Potential Roles for Science Librarians in Research Data Management: A Gap Analysis

Bradley Wade Bishop
Associate Professor
School of Information Sciences
University of Tennessee
Knoxville, TN
wade.bishop@utk.edu

Hannah Rose Collier
Graduate Assistant
School of Information Sciences
University of Tennessee
Knoxville, TN
harmenda@vols.utk.edu

Ashley Marie Orehek
Instructional Librarian
Katie Murrell Library
Lindsey Wilson College
Columbia, KY
oreheka@lindsey.edu

Monica Ihli
Assistant Professor and
Liaison Librarian to ORNL
University Libraries
University of Tennessee
Knoxville, TN
mihli1@tennessee.edu

Abstract

As many sciences move to be more data-intensive, some science librarians are offering more research data services and perform research data management roles. Job analyses provide insight and context to the tasks employees actually do versus what their job descriptions depict or employers assume. Two separate job analyses studies investigated the roles and responsibilities of data services librarians and research integrity officers among the top 10 private and top 10 public higher education institutions. The focus of these interviews was research data

management support roles. Comparing these two groups' responses indicates that the role-based responsibilities for research data services are not always clear within institutions and are predominantly placed on individual researchers or research teams, but science librarians may provide some solutions to address this gap. This paper presents a model for the potential roles of science librarians in research data management.

Introduction

Research data management (RDM) regulations and roles in the sciences have changed greatly over the past 10 to 20 years. The impetus for an increased focus on RDM began when many (and now most) funding agencies across the globe required data management plans (DMPs) with grant proposals. With a heightened focus on reusing research data across all disciplines along with the DMP mandates, academic librarians responded with new research data services, including RDM training and DMP creation tools. With these new regulations altering the traditional research lifecycle for many sciences, science librarians also may need to alter their current services to include more related to RDM.

Fortunately, academic librarians are not alone as many research institutions have research infrastructure that includes dedicated employees to help with both research compliance and RDM. 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. Data Services Librarians (DSLs) across disciplines, but especially in data-intensive sciences, are a newer position in many academic libraries that facilitate RDM-related activities within specific sciences or campuswide. Although these roles exist, compared to the volume of individual researchers generating research data at each institution, the ratio to DSLs and RIOs is very low. This study addresses the need to gain a better understanding of the RDM tasks and perspectives of DSLs and RIOs through job analyses. The job analyses were conducted using interviews with job incumbents. These interviews gather current employee's training, current tasks, and perceptions on DMP implementation and evaluation. This paper compares two sets of job analyses of RIOs (Bishop et al. 2021) and DSLs (Bishop et al. date unknown a; Bishop et al. date unknown b). By reviewing the RDM-related tasks performed by each job, any gaps in DMP evaluation and implementation can be examined. Additional background on DSLs, RIOs, and DMPs are provided in the following literature review.

Literature Review

Data Services Librarians

Librarianship has historically been a support role driven by the information needs of whatever community was being served. Within academic institutions, science 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 assisting with locating resources, instruction, collection development, and expanding to "higher end support" roles such as being project managers and participating in project outputs (Corrall 2014; Brandenburg et al. 2017). In these situations, DSLs in the sciences are being included as personnel in grant budgets. Regardless of job titles, the roles and responsibilities of many liaison librarians, science librarians, and data service librarians are moving toward more collaborative models as research services, digital humanities, user experience, and scholarly communication responsibilities increase (Jaguszewski & Williams 2013).

In particular, DSLs are becoming increasingly responsible for leading RDM efforts at academic institutions (Rice & Southall 2016; Koltay 2019). Read and Cox (2020) investigated the technical competencies of academic librarians supporting scholarly communication. Through semi-structured interviews, the researchers determined that, "participants belonging to larger SC [scholarly communication] units were more likely to be specialized - that is, to have a role with fewer and more specific responsibilities, such as a focus on the IR [institutional repository] or RDM" for specific disciplines (Read & Cox 2020, p. 7). Among the goals of DSLs is to serve as agents to uphold research integrity (Herr 2019). In addition to helping with projects and proposals, many DSLs have begun offering RDM training. These instruction sessions range in detail from general activities to discipline-specific, and range in length from short, one-off sessions to term-length, for-credit courses (Carlson et al. 2015; Schmidt & Holles 2018). RDM training is offered at a number of libraries across disciplines, with some also covering RCR (Herr 2019; Gunderman 2020).

Research Integrity Officers

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 (e.g., RCR). Since most campuses have only one RIO, many researchers and science librarians may not be aware of this integral role in the research enterprise at their institutions. Once the U.S. Public Health Service began requiring RIOs for funded research in 1989, a major task for the first generation of RIOs was to create policies and procedures to handle research misconduct. The position became a necessity to work through any misconduct allegations on campuses (Wright & Schneider 2010). The U.S. Public Health Service regulations initially defined research misconduct as "fabrication, falsification, plagiarism, or other practices that seriously deviate from those that are commonly accepted within the scientific community for proposing, conducting, or reporting research" (Policies of General Applicability 1989). Today, the U.S. Office of Science and Technology Policy definition of research misconduct still includes these three behaviors: (1) Fabrication of results or data; (2) Falsification of data through changing or omitting data or results such that the research is not accurately represented in the research record; or (3) Plagiarism (Steneck 2007).

