Applying Translational Principles to Data Science Curriculum Development

Liz Lyon
School of Information Sciences
University of Pittsburgh
+1 412 624 9436
elyon@pitt.edu

Eleanor Mattern
School of Information Sciences
University Library System
University of Pittsburgh
+1 412 648 5908
emm100@pitt.edu

Amelia Acker
School of Information Sciences
University of Pittsburgh
+1 412 624 4939
aacker@pitt.edu

Alison Langmead
Dietrich School of Arts & Sciences
School of Information Sciences
University of Pittsburgh
+1 412 648 2407
adl40@pitt.edu

ABSTRACT
This paper reports on a curriculum mapping study that examined job descriptions and advertisements for three data curation focused positions: Data Librarian, Data Steward / Curator, and Data Archivist. We present a transferable methodological approach for curriculum development and the findings from our evaluation of employer requirements for these positions. This paper presents "model pathways" for these data curation roles and reflects on opportunities for iSchools to adopt translational data science principles to frame and extend their curriculum to prepare their students for data-driven career opportunities.

General Terms
Training and education.

Keywords
Curriculum development, translational data science, research data curation, iSchools.

1. INTRODUCTION AND CONTEXT
The growing focus on data science, research data management services and associated data curation and preservation strategies represents evidence of the increasing operational impact of the data deluge, the need for data infrastructure development and a realization that significant workforce capacity and capability challenges are emerging. Employers in all sectors are seeking graduates to fill a diverse mix of data-related roles, characterized by a broad range of data literacy skills and competencies, combined with disciplinary knowledge and practical experience. In this complex landscape, Information Schools (iSchools) are reviewing curriculum requirements and developing new data-centric courses to build capacity in the workplace and to support data-driven careers in the 21st century.

This paper will present the outcomes of a curriculum mapping exercise, which has built on translational principles (i.e. "*translating research into practice*") [1] and has recognized the distinctiveness of different data science roles. The methodology is transferable, with a particular emphasis in this paper on career options in data curation and preservation; we highlight opportunities for innovative course offerings and the development of new educational collaborations and partnerships.

2. RESEARCH QUESTIONS
The study addresses three specific areas of interest:

1. What are the skills, competencies, knowledge, experiences and education required for the distinct data science roles?
2. How do these data science role requirements map to current curriculum topics and course offerings?
3. What opportunities emerge for new collaborations and partnerships in developing the data science curriculum?

3. LITERATURE REVIEW
The importance of data for libraries was recognized as early as 2006 [2] [3] and data librarianship was identified as a "gap in the market" in 2008 [4]. The roles and responsibilities of librarians and data were reviewed [5] [6] [7] and the need to re-skill subject and liaison librarians has been described [8], [9]. Surveys of RDM activities in libraries have been published [10], [11] which demonstrate a gradual ramping-up of infrastructure and service delivery.

Two recent reports from the UK have highlighted the requirement for "*a skilled workforce and data-confident citizens*" [12] and "*a severe shortage of UK data talent*" [13]. Whilst these reports have primarily focused on data analytics, earlier reports have described "*a dearth of skilled (data) practitioners*" [14] and "*the current paucity of data scientists*" [15], recognizing contrasting roles and responsibilities. Marketplace analysis of trends data from Indeed.com, demonstrates a steady growth in data-related positions [16], reinforcing the demand : supply ratio imbalance. Examining the range of nomenclature for describing these positions, "a *need to disambiguate and develop definitions for professional roles*" [17] has been recognized. Four discreet data roles were identified by Swan and Brown [15]. In this paper we will draw on the family of (six) data scientist roles described by Lyon and Brenner [18]: data

librarian (University or Research Institute), data archivist (National Archive), data steward (Data Center), data analyst (Corporate sector), data engineer (IT Company) and data journalist (News & Media). Typical employment locations are indicated in brackets; this is a key perspective since the tangible requirements of real-world settings, are positioned to be primary drivers in curriculum development. The current study examines the first three roles in depth and will draw out perceived commonalities and differences.

The range of data-related roles is reflected in the breadth and depth of the data curriculum, since its scope should support the specific requirements of each role. The position, function and value of iSchools in developing data workforce capability and building capacity has been noted [18]. Data-related graduate programs, certificates and courses are already provided by some institutions [19] [20]. Two new graduate courses (Research Data Management and Research Data Infrastructures) have been designed and delivered at the University of Pittsburgh iSchool, alongside data mining, data analytics and data visualization classes and an Advanced Certificate in Big Data. There are also a range of data management training programs positioned towards up-skilling existing library and information professionals e.g. RDMRose [21], immersiveInformatics [22], MANTRA DIY Kit [23]. A recent review of digital curation education and training, notes the development of a Research Data Management MOOC (Massively Open Online Course) by UNC-Chapel Hill in 2015 [24].

There is a growing body of work addressing the core skills, competencies and training requirements for digital curation and research data management. These initiatives have been variously framed as Data Information Literacy (DIL) [25] [26] and Data Management Skills Support Initiative (DaMSSI) [27]. Twelve core competencies were identified in the DIL Project: Data Processing and Analysis, Data Management and Organization, Data Preservation, Database and Data Formats, Ethics and Attribution, Data Quality and Documentation, Data Curation and Reuse, Data Conversion and Interoperability, Data Visualization and Representation, Discovery and Acquisition, Metadata and Data Description, Cultures of Practice.

