in part by recent funding agency requirements for data management planning (Reznik-Zellen et al., 2012; Lyon et al., 2013), new responsibilities in research libraries have resulted in a range of new job titles with increasingly diverse data responsibilities (Bracke, 2011; Xia and Wang, 2013). The expansion in expected expertise for information professionals adds to a continued struggle for LIS identity and recognition (Fisher and Julien, 2009; Gray, 2013; Higgins, 2011), complicated by the fact that data responsibilities are closely intertwined with other towering professional roles in the digital realm, including information gatekeeping (Cox, 2013), and building information and knowledge infrastructure (Edwards et al., 2007; Monteiro, 2012; Soenher, Steeves & Ward, 2010; Edwards et al., 2013).

Library roles in *data* curation are necessarily evolving alongside those of *digital* curation (Gold, 2007; Gold, 2010), as documented in recent studies of job advertisements of library positions (Park and Lu, 2009, Kim, et al., 2012) and investigations of how library professionals feel about the new roles and library preparedness for providing data services (Tenopir et al., 2013). Case studies and the professional discourse are beginning to make explicit the institutional and university level requirements and experiences in building the digital data enterprise necessary for research and digital initiatives (e.g., Walters, 2009; Lage et al., 2011; Prom, 2011; Hswe et al., 2012; Newton et al., 2012; Jahnke, Asher, & Keralis, 2012; Illinois Research Data Initiative, 2013; Reznik-Zellen, Ademick, & McGinty, 2012; Tenopir et al., 2013).

Nonetheless, LIS education capacity is still uneven across schools, although the central role of LIS in digital information management and data curation has been acknowledged as a future thrust in the field (Heidorn, 2011). According to Harris-Pierce and Liu (2012), only a third of the LIS programs offer a course in data curation at the graduate level, with content addressing information resources, information organization, metadata, and technical knowledge and skills. An analysis of national data curation curriculum in LIS schools identified a total of 203 programs at 63 universities offering courses relevant to data curation, but most appeared to be part of digital library curriculum that covers digital content in a more generic way, with only a few schools offering programs concentrating specifically on contemporary demands of the data workforce (Varvel, Bammerlin & Palmer, 2012). While not yet empirically documented, many LIS and iSchools have since made significant progress on new programs. A couple of examples are the Data Curation emphasis within the Post-Masters Certificate at the University of North Carolina and the specialization in Curation and Management of Digital Assets at the University of Maryland.

Activity in continuing education for working professionals has progressed in parallel, offered by a variety of institutions and in a variety of formats. Since 2006, a sustained series of institutes has been offered in data curation at the University of Illinois (Renear et al., 2012), and in digital curation at the University of North Carolina (Hank, Tibbo & Lee, 2010). At the same time, non-LIS schools are quickly building capacity in data science education, including online offerings to accommodate working professionals (see, for example, Howe, 2012). Professional organizations and premier data centers are also providing outreach in best practices and tools for data curation, such as the institutes in data management sponsored by the Inter-university Consortium for Political and Social Research (ICPSR, 2012), the e-Science Institutes offered by ARL and CLIR/DLF, and the extensive and diverse set of activities and resources sponsored by the UK Digital Curation Centre.

Data curation education in LIS has focused on preparing new students as well as extending the skill set of current professionals in the workforce. A diversity of programs exists in terms of length, delivery modes, and level of certification or specialization. The program at Illinois reported here was incubated as the Data Curation Education Program (DCEP), and supported through IMLS grant funds, as were many other educational efforts referenced above. In addition to developing the Specialization in Data Curation in the masters program, DCEP produced research on education and workforce needs, and supported the Summer Institute in Data Curation for working professionals with events focused on the sciences and humanities data. With years of development and delivery of data curation education now completed, we can begin to assess outcomes within the context of the broader trends in LIS and the workforce at large. This report is one piece of the field's coming efforts to determine our goals and document our achievements in data curation as we move into the next generation of information professions, where data expertise will undoubtedly be a major part of what we contribute to our research institutions and society.

4   Methods

The formative evaluation design and survey development was guided by the following research questions:

1. What are current data curation workforce needs and future trends from the perspective of graduates?

2.   What are the work experiences of the graduates of the Specialization in Data Curation program?

3.   How well does the program prepare graduates for their jobs?

In developing the survey, the team reviewed questions from national and large-scale workforce surveys such as WILIS (Marshall et al. 2010), A*Census (Walch, 2006) and 8 R's (Sivak & De Long, 2009). A web-based survey was developed employing both closed and open-ended questions covering program assessment, employment status, job characteristics, career intentions, continuing education needs, and future trends in data curation. It also collected information on the respondents' current employer and on the data resources they are responsible for in their positions.

The survey was distributed in April 2013 to alumni graduating from December 2008 through 2012, applying a census sampling strategy; i.e., all 63 graduates received an email invitation to the survey. After two weeks, a reminder was emailed to non-respondents. The response rate was 37% (N=23). Despite the lower response rates for web surveys (Hayslett & Wildemuth, 2005), web surveys produce higher quality responses than offline methods (Gunter et al. 2002). The survey data provided highly informative and valuable indicators for considering next steps for the program and for iSchools interested in beginning programs.

The survey was also conducted in conjunction with ongoing placement analysis of graduates from the data curation program. Placement information, including current job title and employer, has been recorded for 84% of the 63 students graduating with the Specialization in Data Curation.