In practice, Offices of Research Integrity do any number of oversight activities as they operate with no formal or universal standard for defining their scope of responsibility. One survey covering roles and responsibilities (n=56) found that nearly all RIOs identify themselves as participants in drafting and maintaining policies and procedures dealing with research misconduct. Other common RIO duties include determining whether an investigation is warranted, sequestration of data relevant to an investigation, protection of whistleblowers, reporting allegations to funding agencies and other officials, and RCR training to educate all researchers to avoid misconduct (Wright & Schneider 2010). A smaller (n=12) qualitative study's role-based questions revealed results largely consistent with those previously noted. In addition, the responses indicated that although data are central to research today, it may not be a part of RIOs' training (Bishop et al. 2021). In some instances, subject specialist librarians participate in research misconduct inquiries and investigations to offer their scholarly communication expertise to detect plagiarism, as well as data fabrication and falsification. All RIOs stand to benefit from more awareness of this untapped expertise on their campuses.

For purposes of this study, inquiries and investigations into potential research misconduct are considered the primary activity of all RIOs. Much of research misconduct involves examining research data. DMPs may be used as guides to assist in those research data tasks. A recent survey

inquiring about the most recent investigation of misconduct for 24 RIOs found that poor RDM practices were present in 71% of cases (Kalichman 2020). Additionally, numerous examples of data-related misconduct cases further solidify the RIO's role as a stakeholder in the research ecosystem of RDM.

The National Science Foundation (NSF) Office of the Inspector General provides an annual report on research misconduct cases, but excludes investigations of research misconduct found to be unintentional (https://www.nsf.gov/oig/reports/) (e.g., National Science Foundation 2019; National Science Foundation 2020a; National Science Foundation 2020b). There are likely more instances never reported. Fabricated and manipulated data are common causes of research misconduct findings. The consequences of misconduct findings can be devastating to the careers of individuals and burdensome to each institution.

To help reduce the likelihood of some research misconduct, one study found that DMP audits resulted in an overall positive impact for researchers through improved RDM (Lee et al. 2019). The lack of adequate DMP implementation or evaluation throughout the research lifecycle may lead to a lack of compliance down the road and increase the likelihood of research misconduct.

These issues raised may be addressed by DSLs working in sciences as RDM allies on campuses. Many science librarians have already gained the knowledge and skills to become involved in DMP creation and implementation. This study aims to describe the current practices for both DSLs and RIOs and highlight potential gaps in RDM across campuses that could be addressed with a model for science librarianship that includes additional services and resources. Additional details of DMPs help to contextualize the crux of the model.

Data Management Plans

DMPs or similar documents by other names (i.e., Data Sharing Plans) have been required by the National Institutes of Health since 2011 for grants greater than $500,000 and the NSF since 2011 for all projects. DMPs are short (e.g., NSF limits to two pages), formal documents that describe the ways in which the researchers will manage research data generated during projects. DMPs describe the roles, responsibilities, and activities for managing data during and after research (Bishop & Hank 2020). DMPs typically address topics such as file management, file types, backup and security, metadata, and sharing and access of data. Almost all US federal agencies and most private foundations currently implement a DMP requirement.

Williams et al. (2017) reviewed DMP requirements from different research funders. With differences in requirements and the novelty of DMPs for some researchers, there has been confusion on the proper creation and content for DMPs as well as what services might be needed for implementation. For example, one study evaluated 119 DMPs and found that 51% did not identify the individual(s) responsible for RDM (Van Loon et al. 2017). These studies have identified that researchers need help and guidance to create quality DMPs for their projects. DMP creators must have familiarity with disciplinary norms, extending to typical funding requirements pre and post award. Those writing DMPs must also have knowledge of various DMP tools and templates across funding agencies to capture important details. Those involved in DMP review and support of their implementation must also stay up-to-date on these issues, including relevant data standards and processes (Cox & Verbaan 2018). The two groups of professionals investigated in this study (DSLs and RIOs) are both perched in support roles to help with policy, documentation, and science domain knowledge assisting researchers with creating and implementing DMPs.

Objectives

This study is part of a series of studies examining the current roles and perspectives of DSLs and RIOs on DMPs. Comparing responses from these two groups may provide insights into the role-based responsibilities for research data management services, as well as ways to improve services to meet the needs of their respective university communities. To that end, the authors sought to explore the following research questions in the context of U.S. universities:

1. How are DMPs implemented and evaluated?
2. How do DSLs and RIOs support DMP implementation and evaluation?
3. What institutional support exists for RDM and RCR activities?
4. What similarities and differences exist in RDM support by DSLs and RIOs?
5. What institutional supported RDM and RCR training do DSLs and RIOs receive and/or deliver?

Methods

Ten DSLs and twelve RIOs were interviewed using separate semi-structured interview questionnaires (Bishop et al. 2021; Bishop et al. date unknown a; Bishop et al. date unknown b). This study received Institutional Review Board approval prior to data collection (Bishop 2020a; Bishop 2020b). The DSLs' job titles reflect mostly science librarians or data services librarians that serve science [e.g., Data Librarian (2), Science Data & Engineering Librarian, Sciences Data Librarian, Data Science Librarian, Research Data Librarian, Computational Research Librarian, Research Data Management Librarian, Numeric Data Services and Data Management Librarian, and Senior Research Data Management Consultant]. When asked about their subject area coverage, five participants mentioned that they usually worked with specific subject disciplines such as social sciences (2); engineering, physics and other STEM-related disciplines (2); and life sciences. The remaining five participants supported all disciplines on campus.

