ISSN 2161-3974

JeSLIB 2022; 11(1): e1226
https://doi.org/10.7191/jeslib.2022.1226

Journal of eScience Librarianship
putting the pieces together: theory and practice

Full-Length Paper

Data Services Librarians' Responsibilities and Perspectives on Research Data Management

Bradley Wade Bishop[1], Ashley M. Orehek[2], Christopher Eaker[1], and Plato L. Smith[3]

[1] University of Tennessee, Knoxville, TN, USA
[2] Lindsey Wilson College, Columbia, KY, USA
[3] University of Florida, Gainesville, FL, USA

Abstract

This study of data services librarians is part of a series of studies examining the current roles and perspectives on Research Data Management (RDM) services in higher education. Reviewing current best practices provides insights into the role-based responsibilities for RDM services that data services librarians perform, as well as ways to improve and create new services to meet the needs of their respective university communities.

Correspondence: Bradley Wade Bishop: wade.bishop@utk.edu

Received: August 20, 2021 **Accepted**: February 7, 2022 **Published**: March 4, 2022

Data Availability: Deidentified data for this article can be found in the Tennessee Research and Creative Exchange (TRACE) institutional repository: Bishop, Bradley Wade. 2020. "Data from Data Services Librarians Study." [Data Set] University of Tennessee, Knoxville Libraries. https://doi.org/10.7290/m29yhy5qen

Disclosures: The authors report no conflict of interest.

JeSLIB 2022; 11(1): e1226
https://doi.org/10.7191/jeslib.2022.1226

Abstract Continued

Objectives: The objectives of this article are to provide the context of research data services through a review of past studies, explain how they informed this qualitative study, and provide the methods and results of the current study. This study provides an in-depth overview of the overall job responsibilities of data services librarians and as well as their perspectives on RDM through job analyses.

Methods: Job analysis interviews provide insight and context to the tasks employees do as described in their own words. Interviews with 10 data services librarians recruited from the top 10 public and top 10 private universities according to the 2020 Best National University Rankings in the US News and World Reports were asked 30 questions concerning their overall job tasks and perspectives on RDM. Five public and five private data services librarians were interviewed. The interviews were recorded and transcribed. The transcriptions were analyzed in NVivo using a grounded theory application of open, axial, and selective coding to generate categories and broad themes based on the responses using synonymous meanings.

Results: The results presented here provide the typical job tasks of data services librarians that include locating secondary data, reviewing data management plans (DMPs), conducting outreach, collaborating, and offering RDM training. Fewer data services librarians assisted with data curation or manage an institutional repository.

Discussion: The results indicate that there may be different types of data services librarians depending on the mix of responsibilities. Academic librarianship will benefit from further delineation of job titles using tasks while planning, advertising, hiring, and evaluating workers in this emerging area. There remain many other explorations needed to understand the challenges and opportunities for data services librarians related to RDM.

Conclusions: This article concludes with a proposed matrix of job tasks that indicates different types of data services librarians to inform further study. Future job descriptions, training, and education will all benefit from differentiating between the many associated research data services roles and with increased focus on research data greater specializations will emerge.

JeSLIB 2022; 11(1): e1226
https://doi.org/10.7191/jeslib.2022.1226

Introduction

Research Data Management (RDM) is a broad term that encompasses research data services throughout the research lifecycle from data creation to long-term preservation and dissemination (Lake et al. 2013). Data management plans (DMPs) are a key component of RDM and serve as formal documents describing the roles and activities for managing data during and after research. DMPs are now a required component of grant proposals by most funding agencies and across many countries. The valuable insights into how real-world data services librarians (DSLs) work to buttress the research enterprise through new and existing RDM tasks inform the design of future data policy, services, tools, education, and hiring practices.

As many domains move towards data-intensive transdisciplinary research, academic libraries have responded by offering RDM-related services. Early reports found that the majority of Association of College and Research Libraries (ACRL) were not offering research data services (RDS) but planned to offer RDM services in the next two years once they hired skilled staff or trained current staff to do so (Tenopir, Birch, and Allard 2012; Tenopir et al. 2015). Yet, follow-up surveys over the years from the same team using the same methodology found little increase in those RDS services reported at US academic libraries (Tenopir et al. 2019). In fact, 44 percent of academic libraries from the latest Tenopir et al. (2019) survey stated they are not involved in RDS. Follow-up interviews with some participants from the 2019 survey were conducted to discover the reasons for a lack of RDS growth. Participants acknowledged the library was not the resource viewed by faculty as best suited to fulfil these emerging RDM needs or that academic libraries simply were unable to hire librarians with the required skills to offer RDS services.

Additional research has explored the RDS that some academic libraries outside the US began offering beyond the US in response to the new paradigm in research sometimes referred to as Science 2.0, or Open Science, among others (European Commission 2015; Lord and Macdonald 2003). One survey of academic libraries in Australia, Canada, Germany, Ireland, Netherlands, New Zealand, United Kingdom, and the United States received 209 valid responses asking about RDS such as RDM instruction, managing a data repository, and carrying out long-term preservation (Cox et al. 2017). The academic libraries' role in offering these data-related services varied, but further analysis confirmed librarians have displayed leadership in planning and coordinating across campuses and that several academic libraries are positioned to lead the continued growth and maturity of these services (Kim 2019). In response to these emergent RDM needs, several academic libraries are hiring DSLs (Koltay 2019). With RDM needs growing, DSLs may take on a variety of responsibilities, such as locating secondary research data; reviewing DMPs; conducting outreach; collaborating on funded research projects; offering RDM training; assisting with data curation, dissemination, and preservation; and managing an institutional repository (IR) or digital repository (DR). In addition, other RDM tasks may include data security, data citation, and data sharing via deposit that involves issuing Digital Object Identifiers (DOIs) and necessitates the need for DSLs to have knowledge of copyright and licensing.

JeSLIB 2022; 11(1): e1226
https://doi.org/10.7191/jeslib.2022.1226

This study interviews those currently employed as DSLs doing the RDM work. The 10 interviews captured the overall job responsibilities, including tasks and duties related to several services at a more granular level than any survey could. The gap in the literature of detailed examples of DSL work is partially addressed by this study's qualitative approach. The following research questions guide the 30 interview questions:

1. What typical tasks relate to those in the role of a data services librarian?
2. How are research data services structured at institutions?
3. How are data management plans implemented, evaluated, and funded at institutions?
4. What research data management training are offered at institutions?
5. What education, training, and experience do data services librarians have?

The insight gained into the roles and responsibilities of the DSLs may provide any of those institutions not yet offering RDM or changing their current practices a few concrete examples of the breadth of duties DSLs deliver. The following overview of related research provides context on DSLs, their training, and DMPs.

Related Research

Data Discovery

The RDS or RDM-related librarian jobs emerged with wide-ranging responsibilities across the entire data lifecycle from finding data to planning data management to data analytics and visualization to curating data for long-term preservation. Locating data requires similar soft skills related to reference interviews about any information needed. A data reference interview may be more complex than traditional reference questions due to the complexity of data (i.e., one question leads to another) (Rice and Southall 2016). To help locate data, librarians need to have awareness of existing digital repositories (DRs) and institutional repositories (IRs) across disciplines and how to perform data discovery using them. Secondary data analyses have been the norm for many sciences where data are often collected only once in real-time at global scales requiring data sharing. Anecdotally, secondary data analysis use has increased in human subjects research as some data collection methods were reduced or greatly altered during the pandemic (e.g., direct observation). Sometimes locating data is not enough and users need help transposing data across formats or further assistance to access sensitive data. In reference, across all data services, a librarian's basic understanding of a community of practice's disciplinary norms and standards will help navigate these complex queries (Schmidt and Shearer 2016).

JeSLIB 2022; 11(1): e1226
https://doi.org/10.7191/jeslib.2022.1226

Data Management Plan Review

Another task for many DSLs relates to DMP review. DMPs are documents, which are often brief due to funding agency stipulations that describe the ways in which the researchers will manage research data created during and after their projects. These documents typically address topics such as file management, file types, backup and security, metadata, and sharing and accessing of data. DMPs or similar documents by other names (i.e., Data Sharing Plans) have been required by the US National Institutes of Health since 2011 for grants greater than $500,000 and the US National Science Foundation (NSF) since 2011 for all projects. Since then, almost all US federal agencies, most private foundations, and many other nations have implemented a DMP requirement for publicly funded research. DMPs and similar documentation will only grow in importance. For example, NIH has issued a new Policy for Data Management and Sharing (Notice Number: NOT-OD-21-013), which will require NIH funded researchers effective January 25, 2023 to prospectively submit a plan outlining how scientific data from their research will be managed and shared (NIH 2021).

The familiarity with disciplinary norms for DSLs must extend to funding requirements pre- and post-award. For example, there are several tools with DMP templates to help write DMPs for proposals from most funding agencies (e.g., https://dmptool.org). Still, things do change across agencies, directorates, and foundations, and the attention to detail of DSLs may be a valuable contribution during grant submission. This type of DMP review support requires that librarians stay up-to-date on these issues, including relevant data standards and processes (Cox and Verbaan 2018). DMP review and writing assistance only relates to DSL roles pre-award. To help post-award, a number of other research data services have emerged. A National Academies of Science, Engineering, and Medicine (2019) report recommends that NSF and other funders create code and data repositories that allow for the long-term preservation of digital artifacts. The Data Repository Guidance by scientific data is a good resource for DSLs (Nature 2021). DSLs may help build some of this cyberinfrastructure and assist researchers at other points later in the data lifecycle to help evaluate and implement DMPs.