Quantitative data were loaded into R 3.01 software for analysis. Textual responses were analyzed using ATLASti 7. Analytical codes were developed using both an inductive and deductive approach. Open-ended responses were coded initially to identify emerging themes. Next, the authors reviewed the research questions and literature to generate additional codes. A codebook was created with the final set of codes for analysis. Two team members coded the data in a process for achieving inter-coder reliability. The survey instrument will be archived in the IDEALS repository (Thompson et al., 2013). This paper presents results from the quantitative and qualitative analyses arranged by topic areas.

5   Results

Results are reported for the following areas: respondent demographics, current employment, careers, program assessment, continuing education needs, and future trends.

5.1   Respondent Demographics

Survey respondents graduated between 2008 and 2012 with the majority graduating after 2010 (57%). Graduates were primarily female (63%) with a median age of 34 years (mean 35; std. dev. 9). By comparison, the median age category of recent LIS graduates from the Marshall et al. (2010) study was 31 to 35 years (Marshall et al., 2010), and, as of the year 2010, the median age of the US labor force was 41 years (Toossi, 2012). Seventeen percent of our respondents were non-Caucasian, a somewhat higher percentage of minorities than the 10%-11% from earlier surveys (Marshall et al., 2010), an outcome of the DCEP program's efforts to recruit underrepresented students. All respondents were located in the United States, currently living in 14 different states. Over a third of respondents were working in the Midwest primarily in Illinois and Iowa. California was another prominent location.

5.2   Current Employment

The survey asked graduates if they were currently working for pay. Despite the recent economic recession, 91% of respondents were employed at the time of the survey, with 2 unemployed and seeking work. The survey included questions for employed graduates about their position – whether it was considered full-time and considered a data curation position. Of those employed (n=21), all held full-time positions. Forty-eight percent considered their position to be in the field of data curation. Graduates indicated whether they had

the opportunity to apply the skills that they learned from the specialization. Majority (95%) of those employed agreed that they apply skills learned from the Specialization in Data Curation in their positions. Of those not working specifically in data curation (n=11), 90% had opportunities to apply their data skills.

The survey asked graduates to select the employer type from response options that best described their current employer. Graduates were working in a variety of settings. Among the most frequently reported were positions in an academic setting (55%). A few graduates were working in corporate and non-profit institutions, and one graduate worked in each setting - research center, data center, and government. These findings are fairly consistent with our program's placement information where we found graduates working in academic (49%), corporate (17%), and non-profit (15%) settings. Five graduates are working in government, three students are now employed in research centers, and two are employed in data centers. Of those working in a data curation position (n=10), 6 graduates were working in academic settings, and 1 each in government, research center and data center.

The survey asked for graduates' position titles and whether there was a more appropriate title for their current job. Coders analyzed free text responses for traditional LIS and non-traditional LIS positions, and found that traditional and non-traditional LIS roles were blended in the positions of many graduates. There was a diverse range of traditional library positions with respondents mentioning each of the following at least once: systems department head, reference librarian, subject librarian, index specialist, archivist, internal records manager, preservationist, cataloger, procedural documenter, and project manager. One graduate summarized their job as more 'liaison librarian' to disciplines, illustrating the articulation work involved: "I am currently arranging for data management support in collaboration with other offices."

There were twice as many positions coded as 'non-traditional' LIS roles as compared to the traditional LIS roles. For this survey analysis, non-traditional LIS roles were defined as those that emerged in the last decade such as the management, preservation, or curation of data and digital objects; work with digital repositories; and engagement with metrics and communities through social media. Titles associated with non-traditional roles included Data Manager, Data Management Consultant, Digital Project Analyst, Web Metrics, Social Media Specialist, Systems Architect, and Application Analyst.

In the survey, respondents selected their current annual salary from a list of salary ranges. Forty percent of respondents reported annual salaries between $50,000 and $59,999. A few graduates selected salary ranges of $60,000 - $69,999 and $70,000 - $79,999. All respondents indicated salaries of $30,000 or greater.

Job functions that you perform in your position. In the survey, graduates selected from a list of 20 duties. Overall, the respondents had positions comprised of several duties. Of the response options, most frequently reported duties were liaison and consulting (67%), user instruction (67%), data management (62%), metadata (62%), and policy development (62%). Thirty percent had supervisory responsibilities. With regard to data, 58% of all employed respondents had shared decision-making authority. Approximately 37% had some input to decision-making. Graduates working in data curation (hereafter referred to as data curators) all had duties in data management (100%), with high levels of responsibility in preservation planning (70%), data quality (70%), and compliance (70%). The duties topping the list for graduates not working in data curation (hereafter referred as non-data curators) were training (64%) and consulting and liaison (45%).

The survey gathered descriptions of current job duties in two questions – 'describe the work you do in your current position' and 'elaborate or specify other duties.' From analysis of open-ended responses, a set of categories emerged to describe the work of data curation professionals: technical, service and managerial duties. Technical duties were defined as those associated with development and support of digital technologies including software and hardware components. These duties ranged widely from managing existing data systems to designing new infrastructure. See Table 1 for examples of job duties by type. Service-oriented duties were primarily in community outreach and training. The service audience

included the public, scientists, data users, colleges, and universities. Administration and managerial duties were defined as work overseeing daily operations, policies, planning, and resources.