Both sets of interviews covered topics related to the full range of job responsibilities with 24 total questions for DSLs and 30 total questions for RIOs. Twelve questions from both interview schedules aligned and covered RDM concepts (e.g., DMPs, data storage, data management costs, and training). This study presents a gap analysis of the similarities and differences between responses to those RDM questions to inform a model for research data services to support a RCR campus culture and reduce the potential for research data misconduct. Table 1 lists the RDM questions.

DSLs and RIOs were recruited from the top 10 public and top 10 private universities according to 2020 Best National University Rankings (US News and World Report 2020). Some universities have multiple RIOs or DSLs, but at least one from each institution was contacted.

Table 1. Research data management survey question comparison

Research Integrity Officer Questions	Data Services Librarian Questions
Data Management Plans	
Do you have any oversight of data management plans?	If you assist with data management plan review, please provide a few examples of that work.

Research Integrity Officer Questions	Data Services Librarian Questions
Who is responsible for data management plan compliance?	Who is responsible for data management plan implementation at your institution?
How are data management plans evaluated for compliance?	How are data management plans evaluated at your institution?
If you were creating an office of integrity, what would be the ideal oversight structure and process for data management plans?	What would be the ideal structure and process for data management plan implementation and evaluation?
Data Storage	
Does your institution have any ownership or disposition of data policies?	Does your institution have any data policies?
Does your institution support any institutional repositories for data?	If you manage an institutional or digital repository, please provide a few examples of that work.
Who is primarily responsible for the long-term management of the data for sponsored projects?	What is your institution's commitment to the long-term management of research data? In your institutional repository? In digital repositories? For all other data?
Data Management Costs	
How are data management efforts for sponsored projects at your institution funded?	How are data management efforts for research projects at your institution funded for sponsored projects and/or all other projects?
What budget allocated exists for long-term data management beyond the life of projects and grants?	What budget allocated exists for your institutional or digital repository? Personnel? Other infrastructure?
Training	
Does your office provide responsible conduct of research (RCR) training?	Does someone at your institution offer RCR training? Do you have a role?
Does your office provide data management training?	If you provide research data management training, what types of research data management training do you offer?
Have you received any research data management training? If yes, what types of research data management training did you receive?	What types of research data management training have you received?

After contacting 20 DSLs, interviews via Zoom were conducted with five participants from public and five from private universities. Of the total 20 RIOs contacted, only three RIOs from private universities and nine RIOs from public institutions were interviewed via Zoom or in-person (February 2020 through March 2020). The National Universities Rankings as a sampling

frame was used because those top institutions emphasize faculty research as the result of large research expenditures and are more likely to also have more researchers with RDM needs.

Interviews were recorded, transcribed, and coded in NVivo using a grounded theory application of open, axial, and selective coding to capture their job tasks and perspectives on RDM (Glaser & Strauss 1967). The grounded theory application of coding generates categories and broad themes based on the responses using synonymous meanings. For each question the coding steps involved reading the transcripts, assigning a code for a response, then adjusting the application of the code as differing or similar responses emerged.

The informed consent that participants agreed to included open data language with anonymized and deidentified transcripts that are now available. Informed consent was obtained from participants prior to their interviews, and data from both groups are available through the University of Tennessee institutional repository (TRACE: Tennessee Research and Creative Exchange) (Bishop 2020a; Bishop 2020b).

Results

Data Management Plans

Questions regarding DMPs inquired whether participants assisted with writing DMPs, implementing them, or evaluating them for compliance. Each participant was also asked an open-ended question about their ideal structures regarding DMP implementation, evaluation, and compliance. Responses among DSLs about writing and reviewing DMPs for proposals varied with most not involved in this part of the research process. Seven participants reviewed plans occasionally, two stated they did upon request, and one participant said it was their primary role. Four worked with faculty to review plans prior to grant submission and performed some kind of DMP evaluation. These evaluation procedures involved checking for completeness and being a part of joint or sole peer reviews prior to grant submissions. Five DSLs were not involved with evaluation and one mentioned that it was up to the researcher to determine if they needed outside help with this portion of the grant proposal.

None of the RIOs interviewed were responsible for ensuring DMP writing, implementation, or compliance, and each assumed other entities at their universities had this responsibility. In fact, none of the 12 RIOs were involved in any part of DMPs. Three RIOs stated that this responsibility fell on other parties: the compliance officer (1), the library (1), or the funder (1). DSLs and RIOs both assumed individual researchers were responsible and acknowledged that RDM support was likely underfunded. Two DSLs and eight RIOs did not know how the DMPs were evaluated. Furthermore, some participants stated if a researcher does not request or require input during the proposal process, there is little incentive to revisit or reassess the document after funding commences.

All around, DSLs and RIOs did not formally enforce DMP implementation, leaving that responsibility to Principal Investigators or other researchers. Some DSLs mentioned that journals and funders served as the DMP implementer. One DSL stated that academic departments handled that responsibility, although their library is contemplating if they "should be better at implementing enforcement" (DSL-8). Another DSL was, at the time of interview, working with their Office of Research to standardize the process. A third DSL stated that, "We've had a number of issues where [researchers], the day before [funding agency reports are due], [are] trying to prove they followed through on their DMP. There's a carrot [but] there's no stick yet"

(DSL-9). Similarly, one RIO participant summed it up as an unfunded mandate that no one makes time for because they are not enforceable (RIO-5).