The term "*translational data science*" was introduced by Lyon and Brenner [18] to describe the transition of data skills, software tools and research intelligence from the iSchool to the marketplace. This characterization is particularly relevant to the development of the broad data science curriculum, which aims to equip graduates for new community practice roles in a range of disciplines, organizations and sectors. The implications of a translational approach to the design of training programs in the clinical sciences has been described which highlights the need to understand complementary disciplines and to become immersed in (clinical) practice [28]. These requirements have resonance for data scientists of all flavors, who must combine a portfolio of data informatics skills and competencies with disciplinary knowledge and practice. An immersive approach to research data skills development has been adopted in the immersive Informatics program [22], in the clinical setting [29] and at the University of Pittsburgh [30], where students spend time in the research laboratory. A similar model has been implemented at the University of Illinois Urbana-Champaign, with an intern practicum located in a data center [31].

4. METHODOLOGY

In this small-scale study, we sought to create a transferable methodology that faculty at iSchools may use to examine and review their existing curriculum in order to ready their students for future translational (market-driven/real-life) data preservation and data curation roles.

We selected the three data-related preservation roles outlined by Lyon and Brenner [18] to provide a focus for an analysis of employer requirements: Data Librarian, Data Steward/Curator, and Data Archivist. We searched for recent job postings (published from January 2014 to April 2015) that were of "semantic equivalence" to these roles, using five job banks to locate the positions: indeed.com; *The Chronicle of Higher Education's Vitae;* ALA JobLIST; www.jobs.ac.uk; and the IASSIST Jobs Repository. These job postings are listed in Table 1. We selected these job banks based on both breadth and on tailored focus. We anticipated that indeed.com, a search engine that aggregates listings from multiple job sites, would offer breadth. The academic and library employment sites (*Vitae* and www.jobs.ac.uk) would enable us to search for positions within institutions of higher education and libraries, both of which we anticipated would be major employers for these data curation roles. Finally, IASSIST is an international organization of information professionals focused on social sciences and data; we selected the Jobs Repository because of its narrow and relevant scope.

We used keyword searching for the job banks. In the instance of IASSIST Jobs Repository, the volume of positions published in our studied time frame allowed us to visually scan the job titles for relevance. For our analysis, we aimed to locate ten job positions for each data-preservation role. We were successful in doing so for the Data Librarian and Data Steward positions but discovered a paucity of positions with "Data Archivist" as a job title. In this case, it was necessary to broaden our search and analyze positions that fell outside of the January 2014 to April 2015 time frame, and we drew upon the IASSIST Job Repository, which includes postings from 2005 to present. Even with this resource, it was necessary for us to include in our analysis one position that we read as "archival" in nature and that was located at a data archive, despite the absence of lexical equivalence (that is, the titles were different but the nature of the work similar).[1]

While we were able to access the full job descriptions for the more current positions, the IASSIST postings offered a more abbreviated job advertisement. In the cases in which a URL was available in the IASSIST job advertisements, we attempted to use the WayBack Machine to locate the original job descriptions. We found that while we could access the institutions' human resources websites, the job descriptions were not indexed.

We performed a content analysis on the job descriptions (and, when necessary, job advertisements) for our suite of job positions to identify patterns in employers' requirements for job candidates. We

[1] It should be noted that non-traditional and traditional archive positions are seeing an explosion of new names and classifications, including "digital asset managers," "digital content specialists," "digital services technician," "supervisory IT specialists." In each of these jobs people are responsible for preserving, describing and providing access to data sets at different scales. Forthcoming work by Acker will highlight these changes.

developed a coding scheme that examined five categories:

- *Education:* Academic qualifications
- *Experience*: direct, hands-on practice
- *Knowledge*: understanding of/familiarity with topics/subjects/issues
- *Skills*: ability to do an action well
- *Competencies*: proficiency with specific tools/technologies/programming languages.

For each of the three roles studied in this paper, we sorted requirements articulated in the located job positions within these five pre-set categories. Having done this, we looked for patterns within the requirements that cut across the positions. For example, having grouped required technological "Competencies," we drew specifications that candidates be proficient with Microsoft Access, with MySQL, and Oracle and coded these specifications as "competence with relational database systems" (see Table 3).

While we made note of whether the employers characterized the education, knowledge, experience, skills, and competencies as essential or desirable requirements, we looked for patterns in the coding irrespective of this classification. In doing so, we could assess how the curriculum would best prepare iSchool students for employer consideration for the three data-preservation roles.

We identified all requirements that appeared in at least two of the positions studied for each role and designated these as "Key Requirements". From here, we analyzed course syllabi in the University of Pittsburgh School of Information Sciences graduate Library and Information Sciences (MLIS) program to determine relevant courses offered which would support the required and desired education, experience, knowledge, skills, and competencies. In doing so, we focused on the course description, objectives, and topics outlined in the syllabus. We then identified course topic gaps and opportunities for partnerships, both internal and external to the School of Information Sciences at Pitt.