Outreach and Collaboration

Like many academic librarians, outreach and engagement are common job responsibilities. Promoting current and new services to users, providing instruction, increasing partnerships with faculty, and integrating with courses and curriculum development have all been noted as types of outreach (Silver 2014). It may be difficult to operationalize what works and what does not work as outreach is contingent on each institutional context, organizational culture, and the roles and responsibilities of academic librarians at each institution.

Librarianship has historically been a support role driven by the information needs of each community being served. Many researchers across institutions of higher education and science agencies value the skills librarians and data managers bring

to the table, but they may not always come to mind when building a research team. That is changing as librarians are increasingly taking on more collaborative roles as integral parts of grant projects from proposal to conclusion. Librarians' roles range from traditional responsibilities such as literature searching to more "higher-end support" models such as being project managers and participating in project outputs (Brandenburg et al. 2017; Corrall 2014). In these situations, academic librarians are being included as personnel in grant budgets. Although examples of unfunded research collaborations were not found in the literature, assuredly these exist as well.

Data Visualization

One newer skill appearing in job descriptions that is unique to DSLs is data visualization (Ogier et al. 2018). A guide to good data visualization includes the following: (1) be clear on the question; (2) know your data and start with basic visualizations; (3) identify messages of the visualization, and generate the most informative indicator; (4) choose the right chart type, and (5) use color, size, scale, shapes, and labels to direct attention to the key messages (Lau and Pan 2015). The visualization of data requires that data adhere to metadata best practices, formats, and standards to enable description and accurate representation.

Data must be cleaned and prepared to enable visualization. Cleaning data includes resolving null values, enforcing input format rules (i.e. character, date, string), quality assurance and quality control (QA/QC), and removing duplicates. Fundamental steps such as *Nine simple ways to make it easier to (re)use your data* (White et al. 2013) are necessary during data preparation. "Data preparation is not just a first step, but must be repeated many times over the course of analysis as new problems come to light or new data is collected," as Hadley Wickham (2014, 1) pointed out. Software as Python (https://www.python.org), R (https://www.r-project.org), and scikit-learn (https://scikit-learn.org/stable)—Machine Learning in Python—along with graphical user interface (GUI) tools such as RStudio (https://rstudio.com) enable visualization of data, particularly tidy datasets. "Tidy datasets are easy to manipulate, model and visualise, and have a specific structure: each variable is a column, each observation is a row, and each type of observational unit is a table" (Wickham 2014, p 1). Training in Python, R, Spreadsheets, SQL, and Unix are necessary for DSLs to use many open source data visualization tools.

Kaggle (https://www.kaggle.com/learn/overview) and the Library Carpentry (https://librarycarpentry.org) are two examples of free online education resources to enable data visualization competencies. Additionally, learning how to leverage abstract and citation databases' application programming interfaces (APIs) to transform data into dynamic, interactive visualizations highlighting individual researcher activity (Mischo 2020, 1) is needed. Data visualization competencies include developing programmatic workflows and leveraging APIs (e.g. Scopus) to visualize aggregated data. A lot of resources exist for librarians to both learn and

teach data visualization with much of this work closely related to data curation more broadly.

Training

A common service many DSLs have implemented in the last several years is offering RDM training. These instruction sessions range in detail from general RDM that apply to any research data to discipline-specific skills. They also range in length from short, one-off sessions to term-length, for-credit courses (Carlson, Johnston and Westra 2015; Schmidt and Holles 2018). RDM training is offered at a number of academic libraries as well as some also delivering training related to Responsible Conduct of Research (RCR) (Gunderman 2021; Herr 2019). Select online data management training include:

- DataOne Data Management Skillbuilding Hub
  https://dataoneorg.github.io/Education
- ESIP Data Management Training (2020)
  https://dmtclearinghouse.esipfed.org
- ESIP Commons Data Management Short Courses for Scientists (2016)
  https://commons.esipfed.org/datamanagementshortcourse
- MANTRA Research Data Management Training (2020)
  https://mantra.edina.ac.uk
- Belmont Forum's e-Infrastructures and Data Management Toolkit
  https://bfe-inf.github.io/toolkit

Most disciplines, institutions, and data centers have tailored instructional materials.

Data Curation

Even though the origins of data curation stem from broader work in digital curation with its own long history influenced by many disciplines (i.e., archives, academic research libraries, data management, eScience, and Library and Information Science research communities), the scope and reach of data curation now extends out to assist all disciplines with their data. Data curation is a data lifecycle management process of providing descriptive, annotative, and representative information for research data through metadata (DCMI 2021; Bird et al. 2016). Data curation facilitates the adoption of existing metadata and format standards where appropriate to aid in the organization, access, discovery, and storage of data. Data curation can be applied to analog, digital, digitized, and born digital data existing offline or online.

Data management and curation (DMC) practices include four major data lifecycle management processes that:

1. Fulfill departmental, institutional, organizational policies & data management requirements;
2. Provide data creation (primary, secondary, tertiary data) (Lord and Macdonald 2003, p. 42), data publication, minimal data description;
3. Facilitate added value (metadata), management & storage of archived data over data lifecycle;
4. Integrate a series of technical & strategic actions and consultations to ensure continual data authenticity (Smith II 2014).

Data curation services in academic libraries provide technical infrastructure that supports RDM throughout the research lifecycle and might include metadata creation, persistent storage, and the assignment of unique identifiers (Johnston et al. 2018). RDM services (RDMS) include the aforementioned (1) DMP review; (2) managing active data; (3) data selection and handover; (4) data repositories; and (5) data catalogues interrelated processes enabled through training, guidance, support operating under overarching RDM policy/strategy embedded into higher education operations for sustainability (Jones, Pryor and Whyte 2013). RDMS vary in capacity, infrastructure, resources, and support across institutions, but demand is growing for them whether they are provided by academic libraries or not.

The Data Curation Network (https://datacurationnetwork.org) outlines the eight tasks that informed this study's interview questions on data curation, which include: (1) check files/code and read documentation; (2) understand the data (or try to); (3) request missing information or changes; (4) augment metadata for findability; (5) transform file formats for reuse; (6) evaluate for FAIRness; and (7) document all curation activities throughout the process. The DCC (2007) Curation Lifecycle Model which was reimagined during a presentation at the 15th International Digital Curation Conference in Ireland by Johns Hopkins University in 2020 is an exemplar model in which to further development of data curation efforts (Higgins 2008).

Managing an Institutional or Digital Repository

Many university libraries manage an IR which contains many types of materials. These services have been offered in academic libraries for years and they typically contain articles, posters, reports, theses, and dissertations. The inclusion of research data produced by researchers at their institution is a more recent development. Yoon and Schultz (2017) discovered that more than 60% of the libraries they surveyed offered data deposit services into their IRs. Options for archiving research data range from solutions designed specifically for data (e.g. HUBZero) to solutions designed primarily for textual materials (e.g. Dspace). Likewise, options for managing the repository range from using in-house staff to using a hosted service (Uzwyshyn 2016). DSL roles related to IRs and DRs range from awareness and helping researchers locate the appropriate repository for storing and perhaps sharing their data to assisting researchers with depositing in a library-managed IR, which includes all the data curation steps outlined above.

JeSLIB 2022; 11(1): e1226
https://doi.org/10.7191/jeslib.2022.1226

Emerging Skills and Workforce Development

One source of concern in the DSL landscape is the disconnect between the new skills needed to keep up in these evolving positions and what is being taught in training programs for academic libraries (i.e., iSchools). New data-intensive technical skills are becoming necessary to work with data in all sectors. Since many data services are staffed by librarians who never learned these skills in graduate school, they are unprepared to assist with large datasets and digital artifacts in an open, networked environment. Thus, several efforts have been developed to address this knowledge gap (Kirkwood 2016).

Two examples of these new skills training opportunities are the Library Carpentry series of workshops mentioned earlier as well as Data Carpentry, Software Carpentry, and many more data science training institutes for librarians. To help data librarians learn coding skills, the Carpentries provide hands-on workshops that teach specific coding and data management skills that librarians can put into use immediately (Atwood et al. 2019). Likewise, for data science and visualization skills, many academic librarians did not learn these skills in their library school curricula and the demand for these skills has increased in recent years. Two examples are the Data Science Training for Librarians program (http://www.dst4l.info) and Data Science and Visualization Institute (https://www.lib.ncsu.edu/data-science-and-visualization-institute).

As a few of these authors are current instructors for iSchool programs it is worth noting there have been and continue to be efforts to formalize education and credentials in the DSL arena. A 2012 survey of 52 library and information schools in the United States and Canada found that 16 offered courses of data curation (Harris-Pierce and Liu 2012). A more recent and exhaustive search through the curriculum and course content of the 123 iSchools in 2022 has not been done, but it is worth noting some of the US-based programs that have specialized in data services and curation and offer named programs such as the University of North Carolina at Chapel Hill's Professional Science Master's degree in Digital Curation and Management (https://sils.unc.edu/programs/graduate/digital-curation), or the free online Research Data Management Librarian Academy (RDMLA) created in collaboration with Simmons University and others (https://www.canvas.net/browse/simmonsu/courses/research-data-management), or University of Denver's RDM Concentration (https://morgridge.du.edu/academic-programs/library-information-science/mlis) or the University of Tennessee's RDM Certificate.