As ancillary players in the research enterprise, DSLs and RIOs have useful perspectives on what works and what does not when it comes to successful and compliant DMPs. These entities from the offices of research and academic libraries could work in a more coordinated manner. Both groups shared their insights for the ideal structure for DMPs with vastly different responses among both groups. Overall, many talked about growing the educational awareness of RDM and increasing funding to support these RDM services and resources. Other goals mentioned included decreasing disconnect, streamlining processes to decrease conflicts, and identifying appropriate subject-disciplined repositories for storage rather than defaulting to their institutional repository. Two DSL participants wished for a single intake portal, three others suggested that all researchers take in-depth RDM training beforehand or as a refresher, and two additional librarians suggested that academic libraries should be informed when grants are funded so they may start assisting researchers at the start of a projects to help implement their DMPs rather than attempt to salvage data at the end. Science librarians could assist with many of these tasks. Having more centralized control of data, including data from research misconduct inquiries and investigations would also help streamline any falsification and fabrication investigations through better reporting and tracking of every step in the research process. This appears in one RIO's suggestion for "an advisory office, aware of what federal expectations are for these that could be advisory to the Principal Investigators" (RIO-10). Emphasizing issues and potential pitfalls for RDM and incorporating RDM into RCR training would also help to deter data-related misconduct.

Data Storage

DSLs and RIOs were also asked about research data storage at their institutions. These questions related to (1) the existence of institutional data policies, (2) an institutional commitment to long-term data management, and (3) any institutional support for an institutional or digital repository. All of the RIOs acknowledged the existence of an intellectual property data policy at their universities, whereas only one DSL confirmed a university-wide data policy with the remaining either having no known policy or a data policy that was outdated. These policies set the standard among students and researchers at institutions as to how to store their data. If DSLs are unaware of them, then they will not be able to advise users to avoid problems that may arise over copyright and ownership. These concerns appear to be more aligned with RIO roles, which may explain their increased awareness; however, DSLs and science librarians working with locating and managing research data should be more involved in the crafting and updating of data policies to encourage compliance with them. Academic librarians are involved in intellectual property protection and education for other types of information on campus, therefore data is an adjacent job task and may require more consideration.

A majority of the RIOs said that long-term commitment to RDM was up to the individual researcher. Again, this was another area where the DSLs were varied in their responses. A general consensus was that institutions were responsible, but there were varying definitions of "long-term" storage. Three DSL institutions were enthusiastic about their university-wide long-term data commitment, making sure money was allocated, backup measures were in place in case of an emergency, and keeping detailed records. Three DSL institutions had some commitment, promising storage in-perpetuity but often reevaluating after 5-10 years. One DSL institution had no commitment and three were unsure of any commitment because of loose promises of in-perpetuity, no requirements, or that no institutional repository existed within the

institution. With PIs again held to plan for long-term RDM from the RIOs' perspectives and DSLs split or unsure of institutional commitment, further work needs to be done to preserve research data or at least understand the current workflows used by researchers.

Among the ten DSLs interviewed, eight confirmed the existence of an institutional repository; however, only one of those eight actually managed a repository. This participant hesitated acknowledging a supporting budget's existence saying, "they had one developed or [are] working on it," and assumed the library intended to "manage it in-perpetuity" (DSL-10). Eleven of the twelve RIOs surveyed confirmed an institutional repository, and ten of them stated that it was the duty of the researcher to coordinate long-term RDM. Depositing into institutional repositories sometimes costs money and it is one thing omitted from grant and other university administrative budgets. One way to address any unfunded mandate is to have centralized bodies, like academic libraries and/or information technology, within a university absorb the costs.

Participants mentioned the possibility of having either libraries (4), information technology departments (4), or research departments (3) help with these processes. Two participants hoped their Vice Provost/President for Research or someone in sponsored projects managed this issue. One RIO suggested funding agencies provide a way to identify the already-established data repositories by discipline so researchers would not rely on their institutional repositories for storage. DSLs and RIOs all looked for pragmatic solutions to potential data storage issues, but just like with DMPs, ultimately researchers must have the impetus and take control of their own data storage. This likely includes any incurred costs.

Data Management Costs

Funding questions asked to all participants related to how each institution (1) supported data management efforts for sponsored projects and all other projects and (2) budgeted for the long-term data management beyond the life of projects and grants, including personnel and infrastructure.

Neither DSLs or RIOSs had a clear understanding regarding project data management support budgets. The majority of RIOs assumed projects already supported data management and a majority of the DSLs assumed a university entity (e.g., Office of Research) supported data management, although the libraries' funding sources typically remained unknown.

Although two participants were unsure, 10 out of 12 RIOs assumed that sponsored projects or some other university-level entity controlled RDM support for projects and grants. A majority of DSLs were unsure of a budget amount or even the existence of one. Only one DSL participant gave a specific dollar amount, estimating $150,000 out of a $3 million library budget was set aside for RDM. Some RIOs voiced concerns that RDM costs would exceed the budget of each individual project and some university funds would end up supporting long-term RDM efforts with comments like "my understanding is that grants rarely cover all of it" (RIO-3). In order for library services and resources to successfully help with RDM, these university-wide budgets should be known, if they exist, and include supplemental RDM funds for related data services offered by academic libraries. Both of these populations would better serve in their roles if they knew more about their own institutions budgeting for RDM.