Technical duties
• Managing existing data systems
• Designing new infrastructure
• Generating data
• Preserving data (systems)
• Designing interface
• Developing of digital project workflows
• Consulting on data management plans (infrastructure-oriented)
• Analyzing data
• Ensuring quality control
• Developing applications (e.g., mapping data)
• Using tools for asset management
• Modifying existing software tools
• Documenting procedures
• Ensuring security
• Implementing access level
• Complying with policies

Service duties
• Assisting with data management plans at both planning and implementation stages
• Training scientists and data users on best practices
• Providing support in data access
• Engaging with data management, informatics and design issues
• Explicating data management problems

Admin/managerial duties
• Managing personnel
• Coordinating projects
• Managing databases
• Overseeing collections
• Managing systems
• Allocating resources
• Developing policies
• Overseeing operations
• Writing reports

Table 1: Examples of job duties by type

Describe the data you work with in your position. The survey asked graduates to select the domain areas and formats of data that they encounter in their current job. The most frequently reported domain areas were Life Science (43%) and Physical Science (43%). A few graduates selected Social Science, Business, Government, Technology, and Health. Graduates were responsible predominantly for digital data, such as text (57%), images (52%), presentations (48%), videos (48%), spreadsheets (48%), and databases (48%). Data curators identified responsibility for spatial data (70%), computation models (50%), computational code (60%), and spreadsheets (80%). Interestingly, 27% of those in non-data curator positions had responsibility for spreadsheets.

All but one employed respondent is responsible for digital objects in their current position. As noted in the open-ended responses, most work with both digital and physical objects. Four positions focused primarily on research data, with only one noting responsibility for 'Big Data.' Almost half worked with research data and with non-data objects, such as an internal digital library and audio-visual files. For graduates working with physical objects, formats included books, paper reports, reel-to-reel tape, and vertical files.

Institutional Context. In response to several questions, we found references to institutional context. The importance of making a business case for data curation was a recurrent theme. Six graduates described how they had to get 'buy in' from scientists and administrators within their organization. One noted: "the real challenge is convincing researchers, research administration and even funders that data curation, not just data sharing, is a good return on investment." Resource allocation challenges included "convincing data creators to consider the long-term curation of data in the face of time and budget pressure." One graduate indicated a need for more preparation in "rigorous change management," exclaiming: "It is really, really hard to be the bearer of change!" Communication across domains is vital to making progress. As one graduate stated: "I'm speaking to computer scientists, engineers and PhDs in math/physics. The link between LIS and data curation is not apparent to most and it took me a while to understand where languages of expertise met and where they diverged." The lack of appropriate infrastructure was noted as a barrier: "I'm managing data as part of an assessment project...There is interest in promoting these services but we have to get a few things in place for infrastructure/resources assigned."

5.3   Careers

The survey asked graduates about the number of positions held since graduation and length in their current position. A majority of respondents (96%) have held only one position, holding their current position for a mean of 1.6 years (std. dev. 1.2) with a range from 1 week to 5 years. Respondents were asked to indicate their level of satisfaction with data curation as a career. Most graduates were satisfied with data curation as a career (90%). The survey also gathered graduates' level of agreement with statements about career opportunities. The majority agreed that they have opportunities to develop leadership skills (91%) and to advance their career (100%).

In planning for workforce demands, retention of current employees is an important consideration. Data curators were asked whether they plan to still be working in data curation in 5 years. Majority (90%) planned to still work in data curation. The survey asked non-data curators if they plan to pursue employment in the data curation field in the future. More than half (62%) report an intention to pursue employment in the data curation field.

Each career survey question was followed by an open-text box where many graduates elaborated on their answers. Overall, graduates described data curation careers with positive terms. For instance, a graduate described data curation as "what I love to do." One graduate described data curation as "a super interesting field," while another graduate noted "I like providing access to cool stuff." One graduate enjoyed the diversity and opportunities associated with the work: "I do so many things every day, I am working with lots of totally different people, so much opportunity, I can't imagine ever getting bored or stuck."

Respondents were asked to describe any previous education or experiences that helped them get their current job, responding with comments on formal and informal education and prior positions. A few graduates reported that previous degrees in domain sciences (e.g., geology, biology, agronomy) helped them get their current jobs, of which two had both undergraduate and graduate degrees. A digital librarian mentioned a certificate in digital libraries helped them be competitive for their current job.

Previous work experience, both paid and unpaid, was of interest to employers. Five graduates described internships in data curation, domain sciences, or relevant LIS (e.g., preservation, copyright). Eight respondents reported on the importance of their work experience prior to the LIS program, including

experience in research settings and LIS settings. One credited their previous career in the non-profit sector: "I had a career in nonprofits for about 10 years, 4 of which were in training. My presentation and facilitation skills were a huge part of my getting this job, as well as my general tech-savvy-ness." Five mentioned the value of their graduate student assistantships. For example: "My experience as a research assistant in UIUC's NSF-Funded [project title redacted] helped me learn about creative research methods and collaborative work in the social sciences..."

In open-ended responses, seven graduates mentioned their fieldwork experiences helped them get their current job. A variety of fieldwork sites were reported such as universities, museums, and data centers. A respondent suggested: "More hands-on work would be useful for some of the classes. I know of at least two other people who went through the program and are confident of our grasp on theory, but not so much on our ability to apply that knowledge." They also noted the value to employers of new graduates that have a combination of education and work experience.