As a part of the process, we explored model pathways to the data-preservation roles of Data Librarian, Data Steward/Curator, and Data Archivist, based on the current (2014-2015) curriculum. We also identified ways in which the School of Information Sciences could enhance its preparation for students to meet the expectations of employers seeking applicants for these positions.

Table 1. Job postings examined [28]

Job Postings		
Data Librarian	**Data Steward/Curator**	**Data Archivist**
Data & Visualization Librarian (Dartmouth College Library)	Clinical Data Curator (UnitedHealthGroup)	Collections Development Officer (UK Data Archive)
Data Acquisitions Librarian (The Federal Reserve Board)	Data Steward (InTec, LLC)	Data Archivist (University of Chicago's Center for the Economics of Human Development)
Data Services and Collections Librarian (UC San Diego Library)	Data Curator (DST Systems)	Data Archivist (UC DATA at UC Berkeley) – *past position*
Data Services Librarian (New York University Libraries)	Data Stewardship Coordinator (Stanford University)	Data Archivist (Social and Economic Survey Research Institute, Qatar) – *past position*
Data Services Resident Librarian (The University of Chicago Library)	Data Steward Consultant (Allstate)	
Research Data Curation Librarian (University of Michigan)	Data Steward (University of Virginia)	
Research Data Services (Contract) Librarian (University of New Hampshire Library)	Data Curation Specialist (University of Illinois Urbana-Champaign)	
Research Data Services Librarian (Cornell University)	Data Curator (New York University)	
Research Data Specialist (Purdue University Libraries)	Knowledge and Data Curation Specialist (Cornell University)	
Social Sciences Data Librarian (The University of Texas at Arlington)	Scientific Data Curator (Broad Institute of Harvard University and MIT)	

5. RESULTS

This section presents the analysis of employer-specified job requirements for the selected roles of Data Librarian, Data Steward/Curator, and Data Archivist. Two perspectives are drawn out: a) the common requirements across the three roles and b) the requirements unique to each role. This analysis is followed by the development of specific data-centric model pathways for each role, based on the analysis of job posting requirements. We draw upon the current (2014-2015 academic year) course portfolio of the School of Information Sciences at the University of Pittsburgh.

This analysis is followed by development of specific data-centric model pathways, composed of course "stepping stones." These pathways were developed based on the analysis of job posting requirements and through a review of the current (2014-2015 AY) iSchool course portfolio for Library and Information Sciences. We extended our review to include a consideration of courses in the Information Sciences program at the iSchool at Pitt and in other units on campus to consider stepping stones that may exist outside of the MLIS program as it stands.

The analysis of Key Requirements for each of the three roles (Data Librarian, Data Steward/Curator and Data Archivist) are presented in Tables 2-4. Of these Key Requirements, four were required by all the roles: a) Experience or knowledge or understanding of the researcher perspective, b) Knowledge of metadata standards and schema for data, c) Competence with statistical / analysis software packages and d) Knowledge of disciplinary data.

5.1 Data Librarian

The Data Librarian jobs invite candidates with, at minimum, a graduate degree from an ALA-accredited library and information science program (or an equivalent degree). For tenure-stream faculty librarian positions examined, there was a desire for applicants with a second graduate degree. Notable in the narrative

in the job adverts is an interest in candidates who understand the researcher perspective from their experiences as researchers and who are committed to user-centered library services and resources.

Table 2. Key Requirements for Data Librarian

Data Librarian				
Education	**Experience**	**Knowledge**	**Skills**	**Competencies**
ALA-accredited degree in library and /or information science ALA-accredited degree in library and/or information science **or** advanced degree in relevant discipline ALA-accredited accredited degree in library and/or information science **and** a graduate degree in relevant discipline	Experience conducting qualitative and/or quantitative research Experience designing and delivering RDM training and outreach Experience delivering RDM consultation support Experience assessing user data needs and designing RDM services in response Experience acquiring data resources for a library collection	Knowledge of RDM activities and roles throughout research lifecycle Knowledge of RDM trends/current research, particularly in academic setting Knowledge of metadata standards for data discovery and preservation Knowledge of sources for locating and depositing disciplinary data Knowledge of funders' data management requirements	Ability to work well in collaborative teams Strong oral, written, and interpersonal communication skills Project management effectiveness Analytical and organizational skills	Competence with qualitative and quantitative analysis software packages (e.g. Atlas.ti, NVivo, SPSS, R) Competence with programming languages, (e.g. JavaScript, Python, and PHP) Competence with GIS software Competence with visualization tools

The Data Librarian positions have a unique focus on a) knowledge of research funding agency data management requirements, b) knowledge of RDM activities and roles throughout the research lifecycle and c) experience of designing and delivering research data management (RDM) training and outreach (Table 2). Navigating data-centric model pathways through the current MLIS course portfolio, we propose the following course "stepping stones" for prospective Data Librarians. Together, these stepping stones form curricular pathways.