Each institution likely tracks the employment of their graduates and recruitment efforts for these specialized programs, but that level of granularity for LIS programs does not exist in the Association of Library and Information Science Education (ALISE) Statistical Report and Database (https://www.alise.org/statistical-report).

JeSLIB 2022; 11(1): e1226
https://doi.org/10.7191/jeslib.2022.1226

Methods

The study interviewed 10 current DSLs to address this study's research questions:

1. What typical tasks relate to those in the role of a data services librarian?
2. How are research data services structured at institutions?
3. How are data management plans implemented, evaluated, and funded at institutions?
4. What research data management training are offered at institutions?
5. What education, training, and experience do data services librarians have?

The interview questions included 30 open-ended questions about known tasks associated with these jobs. The study received Institutional Review Board (IRB)-approval. DSLs recruited were recruited via personalized email from the investigator from those identified at the top 10 public and top 10 private universities according to 2020 US News and World Reports (US News and World Report 2020).

The National Universities Rankings as a sampling frame were used because those top institutions emphasize faculty research as the result of large research expenditures and therefore are more likely to also have more researchers with RDM needs. All sampling skews results, but these institutions with the presumably highest demand for RDM services should have the resources and experience to inform best practices for other types of institutions. This study is part of a series of studies that includes the RDM roles of Research Integrity Officers (RIOs) (Bishop et al. 2021). Research Integrity Officers (RIOs) are a federally mandated position at any institution receiving U.S. Public Health Service funding. RIOs promote a Responsible Conduct of Research (RCR) environment, as well as inquire into and investigate potential research misconduct (Steneck 2007). RIOs are in charge of handling research misconduct allegations but spend a great deal of time on campuses promoting ethical scholarly communication practices within research institutions. To allow for a gap analysis between these two important roles on campuses may provide insights into the role-based responsibilities for RDM services, as well as ways to improve services to meet the needs of their respective university communities (Bishop et al., 2021). The same sampling frame of institutions were used for both studies. There are limitations to this qualitative approach, including participants' various experiences limited perspectives of the totality of RDM on their campuses, and these institutions do not represent the likely RDM efforts at many other types of institutions. Still, those 'top' institutions do tend to influence trends in services and resources expected at other academic libraries.

Some of these universities have multiple data librarians with various job titles, but at least one DSL from each institution was contacted. Five public and five private DSLs agreed to participate and were interviewed. The informed consent that participants agreed to included open data language with anonymized and deidentified transcripts that are now available in an IR. Informed consent was obtained from participants prior to their interviews.

The job responsibility questions were informed by the expertise of two DSLs at member institutions of the Association of Research Libraries and the literature reviewed in this paper. The job tasks of most DSLs may include RDM instruction, reviewing DMPs, and overall assessing and supporting campus data needs related to locating data and assisting with data curation. In some instances, these librarians may manage an IR, support open access initiatives, or even assist faculty on grants as dedicated data managers. The data curation section of the questions was informed by the Data Curation Network (https://datacurationnetwork.org). The full interview schedule of the 30 questions appears in Appendix 1.

The interviews were recorded and transcribed. The transcriptions were analyzed using NVivo. Categories and broad themes across responses to the questions emerged using terminology from the questions and responses. Grounded theory application of open, axial, and selective coding in NVivo captured their job tasks and perspectives on DMPs. This qualitative nature of this study allows us to capture the perspectives that quantitative studies cannot capture. These perspectives are meaningful as they provide insights into how these data services librarians feel about the work they do, rather than simply reporting which tasks they do. There are limitations of this study as DSLs' jobs differ across institutions due to historic positioning and status of academic libraries on each campus. In addition, different participants would have introduced a variety of responses, but these top institutions may show best practices with the resources of universities with large research expenditures.

Results

Overview of Responsibilities

To scope each participant's overall responsibilities, interviews began with questions about their primary job duties as DSLs from a list of known tasks, which are displayed in Table 1. As other tasks might not fit under those options, the participants were asked to list any other tasks not asked. Those 'other' responsibilities named include formal research and data management consultations, included coding and programming support to restructure data (n=3); teaching code for statistical analyses (e.g, Python, and R) (n=2); helping track down unavailable resources and data acquisition (n=2); and one participant each mentioned assisting researchers in their active data phases and determining workflows; helping with reproducibility; and performing software carpentry.

JeSLIB 2022; 11(1): e1226
https://doi.org/10.7191/jeslib.2022.1226

Table 1: Primary job duties as DSLs.

Responsibilities	Number of Participants
Assist with locating secondary research data	8
Data management plan (DMP) review	7
Outreach and collaboration	9
Research data management (RDM) training	9
Data curation	5
Manage an institutional or digital repository	1
Other	7

Following the overall job responsibilities section, DSLs were asked about the number of people working at their institution. Participants that mentioned their work entailed any of these job responsibilities were asked the more detailed questions about those aspects of the job. The number of people who worked in research data services at these ten participants' libraries ranged from one to approximately 22 staff members. Two worked alone, but most DSLs had at least one co-worker. The mean was a little more than 6.

To assess the coverage of discipline, departments, and data types, participants were asked more about their specific user groups. Five participants mentioned that they usually worked with specific subject disciplines such as social sciences (n=2); engineering, physics and other STEM-related disciplines (n=2); and one worked with life sciences; and data science and public policy. Of the five, two provided supports for the other disciplines that their libraries serve. All disciplines on campus were supported by the remaining five participants.

Five participants specifically noted working with data types like Tabular data and CSV (n=3), NVivo, GIS applications, and biomedical sequencing data. Four other participants stated their versatility in any data type or format. They all worked with various user populations and the list includes how many participants mentioned each—graduate students (n=9); faculty (n=7); staff (n=4); undergraduate students (n=4); post-doctoral students (n=3); and peer librarians (n=1).

JeSLIB 2022; 11(1): e1226
https://doi.org/10.7191/jeslib.2022.1226

Locating Data

Out of the ten, six participants stated they regularly located secondary research data. For example, Participant 7 identified it as their "primary responsibility with that title, data librarian. It's a little bit of a misnomer, I think, to call me [that]. I think I am a data-discovery librarian because I really only specialize in that one task." This participant was unique in explaining their job as mostly data reference (i.e., updating a LibGuide and helping locate data from common sources). Most of the other five participants explained that the data discovery work was only a portion of their positions. Three participants sometimes located data as Participant 6 described the work, "I mean I have done it, especially for you know some of the social science or health related, health science stuff." Data discovery takes time and this task unlike others might be better addressed by a subject liaison as Participant 9 stated, "I had to draw the line somewhere." Finally, one participant did not locate data at all, rather they referred patrons to their appropriate liaison who handled the transactions for discovery.

Reviewing DMPs

When asked about reviewing DMP responsibilities, seven participants did not review or occasionally reviewed DMPs as part of their position. These DSLs could walk patrons through the DMP tool to draft a DMP, but did not extrapolate further. Two participants reviewed DMPs upon request, with one of those indicating this work was often done last-minute because researchers did not realize they needed a DMP until grant proposal deadlines. "I rely heavily on the various templates in the DMPTool. The other [thing] I've done is collecting successful DMPs. I usually keep the working copy so that I can refer back to that later when I have another researcher who's working with similar data types" (P5). Another librarian sometimes acted as a direct collaborator and was added to research teams in grant proposals since their data services would be central to the research. Participant 10 primarily reviewed DMPs for their position, averaging about 18 per year and performing re-reviews for quick turnarounds. Clearly, the DMP roles vary on each position, institution, and the skills of the DSLs.

As naming a repository is a factor in many DMPs, participants were asked if they have assisted patrons in locating an appropriate repository. Four participants occasionally helped and six helped frequently to locate repositories. They often recommended the local IR as the "first go-to option in terms of being low effort, low barrier, and low costs. They're free to [university] researchers" (P4), "but we're also happy to help people find more of a domain-specific repository" (P3). Participant 5 found it "a very rewarding part of data librarianship because getting your data into a repository is much easier than a lot of people would think." That participant made sure researchers used the appropriate repository as prescribed by a journal, as well as provided other options. Participant 6 used a decision tree with researchers to figure out where to put data. Participant 8 said, "A part of my job is also convincing people that [in addition to storing data] it's worth publishing

JeSLIB 2022; 11(1): e1226
https://doi.org/10.7191/jeslib.2022.1226

data that they don't think needs to be published and finding a good place for it." Participant 10 developed a strategy of suggesting three repositories, "One that would be absolutely perfect; one that is middle of the road, it would get the job done and you'll get some exposure; and one that's really general and wide open." In this scenario, the librarian helps the researcher decide where to deposit data.