5.4   Program Assessment

The survey gathered information on how effective the program was in preparing them to meet their professional obligations. From the response options, most graduates (74%) rated the program as very effective or effective in preparing them. Twenty-six percent reported the program was somewhat effective in preparing them. The survey also asked respondents to select which topics the program prepared them for in their work. Three-fourths of graduates reported that the program prepared them for metadata and documentation. Almost half of respondents felt prepared for preservation planning (61%), modeling and ontologies (57%), data management (52%), and programming (48%). See Figure 1. Overall, the qualitative responses were positive about the program. Graduates were appreciative of the opportunities to pursue data curation in their graduate program. For instance, one graduate responded, "I'm grateful for it. Those courses were what I loved about grad school."
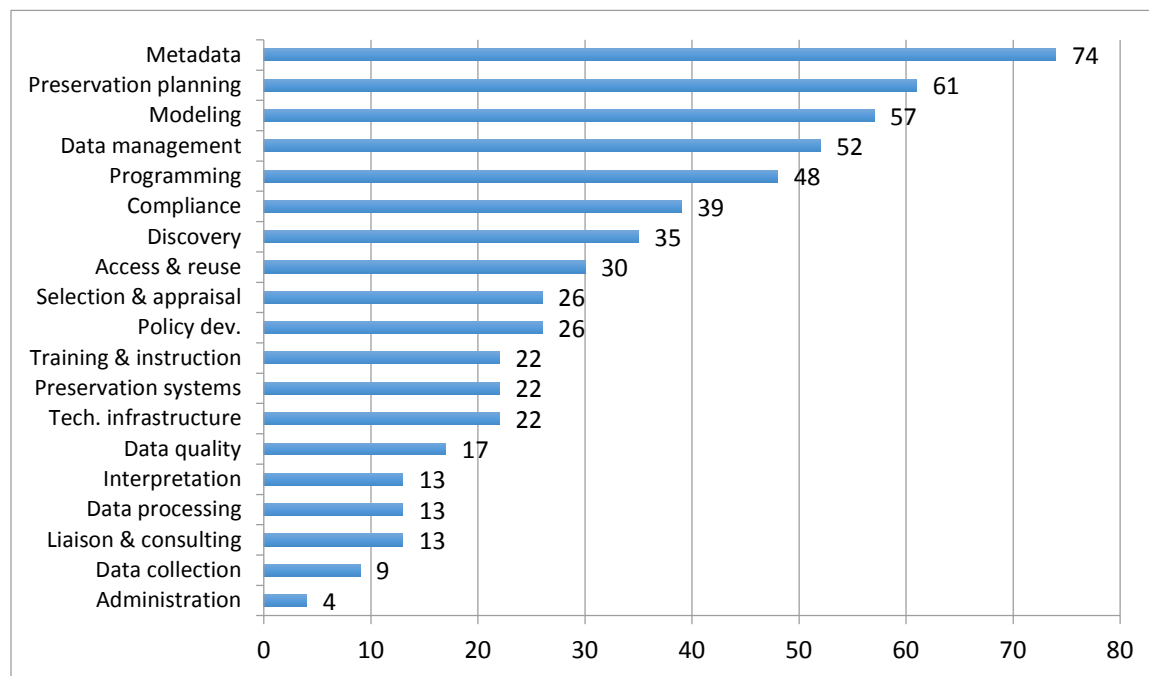

Figure 1: Percent who felt prepared for data curation topics in their professional work

In a series of open-ended questions, the survey asked graduates about the most useful topic in their career, topics missing from curriculum, most valuable aspects of the program and recommendations for program improvements. Graduates' responses broke down into four general categories: useful topics, useful courses,

specific skills, and program organization. The three topics most cited as useful were: current trends and the data curation landscape; metadata and documentation; and computer programming. The data curation courses that graduates frequently mentioned as having been particularly useful in their careers were: Digital Preservation, Metadata in Theory and Practice, Information Modeling, and Systems Analysis and Management. Among specific skills mentioned by respondent were Python, XSLT, XML, Cocoon, RDF, and SQL. One graduate wrote, "I can't think of a single course/area of study that I haven't drawn upon in my work." In addition to coursework, 30% of respondents cited the opportunity to build a strong network of instructors and colleagues as one of the most successful aspects of the program. Other identified strengths included hands-on work, breadth of knowledge and skills, and the rigor of the coursework.

Graduates were also asked to make recommendations on program improvements. The most frequent recommendation, mentioned by four respondents, was providing a greater emphasis on computer programming. Other recommendations offered by two or more respondents included more emphasis on data-specific domains and change management. Respondents reflected on the importance of experiential learning by recommending more hands-on work in the classroom (22%), with a few suggesting that fieldwork should be a program requirement. In open-ended responses, a few graduates requested more engagement with domain communities and data producers either in the classroom or through fieldwork. One insightful comment identified how education will need to change as the field evolves:

> "grow the connection with practitioners, prep data curation students for program development roles in short term, then expand to include both management and detail-oriented worker paths (we will eventually need both, but the short term need is much more for visionary leaders, [in my opinion])."

