

Research Integrity Officers’ Responsibilities and Perspectives on Data Management Plan Compliance and Evaluation

Bradley Wade Bishop, PhD

Associate Professor

School of Information Sciences, University of Tennessee

Robert Nobles, DrPH, MPH, CIP

Vice President for Research Administration, Emory University

Hannah Collier

Graduate Research Assistant

School of Information Sciences, University of Tennessee

Abstract

This paper presents findings from interviews with US Research Integrity Officers (RIOs) on their overall responsibilities as well as perspectives on Data Management Plans (DMPs). DMPs are formal documents describing the roles and activities for managing data during and after research. DMPs are now a required research criterion by many funding agencies globally. A purposive sample of Research Integrity Officers (RIOs) from the top ten US private and public universities were recruited for interviews using an open-ended questionnaire related to their job duties and perspectives on data management plan implementation and evaluation. Responses from 12 participants were transcribed, anonymized, and coded in NVivo. RIO backgrounds, duties, and perspectives varied. The mode number of staff/faculty people dedicated to the RIO role at these institutions was a halftime appointment. All RIOs had some responsibilities related to Authorship, Publication, and Inventorship and Integrity and Information with 11 participants also responsible for offering some Responsible Conduct of Research (RCR) training. Most RIOs assumed that Principle Investigators are responsible for DMP compliance during sponsored projects as well as the long-term data management after a project ends. None of the twelve participants has received any Research Data Management training. Given the sea change in research practices, RIOs should have more training as data-intensive research emerges and DMPs

become commonplace.

Keywords:

Research Integrity Officer, Data Management Plan, Responsible Conduct of Research, Research Data Management.

Introduction

The purpose of this paper is to understand US Research Integrity Officers (RIOs) overall job responsibilities and perspectives on data management plans (DMPs). This study addresses literature gaps for both RIOs' responsibilities in general and their perspectives on DMPs. In 2011, the National Science Foundation (NSF) began requiring DMPs, and by 2016 all federal funding agencies began requiring similar documentation for any data generated from federally funded research activities (Holdren, 2013). DMPs are formal documents describing the roles and activities for managing data during and after research. Several US science funding agencies require researchers to submit a two-page document concerning data curation with data, including a variety of digital objects to enable reproducibility (e.g., notes, code, software, and so forth).

Given the relative newness of DMPs, there are research holes that this study fills related to the implementation and evaluation of DMPs from the research administration perspective.

A RIO is a position at US research institutions that fosters a Responsible Conduct of Research (RCR) environment, as well as someone who responds and marshals research misconduct allegations. The experiential learning of on-the-job training with many administrative academic positions is invaluable and unavoidable, but for RIOs some specialized training does exist through the Office of Research Integrity's (ORI) RIO Boot Camp (n.d.). The new paradigm of data-intensive science and DMP requirements adds another range of topics RIOs will encounter. In order to inform future RIO education and training, this study pursued interviews with leaders who were currently in RIO positions. The study also makes a valuable contribution to inform administrators of their typical roles and contribute to knowledge about this understudied servant for research integrity.

Literature Review

A recent perceived lack of confidence in science prompted Congress to mandate

the NSF to explore issues related to reproducibility and replicability as well as how these issues impacted the public's trust in science (National Academies, 2019). The report highlights some underlying matters that may contribute to a lack of confidence in science, such as a clear misunderstanding of the concepts of consensus and uncertainty. For example, in one survey a large swath of Americans thought "scientists are divided" on the human causes of climate change (37%) and on evolution (29%) (National Academies, 2016). One study showed that when scholars provide uncertainty information this actually leads to readers' distrust and confusion of science since it lacks absolute certainty (Frewer et al., 2003).

Although the perceived lack of confidence stems from these misunderstandings of science, there are other underlying issues in the National Academies report that could impact trust, such as research misconduct. Research misconduct is the leading cause of retracted publications (Campos-Varela, & Ruano-Raviña, 2019). Other research found that research misconduct is being significantly underreported, but unlike the broader distrust issues of science, research misconduct is addressed within the research enterprise (Titus et al., 2008). A myriad of Federal Research Misconduct Policies exist to ensure mechanisms are created for investigations of most federally funded research (<https://ori.hhs.gov/federal-policies>). These current policies derive from 1989 regulations that every research institution receiving U.S. Public Health Service funding must assure to ORI that their institution has policies and procedures to investigate allegations of Misconduct in Research and this has expanded as a requirement for funding from many federal agencies (PHS: 42 CFR 50, 1989; HHS: 42 CFR 50 & 93, 2005).

In 2000, the U.S. Office of Science and Technology Policy adopted a definition of research misconduct to include these three behaviours: (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 (FFP), (Mayer & Steneck, 2011). These behaviours clearly diverge from any concept of research integrity, but norms for responsible conduct vary from field to field and defining good citizenship for even these seemingly clear areas of research misconduct can be difficult (Steneck, 2007). To find some commonality across domains, cultures, and countries at the Second World Conference on Research Integrity, the Singapore Statement was created to provide ethical guidance which research organizations, governments, and scientists can use to develop policies, regulations, and codes of conduct to scope research integrity (Resnik & Shamoo, 2011). The four principles and fourteen responsibilities could be summed up in one word—honesty. Research integrity may serve as a set of honest practices to inoculate scientists from research misconduct; however, central to today's research are data and techniques (i.e.,

machine-learning) that are unable to self-assess their trustworthiness. Whether future research misconduct results from the actions of machines or humans, those investigating allegations work under the common job title—RIO.