Next, participants were asked two questions about who is responsible for DMP implementation at their institutions and how DMPs are evaluated. Five participants said DMP implementation was up to the researcher; three were unsure and assumed the researcher was responsible. One participant's library is currently working with the Office of Research to standardize DMP implementation. Participant 9 said, "I think the Office of Research would like to say that they are, but I don't know how much they do it to be honest. I know that, in medicine specifically, we've had a number of issues where people—researchers—the day before, they're trying to prove they followed through on their DMP." Given these first-hand experiences with compliance issues, Participant 9 wisely remarked, "there's a carrot, [but] there's no stick yet." Two participants shared that other entities were responsible for implementation, such as journals, funders, and other academic departments. DMP implementation may not be a role for DSLs upon this small group or participants, but like many other RDS as one participant put it "there's a lot of discussion among the library in terms of saying we should be better at implementing enforcement" (P8).

Some librarians did evaluate DMPs. Three participants evaluated DMPs in ways like looking for completeness, participating in joint peer evaluations, and sending evaluations back to project PIs. Five participants expressed that they did not evaluate DMPs. Among the responses to these questions, librarians pointed to the lack of enforcement to establish best practices as well as how it was "really up to the individual researcher if they want input from any outside party" (P4). Still, two participants did not specify or were unsure about any DMP evaluation at their institutions because the library was not involved.

Given the details collected through the DMP questions, the interview concluded this section by asking each participant what the ideal structure and process for DMP implementation and evaluation would be. Not surprisingly, participants had vastly different approaches to an ideal structure and process than the current workflow at these institutions with academic librarians more central. Two participants wished for a single intake portal, three others suggested that all researchers take in-depth DMP training beforehand or as a refresher, and two additional librarians suggested that they should be informed of who received grants so they may start working with the researchers at the start of a project rather than at the end. Other unique ideas included: (1) establishing mechanisms for DMP review similar to an IRB review; (2) linking DMPs to a faculty's evaluation processes (i.e. annual review) to increase transparency and accountability, to incentivize good behavior; (3) creating closer connections between any IR librarian and data managers on projects; (4) selecting DRs for storage based on the discipline; and (5) building a data management staff team with campus wide reach

JeSLIB 2022; 11(1): e1226
https://doi.org/10.7191/jeslib.2022.1226

and authority to address the "conflicting opinions and tension for who should be doing data management" (P9).

Doing Outreach and Collaboration

Participants were then asked about any outreach and collaboration aspects of their work. Examples of activities mentioned included brown bag series, interdepartmental work, hosting joint workshops, teaching courses, and "sneaking in" short DMP best practices presentations at the end of other instruction, as well as working with the RCR program manager specifically on responsible RDM workshops. One participant collaborated with other university departments to "further [the university's] mission of open science and better data management" (P6).

Several DSLs provided little detail on collaboration as Participant 7 explained trying to create a local data catalog "that would be a place to make your own local data discoverable for other local researchers," but was several years ago and fizzled out before having a workable solution. Participant 9 explained some barriers to collaboration: "'My subject liaison [...] must be able to help me with that.' And then they get kicked to me and they think, 'Why am I being kicked to another person?'" With organizational structures changing, collaboration may be impeded by users' challenges simply navigating academic libraries' own services.

Data Visualization

Although noted in the literature for over a decade as a service performed by DSLs, only four of the ten participants interviewed currently assisted with data visualization (Primich 2010). Two librarians indicated they assisted in data visualization without elaboration. Participant 3 had a great example of developing an online application that runs on R wherein researchers may upload biomedical data and create graphics. Finally, one librarian previously worked with data visualizations tasks. "[I] used to give a kind of introduction to the data visualization workshop. I've stopped doing that. It's one of my job responsibilities that has kind of fallen to the wayside because I figured you can't be good at everything" (P5). This sentiment was expanded upon: "I also found that a lot of students are using R or Python to create visualizations now and that is…I just don't know how to do that." DSLs do a variety of tasks and, within this sample of ten, data visualization did not appear to be a task for most of them.

Training and Data Policy

Participants, who indicated they did any tasks related to RDM training or data curation, were then asked more details about those job tasks. For training, all but one participant provides RDM training, and librarians used a variety of methods to encourage RDM education from holding one-on-one sessions to collaborating on workshops (i.e. data visualization, GitHub, data curation), or hosting a data week. Participant 1 found office hours ineffective because people only see it as "a

JeSLIB 2022; 11(1): e1226
https://doi.org/10.7191/jeslib.2022.1226

just-in-time service" and not have the foresight to attend a workshop. Participant 2 generalized that their training focused on "recommended practices, tools, compliance, and then how to write a data management plan." These librarians struggle to obtain participation and engagement. As Participant 5 put it, "Nobody's gonna attend a research data management [workshop]. Like, that's the least sexy term, right? I wish we could just, like, totally rename that!" Participant 7's institution used to provide training many years ago, but no longer does.

When asked about their own training, all participants mentioned a lack of formal training in preparation for their positions as data educators, commenting that they learned on the job, were self-educated, and took courses and workshops as listed in the results section for the *Job Analyses* section below. Participant 1 felt training resources were not current. "Not to say that they're not valuable, they're just not realistic to what the researcher needs." Participant 9 felt the support was not there stating, "I think my library assumed that, if you have your own data, you should be good at teaching other people how to manage their data." Participant 4 described early action they and their co-workers took in self-education. In the 2000s, they started receiving RDM questions and later formed a science RDM group to focus on those services. "As data management became more important, in 2010, the National Science Foundation announced that they were gonna require data management plans in their proposal and people in that data management group collaborated with folks [...] to create a series of brown bag workshops about data management."

The role of RCR training was mixed among the participants' institutions. Three participants' Offices of Research typically hosted this training, one commenting how the library recently obtained this responsibility. "We are in development with that right now […] [They] actually asked us to do it" (P2). RCR training was "completely managed by the deans and the librarians," at Participant 8's previous institution, but not their current university. "Mind you, [the university] does like almost exclusively biomedical research and so there just happens to be a lot of biomedical publishers there in the area and so it's very easy for the librarians, who already have relationship with the publishers, bring them in and help people understand the ethics of publishing, visualization of data, things like that" (P8). Three participants shared their libraries had no role in RCR training. Finally, three participants were unsure if any RCR training existed on campus, one commenting, "It depends on how we define research because [our university] has an awful lot of compliance things" (P1). The unsure responses indicate that academic librarians do not do RCR training sense participants were not sure exactly what that topic entails.

Next, participants were asked if their institutions have any data policies. Participant 1 was the only librarian who confirmed a university-wide data policy in place; however, the policy encompasses only research data. Three participants confirmed their institutions did not have data policies in place—"It's all back to the individual researcher" (P4) or the Office of Research, shared Participant 3. Four participants described how data policies were out-of-date, about attempts at

JeSLIB 2022; 11(1): e1226
https://doi.org/10.7191/jeslib.2022.1226

composing policies—"[My institution] has been drafting a research data policy for seven years now" (P10). One institution had a data retention policy and guidance on everything else, but the librarians asserted it read like "we own your stuff but you're allowed to do things with it" (P6).

Data Curation

Only five DSLs indicated data curation was a task they performed during the Overview of Responsibilities questions; many had more to share on the topic. The seven data curation task questions asked did not elicit consistent responses from participants, therefore, librarians each gave an overview of data curation tasks. Two participants said these were among their primary responsibilities, while five other participants explained some tertiary data curation activities of their jobs.

As one of those indicated data curation was their primary responsibility, Participant 3 described all the typical tasks as a member of a research team. They checked codes for completeness and established a workflow for documenting all curation activities. In part, they evaluated data based on the FAIR Data Principles with special attention to the licensing and copyright (Wilkinson et al. 2016). As part of the curatorial process, this DSL encouraged researchers to transform their files prior to compilation in a zip file to avoid errors later and checked all data prior to deposit to make sure it looked good and was well-documented. The other DSLs that state these tasks were their primary job echoed the same duties as well as metadata augmentation; checked if files opened; identified missing data; and troubleshoot any other issues with the researcher as needed.

For those participants that did some data curation, but not as their primary responsibility, the responses varied. Participant 1 participant evaluated for FAIRness on the license side of things and transformed many files. Participant 2 provided metadata curation services and research consultations as well as assisted their IR librarian with complicated cases "as a second pair of eyes" (P2). Participant 5 minted DOIs and usually requested missing information like a readme file. Finally, Participant 9 helped load data into the repository but would not characterize it as data curation given they did not document their own actions; however, they did check files, code, and read documentation.

Institutional Repository and Long-term Commitment to Research Data

In response to their roles in managing an IR, nine out of ten participants did not manage an IR, most noting that other units managed repositories on campus. Eight of those nine confirmed that their university did have an IR. IR arrangement differ across university systems and institutions. For example, Participant 5 had a DRYAD instance run by their state's digital library and their university paid a membership fee. A different sample may have led to more participants that managed IRs, but at these top institutions IR duties fall beyond the academic libraries and information technology offices often manage those operations. Participant 10's library runs an electronic thesis and dissertation (ETD) repository,

JeSLIB 2022; 11(1): e1226
https://doi.org/10.7191/jeslib.2022.1226

which was their "third version of a repository. The first two were homegrown off [other repositories] and they could never develop it to the point of what it needed to be, or anything bigger than about three [gigabytes]." The ETD IR does not take data only text. Participant 3, the only IR manager, managed an outdated instance of Dataverse with their team, but now there is a separate data storage entity on campus "where all the research data goes".