Essential course stepping stones for a prospective Data Librarian will include:

- *Research Data Management*
- *Research Data Infrastructures*
- *Metadata*
- *Academic Libraries*
- *Preserving Digital Collections*
- *Research methods in LIS.*

Desirable course stepping stones will include:

- *Intro to Information Technologies*
- *Managing & Leading Information Services*
- *Digital Repositories (new course already in development)*
- *GIS for Librarians*
- *Information Visualization.*

Course development and collaborative partnership opportunities have been identified:

- *Programming for Librarians (new course already in development)*
- *Intro to Statistical Data Analysis / Data Analytics* (from Graduate Program in Information Science & Technology IST colleagues within the School).

5.1 Data Steward/Data Curator

The Data Steward/Data Curator positions invited applications from individuals with a much wider range of disciplinary training. In addition to information science or library and information science degrees with course work in data modeling and metadata, employers were interested in candidates with computer science, mathematics, and business-related qualifications.

We identified a trend in job titles through our data collection. A search of indeed.com on March 30, 2015, produced 262 results that included the string "data steward." Conversely, there were only five results for "data curator," suggesting that the former, in the United States, is the more common position descriptor.

The Data Steward/Data Curator positions have a unique focus on a) experience of data governance and b) knowledge of data quality assurance practices and c) competency with relational database systems (Table 3).

Our search of the job banks involved using "data steward" and "data curator" as our search terms. What returned to us were positions that were both in the corporate and academic sectors. In the case of the former, these are positions that are data-centric but where the data is more likely to be used for internal research and compliance within the organization.

Table 3. Key Requirements for Data Steward/Curator Positions

Data Steward/Curator				
Education	**Experience**	**Knowledge**	**Skills**	**Competencies**
Unspecified Bachelor's degree Bachelor's degree in discipline relevant to data that is at the focus of work (i.e. health sciences and biological) Bachelor's degree in an "analytical" major such as math, business, computer science Advanced degree in informatics-related field	Experience analyzing and understanding data as a researcher Experience with metadata schemas, structures, and standards Experience with data governance	Knowledge of data management and quality assurance practices Knowledge of metadata schemas and ontologies Knowledge of data governance Knowledge in discipline relevant to data Knowledge of database structure and development	Ability to work effectively in collaborative teams Oral, written, and interpersonal communication skills Ability to communicate effectively with researchers from a variety of disciplines and backgrounds Ability to learn new technologies quickly and to adapt to change Analytical and organizational skills	Competence with relational database systems (e.g. Microsoft Access; MySQL) Competence with Microsoft Excel Competence with data visualization tools Competence with web authoring tools, Drupal

Essential course stepping stones for a prospective Data Steward/Curator will include:

- *Metadata*
- *Research Data Management*
- *Research Data Infrastructures*
- *Information Storage & Retrieval*
- *Digital Repositories (new course already in development)*
- *Preserving Digital Collections*

- *Information Architecture*
- *Corporate knowledge practices*
- *Database Management.*

Desirable course stepping stones will include:

- *Information Security & Privacy*
- *Data Structures*
- *Advanced Topics in Database Management*
- *Information Visualization*
- *Foundations of clinical & public health informatics (if interested in stewardship positions in health)*
- *Digital Curation.*

Course development and collaborative partnership opportunities have been identified:

- *Programming for Librarians (new course already in development)*
- *Data Governance (with Business School or School of Law).*

5.2 Data Archivist

We found a scarcity in current "data archivist" positions; as a result we were only able to code a small set of employer requirements for positions titled "data archivist" and with data archival responsibilities. This is probed further in our Discussion.

In addition, analysis revealed that the term "digital archivist" was out of scope to our analysis. The current job positions with this title that we located were records-focused and did not include any explicit mention of data as an information object under the purview of the candidate. There is, of course, an argument to be made that data is meaningful documentation for research and that, as such, all archivists are data archivists. For the purposes of this paper, we were primarily focused on structured data and as such did not cast our net to include positions without explicit allusion to this.

The Data Archivist positions have a unique focus on a) Experience of data documentation, b) Experience of data preparation and c) Knowledge of how to integrate diverse data resources (Table 4).

Table 4. Key Requirements for Data Archivist Positions

Data Archivist				
Education	**Experience**	**Knowledge**	**Skills**	**Competencies**
Bachelor's degree in discipline relevant to data that is at the focus of work Master's degree in discipline relevant to data that is at the focus of work	Experience creating data documentation Experience with collection development Experience with of data preparation and processing activities Experience using/analyzing data relevant to position	Knowledge data applicable to position and data use in relevant research Knowledge of data collection procedures Knowledge of metadata standards and documentation for datasets Knowledge of how to integrate diverse data resources	Ability to work well in collaborative teams Strong oral, written, and interpersonal communication skills Attention to detail Analytical and organizational skills	Competence statistical software packages (e.g. SPSS, Stata, SAS, R) Competence with Microsoft Office Competence with web authoring tools

Essential course stepping stones for a prospective Data Archivist will include:

- *Research Data Management*
- *Research Data Infrastructures*
- *Metadata*
- *Archives & Records Management*
- *Archival Appraisal*
- *Library & Archival Computing*
- *Preserving Information*
- *Preserving Digital Collections*

Desirable course stepping stones will include:

- *Access Systems, Standards, and Tools*
- *Digital Repositories (new course already in development)*
- *Preserving Digital Culture (course looking at historical development of digital media and theory of digital preservation)*
- *Digital Curation*
- *Information Architecture*
- *Web archiving*

Course development opportunities may encompass:

- *Intro to Statistical Data Analysis / Data Analytics* (from Graduate Program in Information Science & Technology IST colleagues within the School).
- *User experience and systems evaluation.*

6. DISCUSSION

This methodology utilizes the textual analysis of job postings for the three specific data roles, and has proved to be effective in revealing the particular qualifications, experience, knowledge, skills and competency requirements which employers are seeking from graduate students. By using only recent job descriptions, we are seeing the current perspective from the marketplace, where a wide range of organizations are recruiting to fill new data-centric positions. The associated mappings to graduate curriculum components gives an indication of the scope and contribution of the current course portfolio, and its relevance to a translational / market-driven data science environment.

6.1 Comparing the data roles

Previous analyses of data curation / digital curation job descriptions have highlighted a range of job titles with "archivist" featuring as a frequent term [32], great variation in job duties with corporate or research organizations more likely to require domain expertise, most often in science [33]. Job analyses have been used to determine health science and science and technology librarians' competencies for data management [34] and a set of digital curation competencies [35].

The sample size of jobs analyzed in this study was relatively small compared to the previous studies, but was targeted at very specific roles taking a "snapshot" approach. Our results suggested some common requirements of employers across the three job types. Employers were seeking graduates with experience or understanding of the researcher perspective; this knowledge could be attained by carrying out research as a doctoral student or by working closely with a research team, and indirectly implies demonstrating domain expertise. This requirement also emphasizes the value of immersive or embedded sessions or practicums in a research laboratory or other research environment within the curriculum, where the research lifecycle can be observed first-hand through bench science, experimental workflows, field work and the day-to-day perspectives and motivations of researchers exposed.

The common requirement for knowledge of disciplinary data may be related to acquiring research experience in a particular domain. Disciplinary perspectives can also be attained through immersive

classes, skills labs, intensive hands on practicums, or internships [31]. Knowledge of metadata standards was an additional common requirement of all three roles and reflects the need for structured data descriptions using established domain schema wherever possible. An understanding of metadata issues also addresses the need for curators to understand the effort/cost : benefit balance for producers (creators) and consumers (re-users) of data [36].

The final common requirement was competence with statistical and/or analysis software packages such as SPSS, R, Excel and Stata. This is not particularly unique to data science roles or an absent requirement at many iSchools; there are indeed many iSchool programs that already require quantitative skills, including statistics and programming. The results from this study rather demonstrate the continuing relevance of quantitative skills in iSchool graduate students.

Each of the three data roles sought unique requirements for applicants. The Data Librarian roles required ALA-accredited qualifications, with selected requirements for a higher degree; there was clear weight given to evidence of educational trajectories towards careers in library and information work. The Data Librarian roles were unique in stating a need for knowledge of research funding agency data management requirements. This links to the documented development of Research Data Services in academic libraries which focus on providing advocacy on funder policy e.g. US National Science Foundation, National Institutes of Health policy statements for data management or data sharing plans as components of research proposals. The development of data management planning (DMP) tools such as the DMPTool[2] and DMPOnline[3] has enabled libraries to provide consultation and training services as elements within designated "data librarian" roles.

A further unique requirement was a stated focus on RDM activities throughout the research lifecycle. Whilst there are many representations of the research data lifecycle [37], [38], at each stage there are interventions where a data librarian can make a positive contribution e.g. recommending a metadata schema for data description, promoting an established and trusted repository for data deposit and assisting with data identification and citation processes. Therefore a thorough understanding of the whole data lifecycle opens up opportunities for data librarians to craft new data services to support the research community.

The final unique component for data librarians was an emphasis on RDM training and outreach. Academic libraries have long established working relationships with faculty and (graduate) students in departments and schools; frequently this relationship is enacted through liaison / faculty / subject librarians who develop and deliver outreach and training on aspects of information literacy, e-journals, open access publications etc. There is now a significant need for scaling up advocacy and outreach for the many components of RDM; some academic libraries are developing new Research Services portfolios which bring together a mix of novel RDM and digital scholarship offerings delivered by data librarians and others, working in newly sculpted research support teams. This trend may be accompanied by organizational restructuring to optimize resources, service functions and communications.

Reviewing Data Steward / Curator positions, we found some commonalities with the other roles, but also some unique features. These roles sought applicants from a wider range of disciplines and backgrounds. The reference to computer science, mathematics and business-related qualifications positioned these roles more closely with Data Engineer and Data Analyst roles. The trend towards using the term "data steward" perhaps reflects the strong profile of the National Agenda for Digital Stewardship promulgated by the National Digital Stewardship Alliance [39].

The Data Steward / Curator roles also demonstrated unique components: an emphasis on data governance which recognizes the importance of intellectual property rights and other legislative issues associated with data sharing and compliance with open data policy aspirations at both national and institutional levels. These roles also required a knowledge of data quality assurance practices, which reflect the key role of data curators in data selection, appraisal and cleansing workflows as critical elements of data ingest into (trusted) repositories and data centers. The third unique requirement was competency with relational database systems; many large-scale datasets are stored in very large and complex database systems with hundreds of columns and many rows. The ability to import, manipulate, export and manage data in such complex systems is essential in the era of "big data".