Research Integrity Officers

Research Integrity Officers (RIOs) handle research misconduct allegations and promote ethical practices at research institutions (i.e., RCR). This role evolved as a result of the federal requirements to provide a system to investigate misconduct allegations. Interestingly, the RIO position was not mentioned by name in these federal regulations and what the job entailed emerged out of a necessity to address these guidelines (Wright & Schneider, 2010). In response to this knowledge gap, ORI-sponsored RIO Boot Camps in 2016 and held them annually until 2019 to bring RIOs and their legal counsel together for a best practices exchange. A RIO typically responds and performs an assessment on research misconduct that is classified as plagiarism, fabrication, and falsification. Data are central to at least two of the possible misconduct behaviours (i.e., data fabrication and data falsification). In addition, with more publishers requiring data deposit with manuscripts, even plagiarism may involve some exploration of data. Those individuals in the role of RIO may serve in other research capacities at their institutions, but their primary purpose is to ensure compliance with regulations by administering research misconduct allegations and cases.

Research Data Management

Big data presents large-scale challenges as researchers try to navigate massive quantities of data, work across disciplinary boundaries, and keep pace with the requirements of DMPs and preservation needs (Jaguszewski & Williams, 2013). Fields of science focused solely on computation have emerged or expanded, but the training of scientists in Research Data Management (RDM) best practices lags, which may lead to unintentional research misconduct. To help address these problems, the 2019 National Academies report recommends that NSF and other funders create code and data repositories that allow for the long-term preservation of digital artifacts. This is welcome news to the field of data curation, which for over a decade has worked to build this exact infrastructure anticipated by DMP requirements and research trends. In response to DMP requirements, academic institutions, libraries, publishers, and scientific and professional associations from all disciplines have made strides to make data more findable, accessible, interoperable, and re-usable (Wilkinson et al., 2016).

A DMP is a structured, formal document describing the roles, responsibilities, and activities for managing data during and after research (Bishop & Hank 2020). With

the push for more public-facing scientific research and accountability, many funding agencies (86% of UK Research Councils and 63% of U.S. funding bodies) require DMPs within the initial funding application (Smale et al., 2018). Through Horizon 2020, European Union-funded research must make all data accessible to anyone, free of charge, in addition to ensuring Open Access to all peer-reviewed scientific publications relating to its results (Koumoulios et al., 2019). Several academic journals now also require researchers to make public the data and digital outputs associated with a publication (The Royal Society, 2017; PLOS, n.d.).

Despite these external pressures to create and follow DMPs, the compliance with these requirements has lagged. For example, one study evaluated 119 DMPs and found that 51% did not identify the individual(s) responsible for data management, which is consistent with prior research findings (Van Loon et al., 2017). Retraction Watch (2019) reported that 32.5% of the 1,082 retracted publications in one year were the result of data problems (<https://retractionwatch.com/>). One study found that DMP audits resulted in an overall positive impact for researchers through improved data management (Ali, 2019). This lack of adequate DMP implementation or evaluation throughout the research lifecycle may lead to a lack of compliance down the road undermining the intention of DMP efforts. When NSF considers funding tools, training, and activities related to Research Data Management and journal editors consider ways to ensure reproducibility for publications, RIOs need to anticipate the changes to researchers' workflows and gain awareness and training to understand both the responsible conduct and potential research misconduct stemming from DMPs.

Method

This study used a semi-structured interview questionnaire, informed by RCR topics for responsibility questions, and used a modified Data Curation Profile (DCP) protocol for the DMP-related questions. This study received Institutional Review Board (IRB) approval prior to data collection (UTK IRB-20-05623-XP). Informed consent forms included open data language: "This means once responses are anonymized, the data will be openly shared, but only after all possible steps are taken to increase anonymity." The transcripts are available through the University of Tennessee's open repository, the Tennessee Research and Creative Exchange (TRACE). After IRB approval, a purposive sample of Research Integrity Officers (RIOs) were recruited by contacting the RIOs from the top ten National Universities (all private schools) and the top ten Public Schools as listed in the 2020 U.S. News and World Report Rankings (<https://www.usnews.com/best-colleges/rankings/national-universities>). Of the total 20 RIOs contacted, only three RIOs from top universities (private) and nine RIOs

from top ten public schools were interviewed via Zoom and in person (February through March 2020). The National Universities Rankings include those institutions that emphasize faculty research and since they have larger research expenditures, they are more likely to also have more researchers with the required DMPs and RIOs. In fact, ten of the twelve institutions had greater than 640 million total R&D expenditures in the most recent data aggregated (National Center for Science and Engineering Statistics, 2017). This sampling frame of top national universities was used in a parallel study of data librarians at these institutions to conduct a gap analysis on DMPs' implementation and evaluation.