The survey asked respondents about whether they completed and found useful practicum and internship experiences during their data curation program. Fifty-two percent of graduates completed a practicum or internship while studying. More than half the respondents (52%) recommended that students complete a practicum or internship. Of those that did not complete a practicum (n=10), 7 graduates wished they had completed a practicum. Interestingly, graduates that completed a practicum (75%) felt more prepared for their duties of computer programming than those that did not complete a practicum (18%).

5.5   Continuing Education Needs

The survey asked respondents whether they were interested in pursuing continuing education opportunities and whether they had pursued any additional education opportunities. A majority of respondents (87%) indicated that they were interested in continuing education, and 61% had already pursued additional education or professional development since graduation. Respondents not interested in continuing education were asked to select the reason. From the response options, the three respondents not interested in continuing education specified a lack of time and already having the skills that they need.

Those interested in continuing education (n=20) were asked to rank the top three topics that they were most interested in pursuing. The most frequently ranked topics were metadata, modeling, data interpretation, and infrastructure (see Table 2). For the first ranked topics, metadata was the most frequent first choice. Surprisingly, programming was the only topic not ranked by any of the graduates. Additional continuing education topics suggested by respondents included research administration and proposal writing.

Topics	Ranked – any position (1-3)	Ranked #1	Ranked #2	Ranked #3
Metadata	30	20	10	0
Interpretation & analysis	30	10	15	5
Modeling	30	5	5	20

Tech. infrastructure	25	10	0	15
Administration	20	10	10	0
Preservation systems	20	10	10	0
Selection & appraisal	20	0	5	15
Access & reuse	15	5	0	10
Training & instruction	15	5	5	5
Data management	15	0	15	0
Data quality	15	0	10	5
Discovery	15	0	5	10
Data processing	10	5	0	5
Policy development	10	5	0	5
Preservation planning	10	5	5	0
Compliance	10	0	5	5
Liaison & consulting	5	5	0	0
Data collection	5	0	0	5
Programming	0	0	0	0

Table 2: Percent Ranking Top Continuing Education Topics (n=20)

The most preferred delivery method for continuing education was a one-time event of 1-2 working days (30%), followed by a one-time event of 3-5 working days (15%), and a course of 1-5 contact hours per week for one semester (15%). Graduates who have already pursued continuing education opportunities (n=14) reported participating in a range of delivery modes, with a concentration in webinars (36%) and conferences (29%). Additional delivery modes were certification programs, semester-long courses, summer courses, workshops, seminars, and discussion groups. Code Academy and MOOCs were also specified for online options.

Topics of completed continuing education included program management, scientific data processing, big data, computer science, web development, business analysis, semantic technologies, digital humanities, digital archives and records management, Resource Description and Access, and higher education administration. Respondents also mentioned learning specific tools and software, such as GIT, R, SQL, and Drupal, as well as the Python programming language.

5.6   Future Trends

Related to recent discussions of data curation and data stewardship issues and agendas (e.g. Jahnke et al. 2012; RDSA, 2013), the survey asked graduates open-ended questions examining perceptions of emerging issues for data curation professionals. Responses included lively usage of verbs—such as, managing, defining, bridging, educating, drumming up, convincing, and selling—to describe their future work in data curation. As one noted: "to some extent, I don't see that one can escape managing digital files/records of some kind." Many responses echoed the findings of the 2010 Research Data Workforce Summit on the need for engagement with current practice in data centers and the importance of communicating and bridging across domains (Varvel et al. 2010).

Graduates foresee increasing levels of management of complex datasets and anticipate issues with data formats and sources, expressing concerns with video, media production, linked data, and streamed sensory data in "rapidly changing information environments." Continuing education was reported as highly important for practicing data curation professionals. Many aspects in the data curation field are rapidly changing including data formats, standards, and best practices, and there is an urgent need to stay informed and keep their knowledge up-to-date. Two respondents specifically cited the critical need for continuing professional development, not only in data curation but also in the domains where data are generated. The

ability to collaborate and communicate was seen as vital for coordination of activities across communities and institutions. Six respondents referred to the need to keep scientists and administrators informed and to foster "buy-in" to new practices and services. Specific comments noted the importance of "creating awareness of data as something that is useful to be shared" and promoting awareness of data curation more generally. Respondents described having to make the business cases for data curation and clarify how data curation differs from standard operating procedures.

Finally, graduates expressed the need to clarify the role of LIS in relation to data curation. They highlighted the continuing need to address job titles, the meaning of a data curation degree, and the work of LIS because "the link between 'LIS' and data curation was not apparent to most..." Graduates described having to explain to scientists, employers and other research staff what data curation is and how it fits with other data work.

6   Discussion

Similar to respondents in Marshall et al.'s (2010) general study of recent LIS graduates, the Specialization in Data Curation graduates were employed with high levels of career satisfaction. Academic institutions were the top employer for data curation graduates, as seen with LIS graduates (Marshall et al. 2010) and archivists (Walch et al. 2006). While slightly more than half of graduates were not in data curation positions, per se or exclusively, data skills and knowledge were applicable to a wide range of institutional settings and positions as shown by 90% of graduates applying data skills to their current job.