To gauge each university's data preparedness, participants were then asked about their institution's commitment to the long-term management of research data as well as if there was a budget and infrastructure for managing data beyond the life of grants and projects. Three participants confirmed an institutional commitment to the long-term management of any research data. In one instance, a participant stated their institution was a part of a national network and "it's meant to live through [natural disasters because] they've got different nodes in different parts of the country" (P7). Three other participants confirmed some institutional commitment, but without specifics. For example, one institution originally advertised "in-perpetuity" as a time commitment for data but scaled back to ten years with a re-evaluation because "a lot of people [were] nervous about accepting stewardship for anything" (P4). The realization that in-perpetuity meant resources led another participant to remark that their institution guaranteed storage for 5-10 years with re-evaluation down the road. Participant 8 confirmed no institutional commitment, but wished one existed. Three other participants were unsure of the institutional commitment. One thought their university was committed but "we just don't have a data repository" (P1). Since many of these new DMP requirements are just now reaching ten years old, one librarian indicated that next steps were "kind of fuzzy" (P2). Still, one other librarian felt confident that any long-term commitment could use DRYAD to hold data in-perpetuity.

Five of the ten participants acknowledged a budget existed for long-term data management. One institution paid a DRYAD institutional membership with state-level support, but overall none knew what these data curation and data management efforts cost. One participant did assume that these data-related costs were wrapped up in the IT budget due to their "technical support" (P3). The interviewer probed to know more about how RDM efforts were funded for sponsored projects and all other projects. Responses included researcher-funded and other grant-funded sources, free data deposits in the IRs, and funding from other institutional entities or unknown sources. One institution had a grants administration that managed everything pre- and post-award, encouraging researchers to include RDM as a grant line item. Participant 10 said by making RDM a line item, there is time to get appropriate funding. Unfortunately, the librarians' sentiments were that most researchers often neglect RDM because either they assume efforts are a part of other research activities or would be no-cost library services.

JeSLIB 2022; 11(1): e1226
https://doi.org/10.7191/jeslib.2022.1226

Job Analyses

After discussing the job responsibilities, DSLs were asked six questions about their job titles, experience, and educational background. Eight participants held the title of "Librarian"—specific titles including Data Librarian (2) and one each Data Science Librarian, Data Services Librarian, Science Data and Engineering Librarian, Sciences Data Librarian, Data Management and Curation Librarian, and Research Data Management Librarian—and the remaining two held the title of Research Data Management Consultant. In total, six participants held faculty ranks of assistant (3), associate (1), and full (2) and four participants were considered staff. When asked how many years they had been in their current position ranged from less than 1 to 11 years, averaging about 4 years. The participants' years spent in the field ranged from 6 to 26 years, averaging about 13.5 years.

All participants held bachelor's degrees. Four participants had degrees in these STEM-related fields: Biology, Ecology, Geography, and Geoscience. Six participants had degrees in these non-STEM-related fields: Classics, a dual-degree in English and German, Philosophy, Psychology, Sociology, and Women's Studies. Nine participants held master's degrees. Six of them held only MLIS degrees, one held dual master's degrees in LIS and Sociology, and one each in Geography and Marine science and management. One of those nine participants shared they were presently working towards an MLIS. Two participants held doctoral degrees in Biomedical sciences and Geography.

In response to other credentials, eight participants learned additional skills on the job and through conference sessions (i.e. Association of College and Research Libraries), online course offerings (i.e. Coursera and Udemy), and workshops, such as Research Data Management Librarian Academy and North Carolina State University's discontinued program (Data Science and Visualization Institute). Two participants earned certificates in GIS and one completed some copyright training. Other than those mentioned, none of the participants received any formal education in RDM. One participant said data management was not a part of their graduate program's curriculum. Another participant's group started self-educating on all aspects of RDM a long time ago because their department began receiving many RDM inquiries. This was the general trend across all interviewees, and the DSLs collectively were pleased to know that awareness was being raised about formal education and training prior to starting in their field and the strive toward standardization. "I know that there are other people doing similar work but it seems like everyone always has a different job title. I'm really hoping, if I can find some other people, still being new to this, at other big institutions because the problems I'm facing aren't the same at smaller institutions" (P3).

Project Feedback and Pandemic Context

The interviews ended with a request for any additional feedback. As Participant 6 put it there is "an interesting curve where you have people like me who this is all I do, is data management and curation. Then you have other folks who are

JeSLIB 2022; 11(1): e1226
https://doi.org/10.7191/jeslib.2022.1226

expected to do all things." Some DSLs are doing the work of multiple jobs given the increase in data-intensive work across all disciplines. Participants were more interested in learning what other libraries do differently to inform best practices across RDM and data curation work. As Participant 1 wondered, "Is it the same job from place to place? The qualifications and the focus are very different."

Additionally, since the interviews were conducted in March and April of 2020, the pandemic was mentioned. Participant 3 found the transition seamless because consultations moved virtually, and most of their resources were digitized already and accessible online. Overall, participants noted that patrons were not familiar with how the library operated virtually, but once they figured out their work increased. Anecdotally, more researchers had time to deposit data during the initial lockdown and catch up on neglected RDM duties.

Discussion

At these top research universities with large research expenditures, most DSLs had at least one data librarian co-worker with a few institutions housing whole departments of DSLs to allow for specialization and support. With the variety of data across disciplines, a variety of DSLs would be ideal to specialize within domains and better serve unique issues within each domain. This might not be a workable option for most academic libraries given staffing limitations. If more students, faculty, and staff ask for data help, then perhaps more and more academic librarians all will serve in some capacity to provide RDS with expertise for data types emerging over time. This was reflected in the responses to the questions about which disciplines, departments, and data types, each librarian covered. Six of the ten participants assisted with specific disciplines (e.g., Social Sciences), while two of those participants and four others also stated they aided across all academic departments. One or two DSLs may not be able to provide in-depth assistance at large institutions.

The participants in this study worked with all types of data, with tabular data being directly stated most, but also geospatial, biomedical, and qualitative data were mentioned at least once. Future DSLs need to work with an increasing variety of data types and at least four said that this versatility was essential to their jobs. Data science training and electives in iSchool programs may prepare students for this future work with a variety of data types beyond text-based and/or tabular data. All librarians mentioned serving at least one of the following groups: undergraduate students, graduate students, post-docs, faculty, staff, and other librarians. Clearly, these data needs exist across all library users in the 21st Century. The following discussion reviews in greater detail the implications of the job responsibilities described by these participants for job descriptions, training, education, and differentiation for the future of DSLs and other data curation positions.

Locating Data

The task of locating data appears to be central to most DSLs' jobs. Still, these activities were more or less of a focus depending on the participant. Only two participants indicated that they did not assist in locating secondary data and referred users to the appropriate liaison librarian. The demarcation between a liaison librarian and DSL may continue to blur as data needs increase for all users.

Reviewing DMPs

In contrast to locating data, the review of DMPs was not done by a majority of DSLs. As DMPs remain a central aspect of RDM for funded projects, the majority of participants not assisting with review was a surprise. Many DSLs could walk patrons through a DMP tool to draft a DMP even if they do not assist with writing or implementing. DMP work occurs in the planning stage of the research cycle and DSLs involved at the start may be viewed more favorably to assist with roles downstream if they are already integrated into a project.

Locating a Repository

This task was pervasive among this study's DSLs. Using journal and organization recommendations, most DSLs can locate an appropriate data repository. As Participant 8 noted the job is not always just locating one, but promoting the benefits of sharing data. Although other resources exist for other domains, DSLs may use the American Geophysical Union (AGU)'s Repository Finder for most sciences (https://repositoryfinder.datacite.org). DSLs may develop similar tools to the repository finder appropriate for their institution to assist this type of information, especially if more funding agencies request data be shared in Trusted Repositories. One example of a certifying body for data repositories is CoreTrustSeal certification. For repositories to receive this certification they must exhibit and maintain high standards for documentation, usage licensing, access continuity, ethical usage, organizational infrastructure, expert guidance, data integrity, data evaluation, storage procedures, long-term preservation plan, data quality, workflow documentation, data discovery, data reuse, technical infrastructure, and security (https://www.coretrustseal.org/about). As one participant pointed out, there might likely be a few options for each researcher and DSLs can explain the pros and cons of each to inform their repository choices.

DMP Implementation and Evaluation

DMP implementation at each institution was relatively unknown. With the majority of interviewed DSLs either assuming DMP work was up to each researcher or unsure who was responsible. DSLs might not be as involved in day-to-day RDS as hinted at in the literature, but they could be. Although one participant's library is currently working with the Office of Research to standardize DMP implementation, most were not aware of any efforts to do so. As two participants suggested, as more journals, funders, and other academic departments encourage

JeSLIB 2022; 11(1): e1226
https://doi.org/10.7191/jeslib.2022.1226

implementation and evaluation then researchers will need to comply. DMP implementation may not be a current role for the DSLs in this study, but like many other RDS as one participant put it "there's a lot of discussion among the library in terms of saying we should be better at implementing enforcement" (P8).

Although three participants evaluated DMPs, half of the participants in this study did not do this task and two were uncertain if they did, which is likely also not a common task done by data services librarian at any institution. More DMP evaluation may be on the horizon with institutions needing a third party beyond anyone who helped write the DMP for objective feedback on this research outcome.