The interviews consisted of 24 open-ended questions related to RIO duties and perspectives on DMPs implementation and evaluation. The job responsibility questions were informed by U.S. RCR topics (Steneck, 2007). The job tasks of a RIO at research institutions (universities, hospitals, private research companies, and so on) are required by law to have policies that cover various aspects of their research programs if they accept federal funds. The DCP questions were created to capture the step-by-step data lifecycle from scientists for digital curation, but the same approach works for any participant's understanding of data during and after research (Witt et al., 2009). This questionnaire borrows the order of questions on data, storage, costs, and training, to determine what, if any, knowledge RIOs have about the current status of Research Data Management at their institutions. The interview schedule consisted of the following questions:

Responsibilities and Overview

1. Which of the following list relate to your responsibilities?

- Authorship, Publication, and Inventorship
- Integrity and Information
- Conflicts of Interest
- Regulatory Basics for Human and Animal Subjects
- Human Subjects Research and Data
- Use of Human Biological Materials
- Societal Responsibility
- Other:

2. Have you ever used a data management plan in your research misconduct assessment, inquiry, and/or investigative processes?

3. How many people work in your research integrity office?

4. What is your scope of coverage (i.e., certain parts of the university)?

Data Management Plans

5. Do you have any oversight of data management plans?

6. Who is responsible for data management plan compliance?

7. How are data management plans evaluated for compliance?

8. If you were creating an office of integrity, what would be the ideal oversight structure and process for data management plans?

Storage

9. Does your institution have any ownership or disposition of data policies?

10. Does your institution support any institutional repositories for data?

11. Who is primarily responsible for the long-term management of the data for sponsored projects?

12. Who is primarily responsible for the long-term management of the data from research misconduct assessment, inquiry, and/or investigative processes?

Costs

13. How are data management efforts for sponsored projects at your institution funded?

14. What budget allocated exists for long-term data management beyond the life of projects and grants?

15. What budget allocated exists for long-term data management of the data from assessments, inquiries, and/or investigative processes?

Training

16. Does your office provide RCR training?

17. Does your office provide data management training?

18. Have you received any Research Data Management training?

- If yes, what types of data research management training did you receive?

Background

19. What is your current job title?
20. How many years in total have you been working in your current job?
21. How many years in total have you been working with research data (including relevant higher education)?
22. Please indicate your credentials and degrees.
23. Please provide any other educational or training you have received that is applicable to performing your job.
24. Do you have any other feedback about this project?

Interviews were transcribed, anonymized, and indirect identifiers were removed prior to analyses. Grounded theory application of open, axial, and selective coding in NVivo captured their job tasks and perspectives on Research Data Management. For nearly all the questions the responses were dichotomous (e.g., yes/no) and followed with few examples to explain why yes or no. Categories and broad themes were grouped for responses that had synonymous intertwined meanings into the same code (e.g., "I am charged with the research integrity program for all current and former persons of the (...) affiliation" (P2) and "entire university" (P5) were both coded 'Coverage-entire_university'). Given the lack of variance in responses (or potential responses), only a single coder was used and no reliability statistics were calculated. Yes and no responses indicated awareness or responsibility for several questions without a biased way of interpreting them.

The limitations of this study include its sampling, the interview questions used, and coding bias. Although not a representative sample, the participants were all from highly ranked universities with large research expenditures. A different sample with other RIOs from other institutions could have provided different responses. The RIOs participating in this study were either research office staff or very senior faculty, which could vary across institutions depending on how research administration is organized and resourced. Still, regardless of background and education there were clear trends in the responses from all RIOs. DMPs and more broadly the concepts of Research Data Management might fall outside of historic RIO training, and each individuals' research background, if these occurred prior to the big data-paradigm in sciences and related requirements.

The interview questionnaire was piloted with two RIOs and revised for clarity of the questions. The interview questions worked off an assumption that all RIOs had similar job tasks and some awareness of Research Data Management efforts on their campuses. Since this was an exploratory study, with no prior research in this specific

topic to inform the questions, the interview questions asked were answerable during pilot testing, but in practice some required more probing than anticipated for a clear response (e.g., “it kind of depends on what you mean by a data management plan” [P12]).

Finally, as a former RIO and current educator of Research Data Management, inherent biases in the interviews and coding occurred. One example is the assumption that participants understood each question related to DMPs. For example, if a participant asked for clarification on any term, such as an institutional repository, they were given an example. Yet, if a participant responded with a confident yes or no to any question, it was assumed they knew what the topic was and further probing did not occur. During coding, the transcripts are static with an inability to follow up with further questions. Future work may be informed by the following results to refine a questionnaire for a survey to produce more generalizable data.

Results

The results summarize all responses to the open-ended questions concerning RIOs’ responsibilities and institutional overview, perspectives and understanding on DMP compliance and evaluation, and RIO backgrounds. The qualitative data provide some insight into these RIOs with related discussion included in each section.

RIO Responsibilities and Institutional Overview

Table 1 presents the responses to job responsibility questions, which provides an overview of typical RIO work of these participants.

Table 1
RIO Responsibilities

RCR Training	Yes	No
Authorship, Publication, and Inventorship	8	4
Integrity and Information	12	0
Conflicts of Interest	5	7
Regulatory Basics for Human and Animal Subjects	8	4
Human Subject Research and Data	8	4
Use of Human Biological Materials	4	8
Societal Responsibility	6	6

Several RIOs provided other responsibilities, with three mentioning Exports Controls, two stating Radiation Safety, as well as one each for Controlled Substances, Animals, Biosafety, and lab practices. Three discussed training as a responsibility in this part of the interview. One RIO discussed a research rigor and reproducibility initiative that included training. Similarly, two held oversight roles for RCR training at the university for students and faculty.