Our search for recent positions titled "Data Archivist" was challenging, with a significant lack observed in the particular job banks we trawled. We can speculate that whilst the archives community is currently recruiting digital archivists, in most cases, these roles draw on long-established traditions and terminology, and are not yet explicitly framed around data. This does not mean that these positions are not data-focused (primary resources are arguably data), but the absence of reference to datasets placed it outside of our methodology. A major finding of this study is that "Data Archivist" is an uncommon job title today. This points to the fact that archivist roles are being renamed and reclassified in different job sectors.

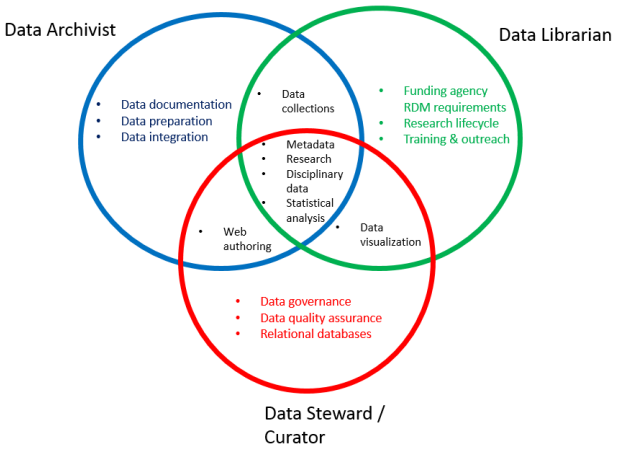
Considering the unique elements of the Data Archivist positions that were located, we identified an emphasis on "data documentation", "data preparation" and "data integration", which add weight to the assertion that established archival principles are reflected in the language which is applied to new digital objects of record i.e. research datasets. It can be argued that there are some similarities in the requirements for Data Librarian and Data Archivist; the emphasis on "data collections" is a common theme which once again reflects the long-established foundations of these fields.

Looking across the three roles, in general the requirements support the categorization and role descriptions of Lyon & Brenner [18], with some equivalence with the Data Librarian and Data Manager roles (the latter possibly equivalent to Data Steward/Curator) described by Pryor and Donnelly [19]. There are particular common Key Requirements across pairs of roles. The Data Archivist and Data Librarian positions emphasized experience related to collections (possibly demonstrating common foundational principles); the Data Archivist and Data Steward / Curator roles featured Web authoring competencies and the Data Librarian and Data Steward/Curator roles required competency with (Data) Visualization tools. The commonalities and unique aspects we observed are summarized in a Data Roles and Requirements Venn Diagram in Figure 1.

Figure 1. Data Roles and Requirements Venn Diagram

[2] DMPTool https://dmp.cdlib.org/

[3] DMPOnline https://dmponline.dcc.ac.uk/



6.2 iSchool Curriculum development

In reviewing our current curriculum, we are adopting a "model pathway" approach to reflect the optimal mix of courses (stepping stones) that a graduate student should/could take to follow specific career trajectories. This navigational process is primarily intended to guide the prospective student, but also serves to highlight potential opportunities to strengthen, broaden and extend the curriculum to support the breadth of data science roles described by Lyon and Brenner [17]. Whilst in some cases a truly radical re-engineering of the curriculum may be appropriate, rather we are adopting the approach of navigating the curriculum in new ways to signpost and showcase primary stepping stones (i.e. courses) to these emerging data science roles.

However given the specific requirements for the three roles explored in this small-scale study, we do suggest that the absence of any data-centric courses would be a perceived gap in an iSchool Library and Information Science curriculum at this time. Furthermore, it is clear that selected new elements are needed to fully meet the expectations of employers seeking data talent. These components may be acquired from other internal iSchool programs such as Information Science & Technologies at Pittsburgh, or from particular external sources e.g. other Schools and Departments. The current focus on "data" opens up many opportunities for new and exciting collaborations and partnerships in curriculum development.

The observed emphasis on disciplinary knowledge and experience may also be addressed through partnerships. For example, a new joint appointment with the Department of Biomedical Informatics paves the way for new focused offerings on text mining and data extraction, ontology development and knowledge organization systems (KOS). The diversity in disciplinary data practice is exemplified by the plethora of standards, schema, formats and cultures in different domains. As educators developing the data curriculum, ensuring graduate student expertise in all these fields is challenging; some would say impossible. Our approach is to aim for balancing these poles (knowledge of all data domains versus knowledge of none), within the curriculum through a mix of RDM courses which heavily feature case studies and domain exemplars, immersive sessions in the research laboratory and discipline/data-type-specific courses e.g. GIS for Librarians, Health Informatics. The results of this study validate the embedded / immersive / practicum components of data courses, since employers have stated their desire for applicants to demonstrate an understanding of research practice and disciplinary expertise in job postings.