Not surprisingly, the ideal structure participants suggested for DMP workflow varied except in the fact that academic librarians would be central. This is odd since, at present, DMP workflow beyond writing did not involve DSLs at many of these institutions. With the requirement at most US funding agencies a decade old, it is odd none had something more formal in place related to these RDM needs. All the actionable suggestions by participants relate to existing DSL roles, but not all DSLs are doing the same jobs. All the suggestions require academic libraries to be a part of each campus's research lifecycle and there seems to be a disconnect at many institutions with the academic library not well connected to RDM efforts across campus or associated research funding. Any of these DSLs could take the initiative to start making their ideal structure a reality.

Doing Outreach and Collaboration

The outreach and collaboration aspects of the participating DSLs' responses map to traditional library tactics with mixed results. Participant 9 mentioned a phenomenon, "There are some [subject librarians] who...don't want anyone talking to their faculty unless they're there, which slows things down." This approach from some liaison librarians may confuse users as to who their library contact is and to whom they should ask questions. With data becoming more central to research, the subject and data librarian roles may need to work in concert when assisting patrons with data.

Data Visualization & Data Science

Of the four librarians that assisted with data visualization, only two gave specific examples. Data visualization is a skill nested in data analytics and, in some programs (i.e., R), the two tasks are rather seamless. The tasks entailed with this potential data service depend on the needs of each institution as well as similar services being offered by other campus entities. Still, the participants in this study did not have much to share on the topic to indicate that academic libraries were doing much. There was a sense that several of the DSLs interviewed need to learn data visualization and data science skills on their own to remain relevant in their field. Participant 2 noted that "in my job to really succeed, I really have to spend some time polishing off my data science skills…I'm seeing more and more of my

JeSLIB 2022; 11(1): e1226
https://doi.org/10.7191/jeslib.2022.1226

colleagues and other people that work in data management could effectively go work in industries as a data scientist if they wanted to. That's a little intimidating!" Data visualization is important for articulating the importance, relevance and significance of data. Many researchers use open source software and tools to visualize data. Some journal subscriptions packages include API features to visualize data. DSLs can assist when needed, if they have proper training in data analytics and visualization. DSLs can learn these data visualization skills from use of software packages such as MATLAB, R, SAS, Tableau, and targeted training supported by academic units, departments, or funded projects. Perhaps, DSLs are not needed in disciplines where RDM is embedded in the program, whereby the DSLs may not have the domain knowledge in which to best assist researchers with data visualization. Thus, collaborations on training support needs across multiple communities of practice can best leverage DSLs' impact, outreach, and support. Professional development coupled with active participation in funded research projects can enable DSLs to develop sustainable data visualization competencies, research engagement, and data science skills. The North Carolina State University Libraries Data and Visualization Services (https://www.lib.ncsu.edu/department/data-visualization-services) is a great example of helping students and researchers develop critical data science skills. However, for other institutions, the data and visualization services may be offered outside of the libraries and/or in collaboration with other campus training (e.g. research computing, discipline-specific).

Training and Data Policy

DSLs explained their own RDM training as self-taught or attended workshops to get up to speed. As Participant 4 indicated, a model of a RDM group formed to address mounting RDM needs. This likely occurs elsewhere when libraries pivot services and reallocate resources to serve their users. Responsible Conduct of Research (RCR) training was not something most DSLs were involved with but, again, one participant had a model where librarians involved in the scholarly communication process could assist with campus wide RCR training. Academic librarians may be an important, untapped piece in the RCR framework across many campuses given their central and longstanding roles of access to information, citation, copyright, and other information literacy activities.

Only Participant 1 confirmed a university-wide research data policy in place and three other DSLs were confident that there were no data policies at their institution. For many universities, a DSL may serve a vital role in the development of a research data policy as a central stakeholder for research data on any campus. Other stakeholders involved in RDM policy would include the Office of Research, Research Computing, Research Compliance Office, Information Technology Department, Researchers, Academic Units, and the Libraries (Erway 2013). From the library perspective, the initial conversations may ensue as outgrowths from the formulation of RDS task forces or working groups that lead to coordinated research data efforts across campus. One use case example includes the University of Florida's (UF) DSL as Chair of the Data Management and Curation Working Group (https://ufdc.ufl.edu/AA00014835/00011/allvolumes), working in

JeSLIB 2022; 11(1): e1226
https://doi.org/10.7191/jeslib.2022.1226

collaboration with UF Research Computing Advisory Committee (RCAC), UF Clinical and Translational Science Institute (CTSI), UF Information Technology (IT), and UF Office of Research to develop a *Supporting data management at UF Proposal to the Office of Research* draft RDM policy in 2019. The Association of Research Libraries (ARL)/Canadian Association of Research Libraries (CARL) Joint Task Force on Research Data Services Final Report (2021) Objective 2: Develop a roadmap with recommendations for the roles of research libraries with regard to research data principles, policies, and approaches to managing research data in the context of the Open Science by Design framework and recommendations section, particularly Recommendation 7: Define an institutional strategy for RDS is relevant to DSLs across higher education institutions (HEIs). Future research could explore the formation and implementation of research data policies, but these participants did not indicate academic libraries were that proactive at their institutions.

Data Curation

If checking code, data quality and data assurance checks, and other issues to create well-documented data become necessary, DSLs need more technical and data policy skills to effectively curate data. Most interviewed did not do these tasks but were aware of licensing and how to ask for missing information in a readme file to improve data reuse. Data curation activities can be labor-intensive, and domain-dependent, so having DSLs specialize in certain data types or domain areas would be ideal to assist with these tasks. A reason for academic libraries to serve in these roles is the relative longevity of these employees compared to assigning data curation responsibilities to a graduate student who may move on in short order.

Institutional Repositories and Long-term Institutional Commitment to RDM

Only one out of ten participants managed an institutional repository (IR), but all interviewed knew there was an IR on their campuses. "A university-based institutional repository is a set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members" (Lynch 2003, p. 2). The findings were especially surprising given that at these institutions RDM tasks related to IRs had been handled by other entities on campus. If academic libraries host the research data services, then why not the research data resources. This likely varies greatly for each institution and more academic libraries may host IRs than this small sample with research infrastructures that support more resources and IT not housed in academic libraries. For DSLs to have a seat at the data table, it might be easier if the table (i.e., IR) was hosted by the academic library or otherwise technically supported.

As RDM matures, initial promises to retain research data "in-perpetuity" may not be sustainable given limited resources and storage at scale. DMP requirements should grow in number and sophistication to enable data reuse, therefore, long-term commitment to RDM should be included in strategic plans and other

JeSLIB 2022; 11(1): e1226
https://doi.org/10.7191/jeslib.2022.1226

university documentation to ensure the considerations of researchers and institutions. This is reflected in the interviews with DSLs stating that a budget existed, but few participants knew the details. Several assumed the data-related expenses were in the IT budget or funded by sponsored projects, but not with commitment from the academic library. DSL's jobs are tied so closely to the continued need and reuse of data that more should be explored on how to secure RDS via soft and hard research funding. As most DSLs are not directly involved but are directly impacted by budget discussions by senior library administrators, budget issues that impact capacity building, infrastructure development, and resources should include libraries with other campus and external partners. DSLs may become an increasingly important part funded research projects to develop long-term institutional commitment to RDM. Long-term institutional commitment to RDM requires communication, collaboration, and cooperation of key stakeholders. For a current example, the UF DSL is involved in a current, multi-million-dollar US Department of Agriculture funded project that generated direct costs monies to the Libraries. The UF DSL facilitated the development of successful DMPs that total over $6 million in research dollars for the first two years of the new DSL position. Tenure-track DSLs with PhD at R1s that promote research are expected to develop grant proposals as part of the tenure seeking process could do something similar.

Emerging Skills and Workforce Development

A recent workforce development report from the National Science Foundation (NSF) Office of Advanced Cyberinfrastructure (OAC) outlined issues that are echoed in this study's results show that the future of DSLs and other data curation positions depend largely on the pipeline between training and hiring (NSF 2019). Job descriptions and preferred requirements for these jobs might be informed by those currently working in these positions. Many current DSL job descriptions contain a long list of skills and duties that may be desirable, but not realistic for one individual to have competence in. In fact, hiring an individual that excels in one of these many RDM tasks could benefit a library more than someone that does a mediocre job on many. In addition, the courses and continuing education for the changing roles of DSLs may benefit from study of those currently crafting the future of this profession on-the-job.

DSL job titles varied, but all include data somewhere in the title. Eight of the ten job titles included librarian with RDM Consultant being the job title for the other two. Job titles are meaningless, but if other librarian titles included the information object it would sound strange (e.g., Science Document Librarian). As data becomes more central to all disciplines, these data jobs could supplant the traditional document-focused librarian lines. As this study and the literature reviewed indicate DSLs do much more than DMPs. While the DMP mandates spawned need and growth of this area, as RDM matures, and a culture change sets in across domains then DSLs' roles will evolve.