Also as seen in Marshall et al. (2010), most data curation graduates (74%) rated the program as very effective or effective in preparing them to meet professional obligations. In terms of job preparedness, respondents reported that the program prepared them highly in the areas of metadata, preservation planning, and modeling. This contrasted with Marshall et al. (2010) where general LIS graduates reported gaining basic knowledge of the field, information seeking, and ethics from their LIS education. The differences would be expected in the two studies of different aspects of the field, but it also suggests that a general LIS education would be far from adequate for current data curation positions.

Respondents cited internships, practicum, and assistantships as key factors in their employability. As seen in Marshall et al. (2010), data curation graduates also reported that practicums or other hands-on experience were beneficial, suggesting practical experience as an area for program improvement. Experiential learning in the classroom and through external fieldwork is clearly advantageous, with three respondents suggesting that some form of fieldwork be a required for completion of the specialization. As one respondent remarked, "just like any field there is a vast difference between theory and practice."

More than half of all respondents are actively engaged in liaison and consulting, user instruction and training, data management, metadata and documentation, and policy development. These are all areas where best practices are actively being developed. As would therefore be expected, respondents strongly recommended that data curation programs sustain a network of students, instructors, and alumni for longer term engagement with other professionals in similar roles. More curricular emphasis on data-driven domains and active domain engagement was also recommended. This is a clear need since data professionals will increasingly provide services directly to researchers who produce data and will work in partnership with them on data management planning, implementation, and development of tools and value-added services.

Technical expertise was viewed as highly important, and it is particularly interesting that only 43% of respondents listed it as one of their duties. Graduates seem to perceive a need for such technical skills even if they are not currently utilizing them. Other common duties, such as data management and policy planning, are likely to remain a prominent feature of data curation job descriptions, though it is possible that focus on these will decrease as functional data infrastructures are established in conjunction with data management practices.

Collaboration and communication across domains will be important for the emergent field of data curation. Graduates emphasized the importance of communicating the role of data curation in the larger research arena. Respondents cited having to define data curation and explain how a curation approach differs from their standard operating procedures. This suggests there is an increasing need to define the jurisdiction of data curators (Abbott, 1988) and disambiguate the various roles within the research process that can best be handled by data professionals (Varvel et al. 2010).

As the field continues to evolve, iSchools will want to consider how to offer continuing education and professional development opportunities to their alumni. At present, graduates are mostly taking advantage of webinars even though nearly half of the respondents stated a preference for options like workshops lasting either 1-2 days or 3-5 days. The ubiquity and affordability of the webinar format will no doubt remain attractive to practicing professionals, but there is also a market for workshops and institutes that can fill the need for continuing development in emerging best practices.

7   Conclusion

The survey and placement analysis shows that the Specialization in Data Curation is meeting workforce needs, as evidenced by level of employment and diversity of job types. According to placement analysis, about half of all positions were outside of academic libraries, with the second largest group in the corporate sector. While satisfaction with the program and with job placements was high, attention is needed in the areas suggested for improvement especially in providing more experiential learning and more applied data curation opportunities. The advice from one respondent "to keep innovating" is important to all iSchools.

The National Data Stewardship Alliance (2014) finds that "studies must be broadened and repeated over time to establish a robust evidence base from which generalizable guidance can be drawn" (p. 23). In support of this agenda, there are several expected next steps for further formative evaluation, including interviews with selected participants and continued tracking for longitudinal analysis of graduates early in their positions and as they have longer tenure in the field. We are also interested in expanding the study across more data curation programs to make progress on larger trends. As more data curation graduates enter the workforce, iSchools will have expanded opportunities to assess gaps in job preparedness, shape curricula based on reported job duties, and build programs that foster long-term professional development.

Evaluation of progress is essential for the field to accurately scope its professional responsibilities in data curation and respond with quality education programs. With appropriate workforce preparation, data curation can move us beyond the perception of data as 'a problem' (Jahnke, 2012) to the opportunities and vision of a culture of knowledge built upon a 21st century foundation of data.

8   References

Abbott, A. (1988). The system of professions: An essay on the division of expert labor. University of
        Chicago Press.
American Council of Learned Societies (ACLS) (2006). Our cultural commonwealth: The report of the
        American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities
        and Social Sciences. Retrieved from
        http://www.acls.org/cyberinfrastructure/ourculturalcommonwealth.pdf
Atkins, D. (2003). NSF-AP Report: Revolutionizing science and engineering through cyberinfrastructure:
Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure.
        Retrieved from http://hdl.handle.net/10150/106224
Bowker, G. C., & Star, S. L. (2009). Cyberscholarship; or, "A rose is a rose is a…". Educause, 44(3), 6-7.
        Retrieved from http://net.educause.edu/ir/library/pdf/ERM0937.pdf
Bracke, M. S. (2011). Emerging data curation roles for librarians: A case study of agricultural data.
        Journal of Agricultural & Food Information, 12(1), 65-74.

Cox, A. M., & Corrall, S. (2013). Evolving academic library specialties. *Journal of the American Society for Information Science and Technology, 64*(8), 1526-1542.

Davenport, T. H., & D. J. Patil (2012). Data scientist: the sexiest job of the 21st century. *Harvard Business Review, 90*(10), 70-77.