Training

Participants were also asked questions related to training and education including, (1) types of RCR training offered; (2) types of RDM training offered; and (3) their own RCR/RDM training. All DSLs and seven RIOs responded that they had received some sort of RDM training. Both populations noted that a majority of this training occurred in informal settings and on their own time. As methods and strategies evolve over time, many said they struggle with keeping up with the changes. As one participant put it, the rapid changes could jeopardize research integrity should the research data manager not stay current with the RDM evolution. One DSL stated, "I think my library assumed that, if you have your own data, you should be good at teaching it" (DSL-9). Another participant felt they were forced to learn because "the NSF [started enforcing] data management plans" (DSL-5). Not receiving the support or encouragement to stay current on training is a common problem among both groups. Surprisingly, five RIOs said they had not received any RDM training. As the research data lifecycle relates to some aspects of potential data fabrication and falsification, RIOs should be more aware of these aspects of the research enterprise. In general, a greater support of such training is needed to encourage parties involved in RDM to stay up to date with the latest RDM practices and requirements.

Nine DSLs and only two RIOs provided RDM training. RIOs that did participate in teaching RDM said that this was merely a part of their provided RCR training. One DSL and six RIOs said their institutions did not offer RDM training. Four other RIOs stated that the training was done in other departments, some mentioning that training occurred in the library or computer science departments. DSLs that did provide RDM training tried various teaching methods including on-campus workshops, one-on-one sessions, and promotional periods (e.g., Data Week). There was an overall sentiment from DSLs who struggle with engaging users because, as important as RDM is, it is not an appealing aspect of research compared to the more exciting parts of the research lifecycle.

Eleven of the twelve RIOs stated that they do participate in RCR training. Three were in charge of this training, and eight were involved though not the lead organizers. The DSLs were less involved in RCR training. Three DSLs said this training was provided by their office of research, three said the library was not involved, and three others were unsure if RCR training was provided at all. Only one DSL said that the library provided this service, and that it was a new development. A couple of participants stated that this responsibility was covered by the various individual departments. For example, one RIO said that RCR training was not completed within their department as it had been decentralized within their institution. One DSL stated something similar, "most of the formal training is done through the departments but the librarians talk about ethics in data and research on a daily basis" (DSL-8). Clearly, RCR training falls more closely to the role of RIO in promoting a culture of compliance and RDM training is more closely related to DSLs job responsibilities supporting aspects of data management on campuses. Overall, more coordination in training across campuses would benefit both groups with data central to nearly all endeavors.

Discussion

This comparison between two different university roles shows some common gaps in RDM practices across top-ranked institutions. There are limitations to this qualitative approach, including participants' various experiences limited perspectives of the totality of RDM on their campuses, and fewer participants overall from private institutions. Still, from 10 DSLs and 12

RIOs, there appears to be an assumption that researchers handle most of their own RDM needs beyond initial training. Researchers also deal with DMP implementation and evaluation throughout the life of any project based on these DSLs. Although many RDM efforts do not relate to their roles, there is a potential for collaboration across the research enterprise and responsibilities and roles for science librarians that complement current and emerging practices. The RDM tasks could be performed and assisted through many entities on campus, such as offices of research, information technology, research labs, academic departments, and individual researchers. With data becoming central to many sciences, academic librarians supporting faculty, students, and staff could expect to collaborate more in RDM.

By asking structurally similar questions, this study compared the perspectives of DSLs and RIOs on topics of RDM, found gaps, and will present a model for the potential roles of science librarians.

To review, RIO job responsibilities stem from the federally mandated nature of their work and this is reflected in their responses. In the interviews, RIOs were more likely than DSLs to mention funding agencies, compliance issues, and federal regulations when describing their ideal conceptualizations of how RDM should operate. This is unsurprising given the scope of a RIO's primary responsibilities to their organization. Conversely, the core values of librarianship emphasize such priorities as preservation, equitable access, and public good (American Library Association 2019). The gaps that emerge from the responses to the RDM-related questions show researchers may need expertise at both the start and end of their projects. DSLs and RIOs both have roles to play during proposal writing, DMP writing, and RCR considerations for planning data collection. They both also have roles to play at the end of projects with compliance and preservation considerations. Several DSLs did speak of DMPs in the context of proposal requirements; however, DSL motivations or goals for DMPs often emphasized such incentives as simplifying workflows and improving success for the researcher, or supporting open science, and not mentioned are the aspects related to access for reuse and other compliance considerations that are the emphasis of RIOs' work.

Across both groups of participants, a desire to improve cross-institutional collaboration among campus functional areas was described as both serve in liaison roles reaching out to other parts of their institutions. These decentralized related research services, such as an office of research or academic library, are not typically embedded into research centers or academic departments and may be seen as outsiders to the actual research enterprise work of grant writing, data collection, and analyses. The idealized goals for strengthening partnerships tended to vary among RIOs versus DSLs. RIOs were more likely to describe a desire for greater collaboration with IT and increased access to information technology-related tools needed to investigate misconduct (e.g., iThenticate). DSLs also cited the importance of relationships with IT, but overwhelmingly described a desire to have the academic library be more integrated with other research-oriented units or administrative workflows, such as offices of research. Several DSL participants described in detail their beliefs that greater integration would increase the use of and streamline their research data services. The variations that emerged from these two groups may simply be yet another reflection of the underlying differences in their units' responsibilities and missions.

Despite any differences in their perspectives, DSLs and RIOs shared many of the same concerns surrounding DMPs, such as the need for better supporting cyberinfrastructure, lack of adequate budget and staffing, and the need for additional training. Generally, many institutions either lack or do not emphasize the importance of RDM training for both DSL and RIO roles.

For institutions to maintain compliance with funding agency requirements, those currently serving in support roles must be able to incentivize RDM training for themselves and those across campus working with research data. This is particularly important to adequately train those employees who provide RDM and RCR training to students or researchers. Several participants from both the DSLs and the RIOs acknowledged that specific university guidelines regarding RDM policy are either missing or poorly advertised within many institutions, indicating that consistent guidelines would be favorably viewed from each of their respective positions.