The RIOs estimation of how many people worked in the research integrity office varied greatly. The mode for this answer is .5 FTE and was mentioned by five participants with an average of 2.83. One outlier mentioned ten people, but there is a chance they listed all individuals in the Office of Research that might support the research misconduct efforts and RCR. One RIO was responsible for misconduct reports for the entire university and affiliated hospitals, but the other 11 only were responsible for the misconduct reported at their university. Although not expressly asked, all RIOs mentioned reporting to a Vice President, Vice Provost, or Vice Chancellor of Research or if they served primarily in one of those roles, as five participants did, that their supervisors were Presidents, Provosts, and Chancellors of Research.

Data Management Plans

When asked if they had ever used a DMP in any research misconduct assessment, inquiry, and/or investigative processes, ten RIOs said no. In practice, none used a DMP with one saying they reviewed data and another saying that they would if necessary. Three RIOs responded to this question concerning their own digital curation practices. For example, organization is key to ensuring clean processes and “assuring chain of custody, version control, review status, metadata, flagging of individual documents” (P7) is an expectation for this work. Table 2 provides an overview of the responses to who bears responsibility for DMP compliance. All RIOs knew that DMP compliance was not their responsibility.

Table 2
Responsible for
DMP Compliance

Responsible for DMP compliance	#
Principle Investigator	8
Chief Compliance Officer	1
IRB	1
Sponsored Programs	1
Nobody	1

The next two questions asked how DMPs were evaluated for compliance and what the ideal oversight structure and process for DMPs should be. Eight RIOs did not know how DMPs were evaluated. One responded, “we are counting on the PI to certify them” (P5), but one each of the remaining participants ascribed this duty to the compliance officer, funder, or the library.

Many of the ideal structures for DMP oversight responses presented by RIOs showed a balance of working with faculty time constraints and the fiscal realities of each institution. Seven participants suggested additional DMP support including best practices, workshops, and tools, just as NSF suggests and scientific organizations and academic libraries have been offering for years. “I think it’s mostly about tools and making sure people know about those tools, and then having controls on those tools and mandating the use of those tools” (P2). There has been little marketing and outreach for existing tools, but one RIO was spot-on that without a mandate, researchers will not use certain tools. Three RIOs suggested new evaluation procedures such as “fully staffed group for quality assurance/quality improvement, where part of their annual audit plan is going out and testing some of the data management plans, and say, ‘You said you were going to do this.. show us!’” (P1). Conversely, two participants thought each department should handle DMP compliance because of disciplinary differences that align with current decentralized oversight structure for all research. Finally, one RIO suggested the academic library because they already serve a liaison-type role across units.

Storage

Table 3 provides responses to the storage section of the interview.

	Storage questions	Yes	No
Table 3 Data Policies and Institutional Repositories	Does your institution have any ownership or disposition of data policies?	12	0
	Does your institution support any institutional repositories for data?	11	1

The next storage question asked who is responsible for the long-term management of data from sponsored projects. Ten participants, like the responsibility of the initial DMPs, responded PIs are also responsible for long-term management. Five of those that indicated the PI mentioned others that could contribute to solving this problem. Four other RIOs mentioned the libraries as part of a solution, but as one pointed out “they have to carve it out of their existing slice” of their budgets (P7). The information

technology (IT) as potential helpers were mentioned by four participants. Three mentioned that departments might help as faculty move and retire. Two participants had the Vice Provost/President for Research or someone in sponsored projects managing this issue.

The final storage question asked who is responsible for the long-term management of the data from research misconduct assessment, inquiry, and/or investigative processes. All twelve RIOs stated that the RIO themselves were responsible for their own data from assessments, inquiries, investigations, with one stating they could consult with the university archivist if need be.

Costs

The funding questions related to how each institution (1) supports data management efforts for sponsored projects, (2) budgets for the long-term data management beyond the life of projects and grants, and (3) preserves data from assessments, inquiries, and/or investigative processes, presented a question out of scope for the RIOs. Although two participants said they did not know, ten RIOs assumed that sponsored projects or some other university-level entity supported data management for projects and grants. Seven RIOs did not know who funded long-term RDM efforts. Two stated that no one funds that, but one participant thought individual PIs would cover those costs and another presumed each department could finance data curation efforts.

RIOs all had a much better handle on responding to the question about their own data management practices and budget. Ten said that there was no separate line item for RIO storage. Two RIOs did indicate that data storage is sometimes needed, and funds are available when needed.

Training

Table 4 shows the different campus approaches to RCR training.

	RCR Training	#
Table 4 Responsible Conduct of Research Training Models	RIO-driven	3
	Research Office-driven, but not RIO	3
	Campus-wide RCR group	3
	RCR course for all graduate students	1
	General online RCR modules only	1
	All RCR education done at department level	1

Participants were also asked if they had given any data management training with six participants reporting no and two stating yes. Four other participants said that data management training was done on campus, but not by the RIO or via RCR. One participant each mentioned library services or computer science as somewhere researchers might go for that training.

All RIOs were asked if they had received any research data management training. Seven said not formally but learned as a part of their career as a researcher or at conferences. Five participants said they had not received any RDM training.