The study also highlights the need to upskill existing practitioners as well as to produce data-savvy and work-ready graduates. Cohorts in the RDM and RDI courses have included practising librarians from both the Pittsburgh University Library System and Health Libraries System, and from Carnegie Mellon University Libraries. In this way, capacity and data capability is being scaled-up to meet the growing demand for Research Services. We hope to see a similar trend with graduates seeking careers within archival science and practicing archivists joining RDM/RDI courses.

This study raises some more general implications for recruitment strategies to graduate Library and information science courses. The employer demand for research experience, disciplinary knowledge and data analysis competencies, highlights the need for LIS programs to review their recruitment base to include STEM graduates who have strong technical and quantitative skills, and are happy manipulating tabular data or performing statistical analyses of datasets using a software package such as R or SPSS.

7. Conclusions

Our study has demonstrated key commonalities and distinct differences between the three data roles investigated. We acknowledge that the work is relatively small in scale and has a strong US focus, but the results indicate helpful directions for developing the iSchool curriculum to help to fill the data "talent gap" [13]. The translational data science approach adopted in the methodology (from iSchool to marketplace), reflects the trends of employers across sectors who are seeking data-savvy and work-ready graduates to fill these different data roles. Finally, we believe there is a great opportunity for iSchools to develop and extend their curriculum to embrace additional data-centric programs, courses and certificates to both educate new-entrants and to upskill existing practitioners to achieve the data-savvy profile, which is currently in high demand.

8. REFERENCES

[1] Woolf, S. H. (2008) The Meaning of Translational Research and Why It Matters. JAMA 299 (2), 211-213, Accessed April 13, 2015 http://jama.jamanetwork.com/article.aspx?articleid=1149350

[2] Carlson, S. 2006. Lost in a Sea of Science Data. *The Chronicle of Higher Education* (June 23). https://chronicle.com/article/Lost-in-a-Sea-of-Science-Data/9136

[3] Hey, T. and Hey. J. 2006. E-Science and its implications for the library community. *Library Hi Tech*, 24(4), 515-528. http://www.emeraldinsight.com/doi/pdfplus/10.1108/07378830610715383

[4] Macdonald, S. and Martinez-Uribe, L. 2008. Data librarianship – a gap in the market. *CILIP Update*, 7(6), 20-21. https://www.era.lib.ed.ac.uk/bitstream/handle/1842/2499/Gap%20in%20the%20market.pdf?sequence=3&isAllowed=y

[5] Corrall, S. 2012. Roles and responsibilities: libraries, librarians and data. In Pryor, G. (Ed.) Managing Research Data. Facet Publishing. pp105-133.

[6] Lyon, L. 2012. The Informatics Transform: Re-engineering Libraries for the Data Decade. *IJDC* 7(1), 126-138. http://www.ijdc.net/index.php/ijdc/article/view/210/279

[7] Jaguszewski, J. M. and Williams. K. 2013. New roles for new times: transforming liaison roles in research libraries. ARL Report. http://www.arl.org/storage/documents/publications/nrnt-liaison-roles-revised.pdf

[8] Auckland, M. 2012. Re-skilling for Research. Research Libraries UK (RLUK) Report. http://www.rluk.ac.uk/wp-content/uploads/2014/02/RLUK-Re-skilling.pdf

[9] Cox, A. 2012. Upskilling Liaison Librarians for research Data Management. *Ariadne* (6 December) http://www.ariadne.ac.uk/print/issue70/cox-et-al

[10] Tenopir, C., Birch B. and Allard, S. 2012. Academic libraries and Research Data Services: Current Practices and Plans for the Future. ACRL White Paper. http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/Tenopir_Birch_Allard.pdf

[11] Cox, A. and Pinfield. S. 2014. JoLIS 46(4), 299-316. http://lis.sagepub.com/content/46/4/299.full.pdf+html

[12] HM Government 2013. Seizing the data opportunity. A strategy for UK data capability. White Paper. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/254136/bis-13-1250-strategy-for-uk-data-capability-v4.pdf

[13] Bakhshi, H., Mateos-Garcia J and Whitby, A. 2014. Model workers: How leading companies are recruiting and managing their data talent. Nesta Report. http://www.nesta.org.uk/sites/default/files/model_workers_web_2.pdf

[14] Lyon, L. 2007. Dealing with Data: Roles, Rights, Responsibilities and Relationships. Consultancy Report. http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/reports/dealing_with_data_report-final.pdf

[15] Swan, A. and 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. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.147.8960&rep=rep1&type=pdf

[16] Larsen, R. 2014. What can we learn from the data? Preparing the workforce for digital curation: The iSchool perspective. International Digital Curation Conference. Presentation available from http://www.dcc.ac.uk/sites/default/files/documents/IDCC14/Panels/RonLarsen_Panel.pdf

[17] Varvel, V. E. et al 2010. Report from the Research Data Workforce Summit. https://www.ideals.illinois.edu/bitstream/handle/2142/25830/RDWS_Report_Final.pdf

[18] Lyon, L. and Brenner, A. 2015. Bridging the Data Talent Gap – positioning the iSchool as an Agent for Change. *IJDC* 10(1), 111-122. http://www.ijdc.net/index.php/ijdc/article/view/10.1.111/384