Among the RIO participants, there appeared to be some confusion over basic RDM terminology: "It kind of depends on what you mean by a data management plan" (RIO-12). This may reflect those faculty or staff assigned this administrative role without actual awareness of this relatively recent research requirement. In addition, five RIOs received no RDM training. This is surprising because a DMP would be a roadmap for any inquiry related to data fabrication or data falsification since a DMP describes the roles and activities for managing data during and after research. With clear regulations for records management related to research misconduct, RIOs know exactly how long storage is expected (e.g., seven years was mentioned most by participants in this study). Similar regulations are needed for each discipline and every institution to inform the preservation of research data. The model illustrated in Figure 1 is an attempt to visualize the gap and potential roles and responsibilities for science librarians.
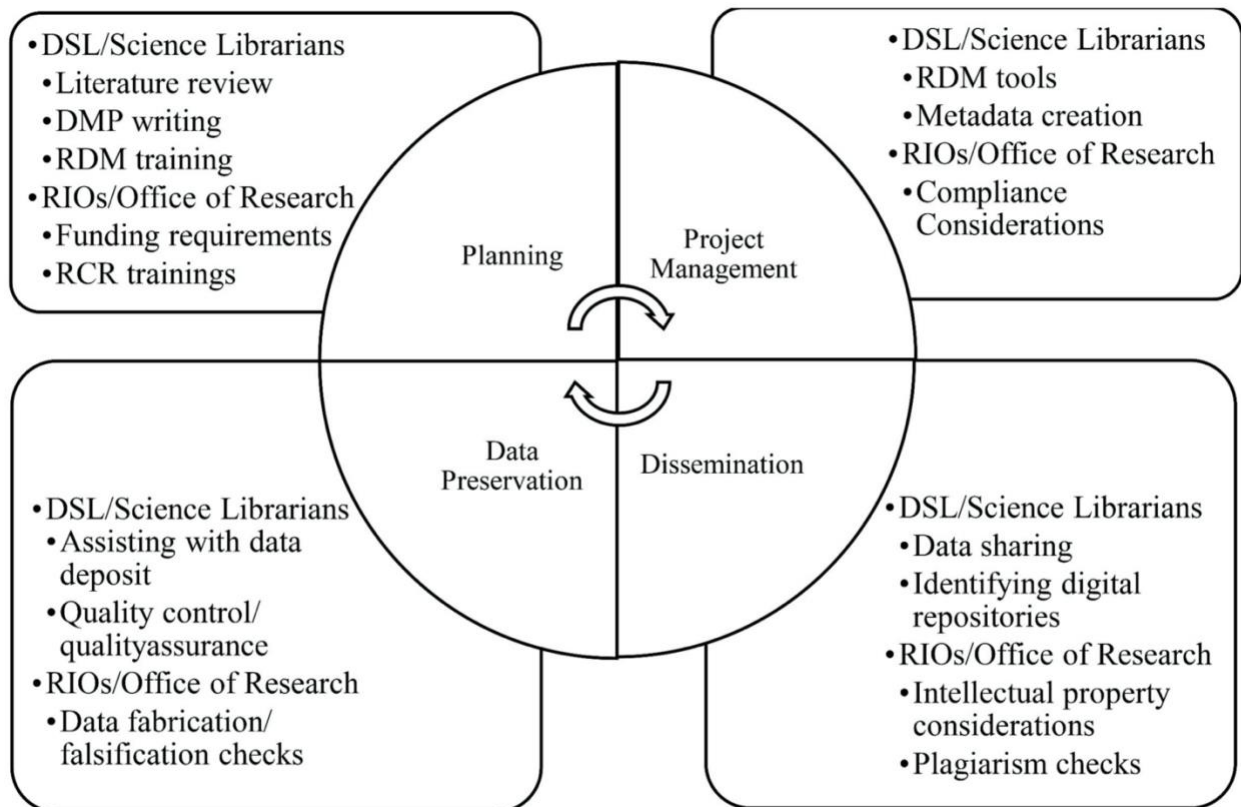

Figure 1. Model of RDM and RCR service responsibilities and roles

With DSLs and RIOs contributing to a culture of compliance in regard to RDM throughout the research lifecycle, there are entry points for both liaisons to provide input to scientists at the right moments in the lifecycle. The planning stage of any externally funded project that triggers the

DMP requirement begins with seeking out opportunities and grant writing, including budgets. The RIO responsibilities are limited in the planning process as are those for librarians, but a few roles exist. Some science librarians may assist by conducting literature reviews and providing access to information resources. Although not directly a RIO role, some office of research staff assist with grant writing and submission. DSLs specifically help with DMPs. Both do and could do more RDM and RCR training together or coordinate to reach students and faculty with these considerations prior to planning any project.

In the project management stage of any research project, the Principal Investigators take the lead and work with a team to collect data. In some instances, DSLs' time could be formally budgeted to ensure data collection with reuse in-mind through standardized metadata creation during a project. Even if not on a research team, science librarians and DSLs may assist scientists through promoting and educating them on metadata standards, common ontologies, and new RDM collaboration tools for them to consider for research projects. A recommendation from this study is that DSLs and RIOs are notified when funding is received as an opportunity to congratulate and remind researchers of the support services offered by these entities at each university. Small interventions at this stage could decrease any RDM and RCR disconnect, which may decrease conflicts or actual research misconduct later in a project.