RIO Backgrounds

The job titles varied due to some RIOs who served in several roles with RIO as one of several nested job titles of participants. For example, five Associate Vice Provosts/Presidents/Chancellors of Research also served as the RIO when needed. Three participants also mentioned their faculty appointments as professors or chairs of departments as their other roles. In six instances, the Director of Research Integrity or Research Policy also served as the RIO and these were the participants that did not have other duties or faculty status.

The average number of years working in the role of RIO was almost 6 years. The range of experience was from one and a half to 16 years. Seven had five or fewer years with a few outliers having 8, 10, and 16 years in that role. The average number of years RIOs had been working with research data was 26.6, which is much higher than time as a RIO because participants were asked to include all relevant higher education. The range of experience with research data spanned from seven to 50 years. Seven of the 12 RIOs were very experienced with over 24 years of experience albeit mostly with data from their own domains.

Six participants had a PhD as their highest level of education, with Biology (4); Civil and Environmental Engineering (1); and Biochemistry (1). These participants also held master's and bachelor's degrees in their areas with one having an additional public health master's degree. Two participants had JDs, with one JD also having many other health-related credentials—a Master's in Public Health, Certified in Healthcare Compliance (CHC), Certified in Healthcare Research Compliance (CHRC), and was a certified Clinical Research Associate. The other JDs had bachelor's degrees, which were in psychology and entomology. Two participants had MBAs and suggested the project management strengths helped run their investigations and rely solely on faculty for domain expertise (P1). One RIO had a master's in genetic counselling with an undergraduate degree in microbiology and molecular genetics. Finally, one RIO

had a bachelor's in science in biology as their highest degree with a Clinical Research Coordinator certificate.

In response to the question concerning other education or training that was applicable to performing their jobs, 11 RIOs mentioned the ORI-sponsored training RIO Boot Camps. Nine participants mentioned other education (e.g., conflict management), other RCR trainings and conferences (e.g., National Council of University Research Administrators), and experience as a faculty member resolving issues, all of which helped them perform these jobs. One participant had a unique background as a lawyer practicing criminal defense, which they state gave them “transferable skills... strong analytical skills, strong communication skills, being able to develop strategies to interact with people, particularly in this context with faculty to develop strategies to keep them in compliance, let's put it that way” (P8). There were no further mentions of other useful education or training and no participant had additional feedback on the study.

Discussion

This discussion provides some context with past research on RIOs and a few suggestions for future work based on the common responses. The RIO perspectives and understanding of DMPs may have implications for the future of research data management and trust in science given their integrity role.

RIO Responsibilities and Institutional Overview

The responses to job responsibility questions provide an overview of typical RIO work and all participants indicated that Authorship, Publication, and Inventorship and Integrity and Information were a responsibility. Those first two responsibilities directly relate to the behaviours defined as research misconduct, FFP, and it may be assumed a part of any RIO's job.

The same is not true for other responsibilities as local contexts determine how resources and responsibilities are assigned. The job of RIO varies most in these potential responsibilities—Conflicts of Interest, Regulatory Basics for Human and Animal Subjects, Human Subjects Research and Data, Use of Human Biological Materials, Societal Responsibility. For example, seven RIOs said that Conflicts of Interest was not a part of their role because others in their office of research handled that specifically, but five did consider that as part of their job. Some responsibilities did fall under the purview of eight RIOs (i.e., Regulatory Basics for Human and

Animal Subjects and Human Subject Research and Data) and four said no unless it is misconduct related that work falls to “other parts of the office that handle them” (P10). The inverse was true for the responsibilities related to Use of Human Biological Materials, with eight no’s as other offices handled those aspects of research and four yes’s. The RIO is central to RCR and research misconduct on their campus, so it is unclear why all of these topics were not unanimous. Perhaps, some RIOs have not had enough experience for these topics to come up in their work, or as later questions reveal, some RIOs focus solely on managing allegations of research misconduct.

For Societal Responsibility, the responses were split with a good deal of misunderstanding about what the associated job tasks might be for that. Indeed, this aspect of RCR is difficult to operationalize into daily or weekly tasks, especially for those only on a half-time appointment. The other responsibilities provided by RIOs reflect their institutions’ research areas—exports controls, radiation safety, controlled substances, animals, biosafety, and lab practices. If not a role for the RIO, with new requirements and data-intensive practice for most research the area of Research Data Management compliance and evaluation should be considered by all Offices of Research.

Although the RIOs responses to the number of people working in their office varied, five participants indicated .5 FTE. For even these highly-ranked universities with large research expenditures, a half-time RIO suffices to watch over a multitude of research projects across disciplines and researchers at all career levels. Still, the average was much higher at 2.83; that may indicate some Offices of Research more fully support all RCR and research misconduct efforts. The question was difficult to answer for some, but even though it is difficult to scope a RIO’s work, the variety of responses indicate the human resource investment into these tasks is not uniform. A few RIOs did state “we don’t have any problem accessing extra support from our IT folks” (P8) and “we assemble a faculty committee that would work under the supervision of their RIO and the dean to carry out their inquiry investigation . . . and desire the faculty committee to have content expertise” (P10). Clearly, when more resources are needed RIOs indicated they are provided. In all cases, each RIO’s coverage was the entire university with one adding the affiliated hospitals. The breadth of potential research misconduct that is never alleged or the volume of unfounded allegations deemed not research misconduct are understudied. Without reliable metrics on these aspects of the research misconduct, it is not possible to project what would be adequate resources for RIOs and RCR activities. These figures through other studies are needed to inform adequate staffing and increase the research integrity of campuses. It would be ridiculous to have .5 FTE to manage and respond to allegations of other types of misconduct that occur on campuses, and with such large research expenditures

these institutions should invest in the prevention and oversight necessary to protect the integrity of these substantial investments.