JeSLIB 2022; 11(1): e1226
https://doi.org/10.7191/jeslib.2022.1226

The majority of participants held degrees from non-STEM-related fields, but data in all disciplines is more vital, STEM backgrounds of individual librarians may matter less. Two participants had doctoral degrees and eight master's degrees indicating advanced training is either needed or helps secure employment for the DSLs. These participants have more personal research experience, which could inform their RDM training and assistance with others. Additionally, the participants explained they gained new skills on the job and through various continued education experience (i.e., webinars). With more iSchool programs offering classes and concentrations, recent graduates should have more exposure to these topics.

The recent release of the Association of Research Libraries (ARL)/Canadian Association of Research Libraries (CARL) Joint Task Force (TF) on Research Data Services (RDS) Final Report (July 16, 2021) recommendations are relevant for DSLs and iSchool programs offering RDM classes and concentrations:

- Recommendation 1: Conduct a cross-campus mapping of existing campus resources and researcher needs for RDS
- Recommendation 2: Define a library portfolio and strategy for RDS
- Recommendation 3: Articulate library and institutional research data services and partnerships
- Recommendation 4: Formalize partnerships through development of a service catalogue
- Recommendation 5: Document services by elements of data management requirements
- Recommendation 6: Evaluate the program on a spectrum of maturity
- Recommendation 7: Define an institutional strategy for RDS

The findings enhanced by the ARL/CARL TF RDS Final Report (2021) and Association of American Universities (AAU) and the Association of Public and Land-grant Universities (APLU) Accelerating Public Access to Research Data Guide (2021) support professional and workforce development for DSLs and graduate students seeking to work academic libraries. This study contributes additional qualitative data that supports these recommendations.

Conclusions

This study's contributions include distinguishing between the various DSLs based on RDM activities, support, and workflows. Future job descriptions, training, and education will all benefit from differentiating between the many associated data services roles those in these data curation positions provide, especially with an increased focus on research data greater specializations will emerge. This article concludes with a proposed matrix of job tasks that indicates different types of DSLs to inform further study. Table 2 presents the potential types of DSLs and descriptions of each DSL type follow.

JeSLIB 2022; 11(1): e1226
https://doi.org/10.7191/jeslib.2022.1226

Table 2: Types of Data Services Librarians (DSLs).

Type of DSL by Job Task	Secondary research data	DMP review	Outreach / Collaboration	RDM training	Data curation	Manage IR
General DSL (P 1, 5 ,8)	3	2	3	3	0	0
RDM DSL (P 2-4, 6, 9, 10)	4	5	5	6	5	1
Discovery DSL (P 7)	1	0	1	0	0	0
Total for all DSLs	9	7	9	9	5	1

Earlier work makes a distinction between the data generalists and domain subject specialists, but this broad typology of either a breadth of knowledge or depth of expertise in a particular discipline is not actionable (Federer 2018) by DSLs with only library information science (LIS) domain knowledge, skills, and abilities (KSA). Still, this model presents one type as the *General DSL* because their tasks most closely relate to traditional library services and resources such as locating data, outreach, and training, with a data focus. Another clear demarcation for the *General DSL* is that beyond finding and teaching about RDM they do not manipulate, or otherwise analyze data like performing curation or managing an IR.

Those DSLs with more RDM training and data curation tasks in addition to the traditional library service and resources tasks were ascribed *RDM DSLs*. These DSLs must collaborate with domain subject specialists to develop data skills. The CODATA-RDA (2021) School of Research Data Science is one model in which DSLs can develop the fundamental requisite research data science skills to better effectuate RDM support services. The *RDM DSLs* do it all. In contrast, the singular *Discovery DSL* participant worked as a reference librarian for data, which with growing demand may be its own professional position at more institutions. Collaboration was one area lacking from these interviews, but as more librarians offer data analytics and visualization services, these specializations may emerge as another type of DSL (i.e., embedded DSLs).

Table 2 presents one attempt based on these 10 participants to demonstrate there are several types of DSLs based on the data services they offer most. Depending on the academic needs of each institution, employers may need a certain set of research data services and this work could inform future job descriptions. The typology-by-job tasks could also inform training to move beyond generalized RDS to focus on the different roles academic librarians do and could play in any academic setting. DSLs are increasingly expected to perform evolving research data science skills. Therefore, each institution likely needs more DSLs to meet these demands with a variety of librarians working the RDM. Research data science encompasses the ensemble of data skills that include (1) principles and practices of Open Science and research data management and curation, including data repositories, (2) the use of a range of data platforms and infrastructures, (3) large scale analysis, (4) statistics, (5) visualization and modelling techniques, (6) software development and annotation, and (7) more (CODATA-RDA 2021).

Academic libraries will need to modify current and future position vacancy announcements that include research data skills for traditional DSLs and non-DSL (i.e. domain subject specialists such as Informatics Librarian, Reproducibility Librarian, and Bioinformatics Librarian with PhDs in non-LIS disciplines). Existing DSLs will need to upskill their KSA through training and workforce development annually via continuous professional development to learn RDM skills. As Participant 6 states some DSLs are "expected to do all things [...] those are jobs for like five people, not one." It would be unscalable and unsustainable if only one librarian at each university was the analog services librarian or journal services librarian and expect adequate services for each campus. In another recent study on RDM services, half of the participants think their library needs full-time RDM experts (Faniel and Connaway 2018). If this is the case, more specialization will be a necessity with additional DSLs supporting research across domains. A few of the academic libraries in this study had several DSLs and this model would best serve any academic library that strives to serve the evolving research data science needs of their researchers, faculty, staff, and students, across all domains. Each institution will make their own administrative choices to where on campus these RDM tasks will be performed. This work shows the strengths of housing and fostering them in academic libraries.

Acknowledgements

The authors greatly appreciate the participants as well as the Spring 2020 Faculty Development Leave of the first author to allow for data collection.

Supplemental Content

Appendix
An online supplement to this article can be found at https://doi.org/10.7191/jeslib.2022.1226 under "Additional Files".

Data Availability

Deidentified data for this article can be found in the Tennessee Research and Creative Exchange (TRACE) institutional repository: Bishop, Bradley Wade. 2020. "Data from Data Services Librarians Study." [Data Set] University of Tennessee, Knoxville Libraries. https://doi.org/10.7290/m29yhy5qen

References

Association of Research Libraries. 2021. "ARL/CARL Joint Task Force on Research Data Services Releases Final Report." Washington, DC. https://www.arl.org/news/arl-carl-joint-task-force-on-research-data-services-releases-final-report

Association of American Universities and Association of Public and Land-grant Universities. 2021. "Guide to Accelerate Public Access to Research Data." Washington, DC. https://doi.org/10.31219/osf.io/tjybn

Atwood, Thea P., Andrew T. Creamer, Joshua Dull, Julie Goldman, Kristin Lee, Lora C. Leigdon, and Sarah K. Oelker. 2019. "Joining Together to Build More: The New England Software Carpentry Library Consortium." *Journal of eScience Librarianship* 8(1): e1161. https://doi.org/10.7191/jeslib.2019.1161

Bird, Colin, Simon Coles, Iris Garrelfs, Tom Griffin, Magnus Hagdorn, Graham Klyne, Mike Mineter, and Cerys Willoughby. 2016. "Using Metadata Actively." *International Journal of Digital Curation* 11 (1): 76–85. https://doi.org/10.2218/ijdc.v11i1.412

Bishop, Bradley Wade. 2020. "Data from Data Services Librarians Study." [Dataset] University of Tennessee, Knoxville Libraries. https://doi.org/10.7290/m29yhy5qen

Bishop, Wade, Hannah Collier, Ashley Marie Orehek, and Monica Ihli. 2021. "Potential Roles for Science Librarians in Research Data Management: A Gap Analysis." *Issues in Science & Technology Librarianship* (98). https://doi.org/10.29173/istl2602

Brandenburg, Marci D., Sigrid Anderson Cordell, Justin Joque, Mark P. MacEachern, and Jean Song. 2017. "Interdisciplinary Collaboration: Librarian Involvement in Grant Projects." *College & Research Libraries* 78(3): 272-282. https://doi.org/10.5860/crl.78.3.272

Carlson, Jacob, Lisa R. Johnston, and Brian Westra. 2015. "Developing the Data Information Literacy Project." In *Data Information Literacy: Librarians, Data, and the Education of a New Generation of Researchers*, edited by Jake Carlson and Lisa R. Johnston, 35-50. West Lafayette, IN: Purdue University Press.

CODATA-RDA. n.d. "Schools of Research Data Science." Initiatives - Data Skills. Accessed March 2, 2022. https://codata.org/initiatives/data-skills/research-data-science-summer-schools

Corrall, Sheila. 2014. "Designing Libraries for Research Collaboration in the Network World: An Exploratory Study." *LIBER Quarterly* 24(1): 17-48. http://doi.org/10.18352/lq.9525

Cox, Andrew M., Mary Anne Kennan, Liz Lyon, and Stephen Pinfield. 2017. "Developments in research data management in academic libraries: Towards an understanding of research data service maturity." *Journal of the Association for Information Science and Technology* 68(9): 2182–2200. https://doi.org/10.1002/asi.23781

Cox, Andrew and Eddy Verbaan. 2018. *Exploring Research Data Management*. London: Facet Publishing.