During the dissemination stage of the research lifecycle, DSLs and RIOs have less involvement. Researchers present, publish, and share data to disseminate their findings through scholarly communication. Science librarians could assist with locating venues for publications and presentations, but many scholars may not need this type of assistance. RIOs and other staff from the Office of Research may refer researchers to data policies and other intellectual property considerations at this time.

Science librarians or even RIOs might offer training on issues related to plagiarism or provide access to plagiarism checkers prior to submissions; however, publication norms vary by discipline and plagiarism checkers have their limitations. Such an approach could serve as a preemptive measure to reduce research misconduct though none of the participants mention such services. Some institutions have added these checks to theses and dissertations in an effort to reduce future issues.

Finally, the data preservation stage of the research lifecycle presents many more responsibilities for both librarians and RIOs. DSLs and science librarians should assist researchers with data deposit through identifying appropriate subject-disciplined repositories or institutional repository for long-term storage. DSLs could coordinate with RIOs at the time of any data deposit as an opportunity to conduct data quality control and quality assurance checks. If issues arose when data are deposited, the DSL could work with the researchers and RIO to resolve. At the very least, such issues could be documented for any future misconduct inquiries and investigations. The work of RIOs inquiries into data fabrication and data falsification issues might be handled entirely at this stage in the research lifecycle of every project sharing data.

Overall, DSLs and RIOs could be more integrated in the research lifecycle given their valuable roles and related RDM responsibilities. If librarians and RIOs were more involved from the onset to the sunset of a grant then all may benefit through a more streamlined process, especially in regard to RDM. The model only shows potential roles for librarians and RIOs, but many others could be enlisted across campus as partners (e.g., IT departments). Throughout the research lifecycle, the opportunity for collaboration between all to coordinate RCR and RDM trainings is also a consideration. In order for library services and resources to successfully help with RDM,

these university-wide assumptions over long-term research data storage and support require further study and financial backing. Investments in the research data infrastructure could come from overhead costs deducted from all external grants and then earmarked for these RDM purposes whether housed within academic libraries or not.

Conclusion

Best RDM practices and procedures differ across institutions and disciplines, but as they form a solid RCR focus, may reduce some potential research misconduct. It was not until the early 2010s that the NSF and other US agencies began requiring DMPs. After a decade of growth, this new step in the research lifecycle should begin to standardize practices. Unfortunately, while standards are evolving, individual researchers or research groups are often responsible for creating and following DMPs for their research. This on-the-fly RDM without expertise and training may lead to unintentional data misuse or mismanagement, as evidenced by investigations into research misconduct cases (Kalichman 2020). In examining the views of both DSLs and RIOs, this study provides a better understanding of how RDM practices, policies, and infrastructure connect across organizational units in an academic research institution. The drivers, responsibilities, and contexts experienced by different parts of a university uniquely contribute to how they interface with both data and the researchers that produce them.

While DSLs and science librarians, given the proper RDM training, are poised to help researchers maintain research data in a compliant manner, researchers may be unaware or unwilling to use these research data services. Gola and Martin (2020) explored the intersection of emotional intelligence and communities of practice, where significant change comes from not rewiring intellectual learning but through behavioral change as a social process familiar to scientists and researchers. This social and behavioral change could begin with RDM and RCR training done in concert and conducted by both librarians and RIOs, either partnering or complementing each other's data strengths. To make this possible, both DSLs and RIOs should be fully and continually trained on RDM and RCR best practices and changing standards. Standardized RDM and RCR training for researchers must then emphasize the importance of creating actionable DMPs within the project proposal stage and following through if funded throughout the research lifecycle. Such training should include warnings, highlighting the dangers of poor data management that have resulted in research misconduct cases. During the training, researchers should be shown how utilizing their academic research library and office of research personnel helps ensure their research project is conducted responsibly and avoids the pitfalls and financial costs associated with research misconduct. The model presented in this paper provides just a few RDM and RCR responsibilities and roles librarians and RIOs could (and in some cases do) offer.

Supporting Information

Deidentified data for this article can be found in the Tennessee Research and Creative Exchange (TRACE) institutional repository (Bishop 2020a; Bishop 2020b).

References

American Library Association. 2019. Core values of librarianship [Internet]. Available from https://www.ala.org/advocacy/intfreedom/corevalues.

Bishop, B.W. 2020a. Data management plan compliance and evaluation [Data Set]. Knoxville (TN): University of Tennessee Libraries. DOI: 10.7290/pebuWHcq7L.

Bishop, B.W. 2020b. Data services librarians. [Data Set]. Knoxville (TN): University of Tennessee Libraries. DOI: 10.7290/M29YHy5qEN.

Bishop, B.W. & Hank, C. 2020. Digital curation. In Kobayashi, A., editor. *International Encyclopedia of Human Geography*, 2nd ed. Amsterdam: Elsevier. p. 323-328. DOI: 10.1016/B978-0-08-102295-5.10531-1.

Bishop, B.W., Nobles, R. & Collier, H. 2021. Research integrity officers' responsibilities and perspectives on data management plan compliance and evaluation. *Journal of Research Administration.* 52(1):76-101.

Bishop, B.W., Orehek, A.M. Eaker, C. & Smith, P. [date unknown a]. Data services librarians' responsibilities and perspectives on research data management: Discussion of the results, challenges, and opportunities. *Journal of eScience Librarianship.* [Under Review].