Data Management Plans

In an era of big data, and nearly a decade since DMPs were required by NSF, the absence of DMPs in any research misconduct assessment, inquiry, and/or investigative processes is telling. This is likely due to researchers not updating DMPs once funded. One RIO said that they would use one if it was related to research misconduct. Another RIO stated they used data, which may or may not have derived from a DMP specifically, but said we “review data as a result of findings of misconduct or findings of questionable research practices or other things like that” (P4). RIOs should know to ask for DMPs as they could be used as a roadmap for the data generated and indicate points of contact and steps in processes where misconduct or falsification could occur. A DMP describes the roles and activities for managing data during and after research that would help any inquiry or investigation. Also, there appeared to be some confusion over the terminology: “it kind of depends on what you mean by a data management plan” (P12). This may reflect those faculty or staff assigned this administrative role without actual awareness of this relatively recent research requirement. With additional study of data curation behaviours across disciplines, RIOs could know what information organization practices to expect in different fields and when to spot risky data curation approaches. RIOs were very confident in their own digital curation practices, which is paramount to any investigative position. Personal information management and data workflows for RIOs could be standardized across the profession. Data standards (e.g, naming conventions, controlled vocabulary, and so forth) would help in aggregating data for reporting purposes and assist during onboarding of new RIOs.

The responses to who was responsible for DMP oversight varied, with the majority indicating the Principle Investigator (PI) would be responsible with presumed university support. As participant 1 put it, “that’s kind of a void right now, and that’s one where I would say, ultimately, the researchers. But we also always tend to add, we as the universities tend to add a lot on the researchers, so I think the real answer is yes, that it’s their responsibility, but it’s our responsibility to help them do that or find means/ways to do those things”. One participant summed up the need to dodge this potential area of non-compliance and misconduct succinctly: “you know what it is, it’s an unfunded mandate, and nobody has time” (P5). This forthright statement should resonate with anyone that has had to write or implement a DMP, but ignoring the data piece of the research lifecycle prevents reuse and reduces reproducibility. From

the seat of a RIO, poor DMPs or non-implemented DMPs complicate investigations related to data fabrication and falsification.

It is understandable that most RIOs did not know how DMPs were evaluated as this work is far from the RCR arena. Funding agencies, proposal reviewers, and researchers themselves see a DMP briefly and once funded, there is little incentive to revisit or reassess the document. One participant responded, “we are counting on the PI to certify them” (P5), but one each of three RIOs ascribed this duty to the compliance officer, funder, or the library. For now, DMP compliance and evaluation is up to PIs without any oversight from the funding agencies or locally at institutions. Academic libraries are poised to assist, with many hiring multiple data librarians since DMP requirements became many funders’ expectations. We do not expect RIOs to ever have a role in these processes, but this study indicates through a small sample that DMPs are not currently on the radar for RIOs even as they may relate to RCR instruction, if not inquiries/investigations.

Despite this tertiary role for RIOs and DMP oversight, the participants did have imaginative solutions for this piece of research administration. As NSF suggests more DMP support including best practices, workshops, and tools match the calls from many scientific organizations and academic libraries. One RIO was very detailed in a plan for DMP assessment saying that they would “pull out a sample of about 33%, depending on the numbers, and spread those across departments to see what we find, and we would have a monitoring tool that we would go out and we would monitor to see... then depending on that initial sample base would dictate the types of education and future monitoring that we would deem required” (P11). With more centralized control of data or these types of audits, the falsification and fabrication misconduct investigations would be streamlined. 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 PIs” (P10).

Ultimately, an ideal structure differs for each institution even among these similar research universities. The preventative efforts of RDM and RCR training will also benefit from a research data infrastructure built to deter research misconduct (i.e., built-in safety measures and warnings for misuse of data). Perhaps, RIOs have some educational role if not in actual oversight.

Storage

All twelve RIOs said that their institution has intellectual property policies where each university owns the data produced there. A review of those policies was not

conducted and beyond the scope of this study, but data disposition is another avenue to explore related to Research Data Management. To assess awareness of where the data are stored, participants were asked if their institutions support any institutional repositories (IRs). Eleven said yes, but RIOs varied on their familiarity with them. One participant said they had an IR, but it was not free. Only one said no, but there is a chance they were not aware of IRs as most institutions in the U.S. have them. For example, participant 8 said “there are policies and procedures related to which data go where and get backed up in, in those repositories” and these types of responses might indicate more training is needed of the data lifecycle of present data-intensive sciences on basic data curation terminology. One RIO suggested that funders provide a repository finder as many data repositories already exist by discipline and researchers would not need to use the university IR. In fact, similar tools do exist in some disciplines. Currently, the American Geophysical Union’s (AGU) Repository Finder has a searchable database of 222 repositories (<https://repositoryfinder.datacite.org/>).