DCC. 2007. "Curation Lifecycle Model." Guidance. https://www.dcc.ac.uk/guidance/curation-lifecycle-model

DCMI. 2021. "Dublin Core Metadata Initiative." Metadata Basics. https://www.dublincore.org/resources/metadata-basics

Erway, Ricky. 2013. "Starting the Conversation: University-wide Research Data Management Policy." Educause Review. Last updated December 6, 2013. https://er.educause.edu/articles/2013/12/starting-the-conversation-universitywide-research-data-management-policy

European Commission. 2015. "Final report of public consultation on Science 2.0/open science: Validation of the results of the public consultation 2.0: Science in Transition." https://tinyurl.com/yzf3zvnm

Faniel, Ixchel M., and Lynn Silipigni Connaway. 2018. "Librarians' perspectives on the factors influencing research data management programs." *College and Research Libraries* 79(1): 100–119. https://doi.org/10.5860/crl.79.1.100

Federer, Lisa. 2018. "Defining data librarianship: a survey of competencies, skills, and training." *Journal of the Medical Library Association: JMLA* 106(3): 294–303. https://doi.org/10.5195/jmla.2018.306

Gunderman, Hannah. 2021. "Developing Lesson Plans for Teaching Spatial Data Management in Academic Libraries through a Lens of Popular Culture." *Journal of Map and Geography Libraries* 16 (3): 239-253. https://doi.org/10.1080/15420353.2021.1944948

Harris-Pierce, Rebecca L., and Liu, Yan Quan. 2012. "Is data curation education at library and information science schools in North America adequate?" *New Library World* 113(11/12): 598-613. https://doi.org/10.1108/03074801211282957

Herr, Melody. 2019. "Responding to research misconduct: A primer for LIS professionals." *Science & Technology Libraries* 38(3): 272-287. https://doi.org/10.1080/0194262x.2019.1644268

Higgins, Sarah. 2008. "The Lifecycle of Data Management." In *Managing Research Data*, edited by Graham Pryor, 17-46. London: Facet Publishing.

Johnston, Lisa R., Jacob Carlson, Cynthia Hudson-Vitale, Heidi Imker, Wendy Kozlowski, Robert Olendorf, and Claire Stewart. 2018. "How Important Is Data Curation? Gaps and Opportunities for Academic Libraries." *Journal of Librarianship and Scholarly Communication* 6: eP2198. https://doi.org/10.7710/2162-3309.2198

Jones, Sarah, Graham Pryor, and Angus White. 2013. "How to Develop Research Data Management Services - A Guide for HEIs." DCC How-to Guides. Last updated May 15, 2015. https://www.dcc.ac.uk/guidance/how-guides/how-develop-rdm-services

Kim, Jeonghyun. 2020. "Academic library's leadership and stakeholder involvement in research data services." Paper presented at the *83rd Annual Meeting of the Association for Information Science and Technology, Virtual Conference, October 22-November 1, 2020*. https://doi.org/10.1002/pra2.304

Kirkwood, Rachel Joy. 2016. "Collection Development or Data-Driven Content Curation? An Exploratory Project in Manchester." *Library Management* 37(4/5): 275-284. https://doi.org/10.1108/LM-05-2016-0044

Koltay, Tibor. 2019. "Accepted and Emerging Roles of Academic Libraries in Supporting Research 2.0." *The Journal of Academic Librarianship* 45(2): 75-80. https://doi.org/10.1016/j.acalib.2019.01.001

Lake, Sherry, Andrew Sallans, Barbara Pralle, David Fearon, and Betsy Gunia. "SPEC Kit 334: Research Data Management Services." Washington, DC: Association of Research Libraries, 2013. Available: ISBN:1-59407-902-1. http://publications.arl.org/Research-Data-Management-Services-SPEC-Kit-334/~~FreeAttachments/Research-Data-Management-Services-SPEC-Kit-334.pdf

Lau, Georgin, and Lei Pan. 2015. "A 5-step guide to data visualization." Elsevier Connect. Last modified January 28, 2015. https://www.elsevier.com/connect/a-5-step-guide-to-data-visualization

Lord, Philip, and Macdonald, Alison. 2003. "e-Science Curation Report - Data curation for e-Science in the UK: an audit to establish requirements for future curation and provision." *The JISC Committee for the Support of Research (JCSR)*. https://tinyurl.com/ydn6w6p3

Lynch, Clifford. 2003. "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age." *ARL: A Bimonthly Report on Research Library Issues and Actions from ARL, CNI, and SPRAC* 226: 1-7. https://bit.ly/3wUDwEI

Mischo, Bill. 2020. "Making metrics come alive with interactive data visualizations." UIUC Libraries Scopus APIs. Scopus. Accessed March 10, 2021. https://tinyurl.com/y8gj85az

National Academies of Sciences, Engineering, and Medicine. 2019. "Reproducibility and Replicability in Science." Washington, DC: The National Academies Press. https://doi.org/10.17226/25303

National Institutes of Health (NIH). 2020. "Final NIH Data Management and Sharing and Supplemental Information." Federal Register 85(211): 68890-68900. https://www.federalregister.gov/d/2020-23674

Nature. 2021. "Data Repository Guidance". Scientific Data Policies. https://www.nature.com/sdata/policies/repositories#general

National Science Foundation (NSF). 2019. "Transforming Science Through Cyberinfrastructure." Office of Advanced Cyberinfrastructure Vision. https://www.nsf.gov/cise/oac/vision/blueprint-2019

Ogier, Andrea, and Michael Stamper. 2018. "Data Visualization as a Library Service: Embedding Visualization Services in the Library Research Lifecycle." *Journal of eScience Librarianship* 7(1): e1126. https://doi.org/10.7191/jeslib.2018.1126

Primich, Tracy. 2010. "A Semester-Long Seminar in Statistica Visualization for Undergraduates as Taught by a Science and Engineering Librarian." *Science & Technology Libraries* 29(3): 181-188. https://doi.org/10.1080/0194262X.2010.497702

Rice, Robin, and John Southall. 2016. *The Data Librarian's Handbook*. London: Facet Publishing.

Schmidt, Birgit, and Kathleen Shearer. 2016. *Librarians' Competencies Profile for Research Data Management.* Joint Task Force on Librarians' Competencies in Support of E-Research and Scholarly Communication, Confederation of Open Access Repositories (COAR). https://www.coar-repositories.org/files/Competencies-for-RDM_June-2016.pdf

Schmidt, Lawrence, and Joseph Holles. 2018. "A Graduate Class in Research Data Management." *Chemical Engineering Education* 52(1): 52-59. https://journals.flvc.org/cee/article/view/105451

Smith II, Plato L. 2014. "Exploring the Data Management and Curation (DMC) Practices of Scientists in Research Labs within a Research University." PhD Dissertation, Florida State University. http://purl.flvc.org/fsu/fd/FSU_migr_etd-9095

Silver, Isabel D. 2014. "Outreach Activities for Librarian Liaisons." *Reference & User Services Quarterly* 54(2): 8-14. https://doi.org/10.5860/rusq.54n2.8

Steneck, Nicholas H. 2007. "ORI Introduction to the Responsible Conduct of Research." Revised ed. Washington (DC): Department of Health and Human Services. https://ori.hhs.gov/sites/default/files/2018-04/rcrintro.pdf

Tenopir, Carol, Ben Birch, and Suzie Allard. 2012. "Academic Libraries and Research Data Services: Current Practices and Plans for the Future." Association of College and Research Libraries (ACRL) White Paper. http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/Tenopir_Birch_Allard.pdf

Tenopir, Carol, Dane Hughes, Suzie Allard, Mike Frame, Ben Birch, Lynn Baird, Robert Sandusky, Madison Langseth, and Andrew Lundeen. 2015. "Research Data Services in Academic Libraries: Data Intensive Roles for the Future?" *Journal of eScience Librarianship* 4(2): e1085. https://doi.org/10.7191/jeslib.2015.1085

Tenopir, Carol, Jordan Kaufman, Robert J. Sandusky, and Danielle Pollock. 2019. "The Time Has Come… To Talk About Why Research Data Management Isn't Easy." Paper presented at the *Charleston Library Conference, Charleston, SC, November 4-8, 2019*. https://doi.org/10.5703/1288284317185

U.S. News and World Report. 2020. "Best national university rankings [Internet]. Accessed March 20, 2020. https://www.usnews.com/best-colleges/rankings/national-universities

Uzwyshyn, Ray. 2016. "Research Data Repositories: The What, When, Why, and How." *Computers in Libraries* 36(3): 18-21. https://digital.library.txstate.edu/handle/10877/7597

White, Ethan P., Elita Baldridge, Zachary T. Brym, Kenneth J. Locey, Daniel J. McGlinn, and Sarah R. Supp. 2013. "Nine simple ways to make it easier to (re)use your data." *Ideas in Ecology and Evolution* 6(2): 1-10. https://ojs.library.queensu.ca/index.php/IEE/article/view/4608

Wickham, Hadley. 2014. "Tidy Data." *Journal of Statistical Software* 59(10): 1-23. http://doi.org/10.18637/jss.v059.i10

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. "The FAIR Guiding Principles for scientific data management and stewardship." *Scientific Data* 3: 160018. https://doi.org/10.1038/sdata.2016.18

Yoon, Ayoung, and Teresa Schultz. 2017. "Research Data Management Services in Academic Libraries in the US: A Content Analysis of Libraries' Websites." *College & Research Libraries* 78(7): 920. https://doi.org/10.5860/crl.78.7.920