Edwards, P. N., Jackson, S.J., Chalmers, M. K., Bowker, G. C., Borgman, C. L., Ribes, D., Burton, M., & Calvert, S. (2013). Knowledge infrastructures: Intellectual frameworks and research challenges. Ann Arbor, Report of a workshop sponsored by the National Science Foundation and the Sloan Foundation, Ann Arbor, Michigan. Retrieved from http://hdl.handle.net/2027.42/97552

Edwards, P. N., Jackson, S., Bowker, G., & Knobel, C. P. (2007). Understanding infrastructure: dynamics, tensions, and design. Final report of the workshop, "History and Theory of Infrastructure: Lessons for New Scientific Cyberinfrastructures", Ann Arbor, Michigan. Retrieved from http://deepblue.lib.umich.edu/handle/2027.42/49353

Fisher, K. E., & Julien, H. (2009). Information behavior. *Annual Review of Information Science and Technology, 43*(1), 1-73.

Gold, A. (2010). Data curation and libraries: short-term developments, long-term prospects. Office of the Dean (Library). Retrieved from http://works.bepress.com/agold01/9

Gold, A. K. (2007). Cyberinfrastructure, data, and libraries. D-Lib Magazine, 13(9/10), 1-11

Gray, J., Szalay, A.S., Thakar A., Stoughton C., & vandenBerg J. (2002). Online scientific data curation, publication, and archiving. Retrieved from: http://arxiv.org/abs/cs/0208012

Gray, S. W. (2013). Locating librarianship's identity in its historical roots of professional philosophies: towards a radical new identity for librarians of today (and tomorrow). IFLA Journal, 39(1), 37-44.

Griffiths, J. (2009). The future of librarians in the workforce. Retrieved from http://libraryworkforce.org/tiki-index.php

Gunter, B., Nicholas, D., Huntington, P., & Williams, P. (2002). Online versus offline research: implications for evaluating digital media. In Aslib Proceedings (Vol. 54, No. 4, pp. 229-239). MCB UP Ltd.

Hank, C., Tibbo, H. R., & Lee, C. A. (2010). DigCCurr I Final Report, 2006-09: Results and recommendations from the Digital Curation Curriculum Development Project and the Carolina Digital Curation Fellowship Program. Chapel Hill, NC: School of Information and Library Science, University of North Carolina. Retrieved from http://www.ils.unc.edu/digccurr

Harris-Pierce, R. L., & Liu, Y. Q. (2012). Is data curation education at library and information science schools in North America adequate? New Library World, 113(11/12), 598-613.

Hayslett, M. M., & Wildemuth, B. M. (2005). Pixels or pencils? The relative effectiveness of Web-based versus paper surveys. Library & Information Science Research, 26(1), 73-93.

Heidorn, P. B. (2011). The emerging role of libraries in data curation and e-science. *Journal of Library Administration, 51*(7/8), 662-672.

Hey, T., Tansley, S., & Tolle, K. (Eds.). (2009). The fourth paradigm: Data-intensive scientific discovery. Redmond, Washington: Microsoft Research. Retrieved from http://research.microsoft.com/en-us/collaboration/fourthparadigm/

Higgins, S. (2011). Digital curation: The emergence of a new discipline. *International Journal of Digital Curation, 6*(2), 78-88.

Howe, B. (2012). Data science curricula at the University of Washington. Presented at the Microsoft eScience Institute, Chicago, Illinois. Retrieved from http://research.microsoft.com/apps/video/default.aspx?id=174393

Hswe, P., Giarlo, M. J., Belden, M., Clair, K., Coughlin, D., & Klimczyk, L. (2012). Building a community of curatorial practice at Penn State: A case study. *Journal of Digital Information*, *13*(1). Retrieved from http://journals.tdl.org/jodi/index.php/jodi/article/view/5874

ICPSR Inter-university Consortium for Political and Social Research. (2012). Summer program offering courses on data science. Retrieved from http://www.icpsr.umich.edu/icpsrweb/ICPSR/support/announcements/2012/02/summer-program-offering-courses-on-data

Illinois Research Data initiative (2013). eResearch Services Task Force Final Report. Retrieved http://www.library.illinois.edu/committee/exec/supplement/s20122013/eresearch_services_final_report.html

Jahnke, L. M., Asher, A., & Keralis, S. D. C. (2012). The problem of data: Data management and curation practices among university researchers. Council on Library and Information Resources. Retrieved from http://digitalcommons.bucknell.edu/fac_pubs/52/

Kim, J., Warga, E. and Moen, W. (2012), Digital curation in the academic library job market. Proc. Am. Soc. Info. Sci. Tech., 49: 1–4. doi: 10.1002/meet.14504901283

Lage, K., Losoff, B., & Maness, J. (2011). Receptivity to library involvement in scientific data curation: A case study at the University of Colorado Boulder. *Libraries and the Academy*, *11*(4), 915-937.

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. JISC Committee for the Support of Research. Retrieved from http://www.jisc.ac.uk/uploaded_documents/e-scienceReportFinal.pdf

Lord, P., MacDonald, A., Lyon, L., & Giaretta, D. (2004). From data deluge to data curation. Proceedings of the UK e-Science All Hands Meeting, Nottingham, September 2004, Retrieved from: http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/150.pdf

Lyon, L., Wright, S., Corti, L., Edmunds, S., & Bennett, F. (2013). What is a data scientist? Panel presented at 2013 International Digital Curation Conference, January 14-17, 2013, Amsterdam, Netherlands.

Maatta, S. L. (2012). A job by any other name: A few bright spots shine for the class of 2011. *Library Journal*, *136*(17), 18-25.