The final storage questions related to who is responsible for the long-term management of data from sponsored projects and research misconduct assessment, inquiry, and/or investigative processes. The majority assumed that PIs would be responsible for not only the DMP during a project, but the only clear choice for long-term data management. Half of the ten that mentioned the PI as the responsible party suggested others at each institution that may help. Academic libraries and the data librarians that work in them are positioned to take on these roles but might not be connected to the research enterprise. It might be possible for RIOs as part of RCR training roles to actively involve librarians to appropriate their expertise. Others in IT roles could also be brought in to augment training on campuses. One way to address any unfunded mandate is to have centralized bodies, like academic libraries and IT offices, within a university absorb the new costs. This may impact the quality of data sharing and call for a reallocation of overhead to supplement data curation costs. As faculty move on and/or retire having departments or persons in offices of research hold data might be an undue burden and not necessarily the proper infrastructure for Research Data Management. At institutions where the researchers do not retain ownership of data, it is odd that the university does not seem to know where their data are located or might be held in-perpetuity (or lost). These broader research administration concerns are beyond the scope of most RIOs, but in research misconduct investigations it may be useful to have some prospects in how data sharing and data management occurs or may occur on their campuses.

On the contrary, all twelve RIOs stated that the RIO themselves were responsible for their own data from assessments, inquiries, investigations, with one stating

they could consult with the university archivist if need be. With clear regulations for records management related to research misconduct, RIOs know exactly how long storage is expected (i.e., 7 years). Similar regulations are needed for each discipline and every institution to inform the preservation of research data.

Costs

The costs questions were beyond the concerns of RIOs and all costs incurred for storage and other curation efforts do not relate to current RCR topics. Ten of the RIOs assumed that sponsored projects or another university-level entity would assist in the long-term data management beyond the life of sponsored projects and grants. The concern of some was apparent that costs would exceed the budget of each project and some university funds would end up supporting Research Data Management efforts with comments like “my understanding is that grants rarely cover all of it” (P3). Ultimately, all data curation beyond the life of project and grants forces data preservation costs onto other entities. Retiring faculty may be given the option to leave their research data, but also asked to cover the curation costs either paid by the individual, department, or funder as long-term management requires cleaning to make data interoperable and enhancing data for discoverability and reuse. As digital objects become the norm for research practices, costs considerations should increase to avoid a total loss of the huge investments in careers of data collection.

Ten RIOs said that there was no separate line item for RIO storage. Statements such as “I mean other than my own effort and cabinet here” (P8) indicate some gap in digital preservation approaches might impact future access. Still, others point out that “once the inquiry or investigation is done, we’re not looking at it anymore” (P10). These responses match those of data storage in that RIOs know their own data and associated costs.

Training

In response to the specific RCR training question, three RIOs explained that their office did offer training that was RIO-driven with one participant stating, “I teach three classes, and I mean, entire classes, not lectures” (P5). Eight others said their office did give RCR training and RIOs were involved, but not as lead organizers. The RCR training described follows with the number of RIOs that described each framework: RCR training is done by someone from the research office by visiting departments, but not the RIO (3); a campus-wide RCR group that offers more discipline-specific training upon request, but not coordinated through the research office (3); a required

RCR course for graduate students (1); or general online RCR modules not created by the institution (1). The required course for students is one way to ensure all have some consistent exposure to RCR from people outside their department, but each institution has their own approach. Only one participant said they were not involved in any RCR training at all with all RCR-education decentralized and nothing across campus. For the most part, RCR training is preventative of unintentional research misconduct. On most campuses, it appears that RIOs take the lead or contribute to other RCR efforts on campuses to promote research integrity.

In most instances, RIOs do not give any data management training. Two RIOs did say yes, but it was “out of 11 or 12 sessions one covers data management” (P12). Similar to other data storage and costs, RIOs are aware of research data management offered across their institutions usually at the academic library or other IT units. Research Data Management is more central to many data-intensive sciences now, so perhaps greater experience with these areas could lead to more focus on RCR training in these areas.

It might be problematic that most RIOs had only received informal Research Data Management training, from their own research careers or at conferences, given that methods and data change over time. As one participant expressed concern over incidental misconduct in this way “things have changed immensely, and I would say, I mean I think in the research integrity, or misconduct world, there is kind of the need of the PI who entered the field ten years ago, say, before the big data explosion, and it is now running a lab, full-borne in the big data explosion, without a solid statistical training, without solid scripting ... That’s a good way to get in trouble” (P10). A lack of familiarity with this new paradigm might also present challenges for investigations into allegations of research misconduct. As the data lifecycle relates to some aspects of potential fabrication and falsification, RIOs should have additional training on these aspects of the research enterprise if not to train others at least for their own responsiveness to new research practices.

RIO Backgrounds

RIOs have various backgrounds that reflect some pipelines into these administrative roles. Research administrators seemingly collect job titles (i.e., wear many hats) and the role of the RIO is often one of multiple jobs for most of these individuals. In this study, nine participants mentioned faculty appointments whether as professors in departments or that they retained faculty status in their administrative roles. This matches a prior survey that found 42% of 56 RIOs were tenured faculty (Wright & Schneider, 2010). Clearly, some institutions value the faculty status of a RIO and

others do not, but there are pros and cons to either model. A pro for faculty status may be peer respect throughout the research misconduct procedures. A con for faculty status may be duration of proceedings due to faculties' limited availability. A dedicated staff person handling operations is a model for many to facilitate a more streamlined process.