[19] Pryor, G. and Donnelly, M. 2009. Skilling up to do data: Whose role, whose responsibility, whose career? *IJDC* 4(2), 158-170. http://www.ijdc.net/index.php/ijdc/article/view/126/133

[20] Creamer, A.T., Morales, M.E., and Kafel, D. 2012. A sample of Research Data Curation and Management Courses. *J. eScience Librarianship* 1(2), 88-96. http://escholarship.umassmed.edu/cgi/viewcontent.cgi?article=1016&context=jeslib

[21] Cox, A., Verban, E. & Sen, B. (2014) A Spider, an Octopus or an animal just coming into existence? Designing a curriculum for librarians to support research data management. *J eScience Librarianship* 3(1) 15-30, accessed April 15, 2015. http://escholarship.umassmed.edu/jeslib/vol3/iss1/2/

[22] Shadbolt, A., Konstantelos, L., Lyon, L. and Guy, M. 2014. *IJDC* 9(1), 313-323. http://www.ijdc.net/index.php/ijdc/article/view/9.1.313/360

[23] Macdonald, S. and Rice, R. 2012. "DIY" research Data management Training Kit for Librarians. Available from http://ceur-ws.org/Vol-1016/paper27.pdf

[24] Tibbo, H. R. 2015. Digital curation education and training: From digitzation to graduate curricula to MOOCs. *IJDC* 10(1) 144-153. http://www.ijdc.net/index.php/ijdc/article/view/10.1.144/387

[25] Carlson, J. R. et al 2011. Determining Data information literacy needs: a study of students and research faculty. *Portal: Libraries and the Academy* 11(2), 629-657. http://muse.jhu.edu/journals/portal_libraries_and_the_academy/v011/11.2.carlson.pdf

[26] Carlson, J. et al, 2013. Developing an Approach for data management education: a report from the Data Information Literacy Project. *IJDC* 8(1) 204-217. http://www.ijdc.net/index.php/ijdc/article/view/8.1.204/306

[27] Molloy, L. and Snow, K. 2011. DaMSSI Project Final Report. Available from http://www.academia.edu/2808837/DaMSSI_Data_Management_Skills_Support_Initiative_Final_Report

[28] Rubio, D. M. et al 2010. Defining Translational research: Implications for Training. *Acad Med*. 85(3) 470-475. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2829707/

[29] Lyon, L. and Webster, K. 2014. Embedding immersive informatics research data management within the iSchool curriculum: a laboratory-based action research Case Study. *Library Research Seminar VI,* University of Illinois Urbana-Champaign. Abstract available at http://www.library.illinois.edu/lrs6/Library_Research_Seminar_VI_Program.pdf

[30] Martin, E. R. (2013) Highlighting the Informationist as a data librarian embedded in a research team. (Editorial) *J eScience Librarianship* 2(1) 1-2, accessed April 13, 2015. http://escholarship.umassmed.edu/cgi/viewcontent.cgi?article=1044&context=jeslib

[31] Mayernik, M. S. et al 2015. Enriching education with exemplars in Practice: Iterative development of data curation internships. *IJDC* 10(1) 123-134. http://www.ijdc.net/index.php/ijdc/article/view/10.1.123/385

[32] Lee, C. (2008) What do Job Postings indicate about Digital Curation Competencies? Presentation at the Society of American Archivists Research Forum, August 26, 2008, San Francisco, accessed April 13, 2015. http://ils.unc.edu/digccurr/digccurr-saa-research-forum-2008.pdf

[33] Cragin, M. H. et al (2009) Analyzing Data Curation Job Descriptions. Poster: 5th Int. Dig. Curation Conference, London, 2-4 December 2009, accessed April 13, 2015. https://www.ideals.illinois.edu/bitstream/handle/2142/14544/Cragin_poster_abstract_DCC_09.pdf?sequence=2

[34] Creamer, A., Morales, M & Crespo, J. et al (2012) An assessment of needed competencies to promote the data curation and management librarianship of health sciences and science and technology librarians in New England. *J of eScience Librarianship,* 1(1) 18-26. http://escholarship.umassmed.edu/cgi/viewcontent.cgi?article=1006&context=jeslib

[35] Kim, J., Warga, E. & Moen, W. (2013) Competencies required for digital curation: An analysis of Job Advertisements. IJDC 8(1) 66-83, accessed April 13, 2015. http://www.ijdc.net/index.php/ijdc/article/view/8.1.66

[36] Michener, W.K. et al (1997) Non-geospatial metadata for the Ecological Sciences. *Ecological Applications* 7(1), 330-342, accessed April 13, 2015. http://lits.bio.ic.ac.uk:8080/litsproject/Michenereta1997.pdf

[37] Higgins, S. (2012) The Lifecycle of data management. Chapter 2 in Ed. Graham Pryor. Managing Research Data Facet Publishing, 239pp.

[38] Corti, L, et al (2014) The Research Data Lifecycle. Chapter 2 in Managing and Sharing Research Data. A Guide to Good Practice. Sage Publications, 222pp.

[39] NDSA (2014) National Agenda for Digital Stewardship 2015 accessed April 13, 2015. http://www.digitalpreservation.gov/ndsa/documents/2015NationalAgenda.pdf