Bishop, B.W., Orehek, A.M., Eaker, S. & Smith, P. [date unknown b]. Data services librarians' responsibilities and perspectives on research data management: The context and data collection. *Journal of eScience Librarianship.* [Under Review].

Brandenburg, M.D., Anderson Cordell, S., Joque, J., MacEachern, M.P. & Song, J. 2017. Interdisciplinary collaboration: Librarian involvement in grant projects. *College & Research Libraries.* 78(3):272–282. DOI: 10.5860/crl.78.3.272.

Carlson, J., Johnston, L.R. & Westra, B. 2015. Developing the data information literacy project. In Carlson, J. & Johnston, L.R., editors. *Data Information Literacy: Librarians, Data, and the Education of a New Generation of Researchers.* West Lafayette (IN): Purdue University Press. p. 35–50.

Corrall, S. 2014. Designing libraries for research collaboration in the network world: An exploratory study. *LIBER Quarterly.* 24(1):17–48. DOI: 10.18352/lq.9525.

Cox, A. & Verbaan, E. 2018. *Exploring Research Data Management.* London (UK): Facet Publishing.

Glaser, B.G. & Strauss, A.L. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research.* Chicago (IL): Aldine.

Gola, C.H. & Martin, L. 2020. Creating an emotional intelligence community of practice: A case study for academic libraries. *Journal of Library Administration.* 60(7):752–761. DOI: 10.1080/01930826.2020.1786982.

Gunderman, H. 2020. Lesson plans for teaching spatial data management in academic libraries through a lens of popular culture [Institutional Repository]. Pittsburgh (PA): Carnegie Mellon University; [accessed 2021 Mar 23]. DOI: 10.1184/R1/13350428.v1.

Herr, M. 2019. Responding to research misconduct: A primer for LIS professionals. *Science & Technology Libraries.* 38(3):272–287. DOI: [10.1080/0194262X.2019.1644268](10.1080/0194262X.2019.1644268).

Jaguszewski, J. & Williams, K. 2013. New roles for new times: Transforming liaison roles in research libraries [Report]. Washington (DC): Association of Research Libraries. Available from https://hdl.handle.net/11299/169867.

Kalichman, M. 2020. Survey study of research integrity officers' perceptions of research practices associated with instances of research misconduct. *Research Integrity and Peer Review.* 5:17. DOI: [10.1186/s41073-020-00103-1](10.1186/s41073-020-00103-1).

Koltay, T. 2019. Accepted and emerging roles of academic libraries in supporting research 2.0. *The Journal of Academic Librarianship.* 45(2):75–80. DOI: 10.1016/j.acalib.2019.01.001.

Lee, C.L., Goh, G.S. & Ali, Y.A. 2019. Effectiveness of data auditing as a tool to reinforce good Research Data Management (RDM) practice. *Abstract Book of the 6th World Conference on Research Integrity.* 2019 Jun 2–5; Hong Kong. Available from https://wcrif.org/images/2019/PDF/Abstract_book.pdf.

National Science Foundation. 2019. Semiannual report to Congress: April 1, 2019 – September 30, 2019 [Report]. Report No.: NSF-OIG-SAR-61. Washington (DC): National Science Foundation, Office of the Inspector General. Available from https://www.nsf.gov/oig/_pdf/NSF_OIG_SAR_61.pdf.

National Science Foundation. 2020a. Semiannual report to Congress: October 1, 2019 – March 31, 2020 [Report]. Report No.: NSF-OIG-SAR-62. Washington (DC): National Science Foundation, Office of the Inspector General. Available from https://www.nsf.gov/oig/_pdf/NSF_OIG_SAR_62.pdf.

National Science Foundation. 2020b. Semiannual report to Congress: April 1 – September 30, 2020 [Report]. Report No.: NSF-OIG-SAR-63. Washington (DC): National Science Foundation, Office of the Inspector General. Available from https://www.nsf.gov/oig/_pdf/NSF_OIG_SAR_63.pdf.

Policies of General Applicability – Definitions. 42 C.F.R. Sect. 50.102 (1989).

Read, A. & Cox, A. 2020. Underrated or overstated? The need for technological competencies in scholarly communication librarianship. *The Journal of Academic Librarianship.* 46(4):102155. DOI: 10.1016/j.acalib.2020.102155.

Rice, R. & Southall, J. 2016. *The Data Librarian's Handbook.* London (UK): Facet Publishing.

Schmidt, L. & Holles, J. 2018. A Graduate class in research data management. *Chemical Engineering Education.* 52(1):52–59.

Steneck, N. 2007. ORI introduction to the responsible conduct of research [Report]. Revised ed. Washington (DC): Department of Health and Human Services. Available from https://ori.hhs.gov/sites/default/files/2018-04/rcrintro.pdf.

U.S. News and World Report. 2020. Best national university rankings [Internet]. [accessed 2020 Mar 20]. Available from https://www.usnews.com/best-colleges/rankings/national-universities.

Van Loon, J.E., Akers, K.G., Hudson, C. & Sarkozy, A. 2017. Quality evaluation of data management plans at a research university. *IFLA Journal.* 43(1):98–104. DOI: 10.1177/0340035216682041.

Williams, M., Bagwell, J. & Zozus, M.N. 2017. Data management plans: The missing perspective. *Journal of Biomedical Informatics.* 71:130–142. DOI: 10.1016/j.jbi.2017.05.004.

Wright, D.E. & Schneider, P.P. 2010. Training the research integrity officers (RIO): The federally funded "RIO Boot Camps" backward design to train for the future. *Journal of Research Administration.* 41(3):99–117.