Marshall, J. G., Morgan, J. C., Rathbun-Grubb, S., Marshall, V. W., Barreau, D., Moran, B. B. & Thompson, C. A. (2010). Toward a shared approach to program evaluation and alumni career tracking: Results from the workforce issues in library and information science 2 study. *Library Trends*, *59*(1), 30-42.

Monteiro, E., Pollock, N., Hanseth, O., & Williams, R. (2012). From artefacts to infrastructures. *Computer Supported Cooperative Work*, 1-33.

National Data Stewardship Alliance. (2014). 2014 National Agenda for Digital Stewardship. Retrieved August 15, 2013 from: http://www.digitalpreservation.gov/ndsa/documents/2014NationalAgenda.pdf

National Science Board (NSB) (2005). Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century. NSB-05-40. Retrieved from http://www.nsf.gov/pubs/2005/nsb0540/.

Newton, M. P., Miller, C. C., & Bracke, M. S. (2010). Librarian roles in institutional repository data set collecting: outcomes of a research library task force. *Collection Management*, *36*(1), 53-67.

Palmer, C. L., Weber, N., Muñoz, T., & Renear, A. H. (2013). Foundations of data curation: The pedagogy and practice of 'purposeful work' with data. *Archives 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/

Park, J., & Lu, C. (2009). Metadata professionals: Roles and competencies as reflected in job announcements 2003---2006. *Cataloging and Classification Quarterly*, 47(2), 145-160.

Prom, C. (2011). Making digital curation a systematic institutional function. *International Journal of Digital Curation*, *6*(1), 139-152.

Renear, A., Cragin, M., Heidorn, B., Palmer, C. L., Smith, L., Unsworth, J., & Senseney, M. (2012). Centuries of knowledge: Graduate School of Library and Information Science data curation education program. Final Report to IMLS. Retrieved from http://hdl.handle.net/2142/30845

Reznik-Zellen, R. C., Adamick, J., & McGinty, S. (2012). Tiers of research data support services. Journal of eScience Librarianship, 1(1). Retrieved from http://dx.doi.org/10.7191/jeslib.2012.1002

RDSA, 2013. Research Digital Stewardship Agenda 2014, Research Digital Stewardship Alliance. Retrieved from http://www.digitalpreservation.gov/ndsa/nationalagenda/index.html.

Rusbridge, C. (2007). Create, curate, re-use: the expanding life course of digital research data. EDUCAUSE Australasia 2007 Advancing Knowledge Pushing Boundaries. CAUDIT. Retrieved from http://hdl.handle.net/1842/1731

Sierra, T. (2012). Staffing for the future: ARL university library hiring in 2011. *Proceedings of the ARL Fall Forum,* Washington D.C.

Sivak, A., & De Long, K. (2009). The blind man describes the elephant: The scope and development of the 8Rs Canadian Library Human Resource Study. *Library Trends*, *58*(2), 167-178.

Soehner, C., Steeves, C., & Ward, J. (2010). E-Science and Data Support Services: A Study of ARL Member Institutions. Washington, DC: Association of Research Libraries.

Stanton, J., Palmer,C.L., Blake,C., & Allard,S. (2012). Chapter 6 Interdisciplinary Data Science Education, In Xiao, N. & McEwen, L.R. (Eds) Special Issues in Data Management (Vol 1110, pp. 97-113). ACS Symposium Series; American Chemical Society: Washington, DC, 2012

Steffen, N., Lance, K. C., Russell, B., & Lietzau, Z. (2004). Retirement, retention and recruitment: The future of librarianship in Colorado. Denver, CO: Colorado State Library, Library Research Service.

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. Truro, Cornwall, UK.

Tenopir, C., Sandusky, R. J., Allard, S., & Birch, B. (2013). Academic librarians and research data services: Preparation and attitudes. *IFLA Journal*, *39*(1), 70-78.

Thompson, C. A., Baker, K. S., Senseney, M, Varvel, V. E., & Palmer, C. L. (2013). Instrument for the GSLIS Specialization in Data Curation alumni survey, 2008-2012. UIUC Ideals Repository. http://hdl.handle.net/2142/46085

Toossi, M (2012). Labor force projections to 2020: a more slowly growing workforce. *Monthly Labor Review, 135*(1), 43–64.

Varvel Jr, V. E., Bammerlin, E. J., & Palmer, C. L. (2012). Education for data professionals: a study of current courses and programs. Proceedings of the 2012 iConference, ACM.

Varvel, V. E., Palmer, C. L., Chao, T., & Sacchi, S. (2010). Report from the Research Data Workforce Summit. Retrieved from http://hdl.handle.net/2142/25830

Walch, V. I. (2006). A*CENSUS results. Retrieved from http://www2.archivists.org/initiatives/acensus-archival-census-education-needssurvey-in-the-united-states

Walters, T. O. (2009). Data curation program development in US universities: The Georgia Institute of Technology example. *International Journal of Digital Curation*, *4*(3), 83-92.

Xia, J., & Wang, M. (2013). Competencies and responsibilities of social science data librarians: An analysis of job descriptions. College & Research Libraries. Retrieved from http://crl.acrl.org/content/early/2013/02/06/crl13-435.full.pdf+html

9   Table of Figures

10   Table of Tables