These participants averaged six years, but seven participants had served five or fewer years. These results match closely with a prior survey of 56 RIOs that found an average length of service of five years (Wright & Schneider, 2010). RIO Boot Camps serve as continuing education for these newer RIOs as the institutional knowledge and experience gained from this type of work only can be acquired through personal practice or learned from senior mentors. Turnover in these positions underscores the importance of data curation standardization in the profession.

Half of the participants had a PhD and many were from the life sciences. This also matches a prior interview study of 79 RIOs that found 60% of RIOs self-identified as researchers with over half possessing a PhD (Bonito et al., 2012). Although this was a small qualitative study, these consistencies with prior research may indicate the sample reflects RIOs more widely. Prior work did not gather discipline-specific information, but future studies should. With many regulations related to the life and health sciences a great portion of all RIOs likely match these results and future hires would benefit from these educational backgrounds.

Again, eleven RIOs mentioned the ORI-sponsored training RIO Boot Camps as critical to their success. The Boot Camps allowed for a best practice exchange and as participant 8 put it, "fellow colleagues at other institutions are very, very valuable". Many former and current RIOs likely agree with participant 4's thought that "I'm not really sure there's training for this job". Despite educational efforts, RIOs are not uniformly trained for their positions, but what training does exist would benefit from more Research Data Management scenarios and at the very least data curation terminology and concepts.

Conclusion

The study provided some baseline results on RIOs' overall responsibilities and perspectives on DMPs. RIOs included staff and senior faculty from a variety of backgrounds, but consistencies emerged from their lack of RDM training and understanding of data management across their institutions. As DMPs become more routine following funding agency requirements, RIOs will encounter more research misconduct that relates to data and DMPs. Although not a representative sample,

these participants were all from highly ranked universities. Most RIOs participating in this study were either research office staff or very senior faculty, so DMPs and more broadly the concepts of Research Data Management would be unknown as most of them received training prior to the big data-paradigm in sciences and related requirements.

In this study, no one had used a DMP in any research misconduct activity. The DMP as a static document may not assist with some assessments, inquiries, and investigations, but knowing how data are created, stored, and made available during and beyond the life of a project certainly could be useful. The DCP questionnaire itself is a tool from Information Sciences to gather a data story and in the event that a DMP does not exist or is outdated, a DCP might be an additional useful instrument for a RIO. At each step in the data lifecycle, different processes and people create, interact, transform, and use data; a DCP highlights these steps and that alone could be relevant to a RIO's work. The misuse potential in data reuse presents other probable considerations for RCR trainings. Perhaps, a RIO's awareness of these potential tools and reuses of data are low due to a lack of experience. More broadly, if not a RIO, some research administration entity should conduct DMP oversight as proper data curation practices prevent misuse, including fabrication and falsification. With advancements in artificial intelligence and machine-learning, research misconduct propagation may occur, but these new tools (e.g., iThenticate) may also become invaluable tools to assist RIOs. Plagiarism software needs an equivalent for data.

Recruiting participants from other locations may lead to other findings, but qualitative research is a good first step to explore understudied areas. The interview responses could inform future survey work to produce more generalizable findings. Still, there was some saturation in responses from the participants and clear themes emerged about RIO backgrounds, training, and perspectives on DMPs. Prior to the needed creation of RDM training at future RIO Boot Camps, academic data librarians may serve as a resource to help all the "faculty out there who really could use some help setting up data management plans for their research" (P6) as well as the RIOs who may need to speak with them.

Authors' Note

Material in this paper is the result of data collection done for the Spring 2020 Faculty Development Leave of the first author. The first author served as one of two deputy RIOs at the University of Tennessee under the second author who was the primary RIO and Interim Vice Chancellor for Research during that time. Bishop's primary research interests are Research Data Management and Data Discovery behaviours

of scientists. Nobles' prior work focused on adolescent health, but now he serves as a national leader in RCR and a catalyst to enhancing research culture globally. We greatly appreciate the participants and upon publication will deposit the anonymized and deidentified transcripts in the Tennessee Research and Creative Exchange (TRACE), which serves as the University of Tennessee's institutional repository.

Wade Bishop, PhD

Associate Professor
School of Information Sciences
University of Tennessee
1345 Circle Park Dr. Room 454
Communications Bldg.
Knoxville, TN 37996
Tel: (865) 974-2775
Fax: (865) 974-7878
Email: wade.bishop@utk.edu

Robert Nobles, DrPH, MPH, CIP

Vice President for Research Administration
Emory University
1599 Clifton Road NE
Atlanta, GA 30322
Tel: (404) 727-3889
Email: robert.e.nobles@emory.edu

Hannah Collier

Graduate Research Assistant
School of Information Sciences
University of Tennessee
1345 Circle Park Dr. Suite 451
Communications Bldg.
Knoxville, TN 37996
Tel: (865) 974-2148
Fax: (865) 974-7878
Email: harmenda@vols.utk.edu

Correspondence concerning this article should be addressed to Wade Bishop, PhD, Associate Professor, School of Information Sciences, University of Tennessee, 1345 Circle Park Dr. Room 454, Communications Bldg., Knoxville, TN 37996, wade.bishop@utk.edu.

References

- Ali, Y. (2019). Effectiveness of data auditing as a tool to reinforce good Research Data Management (RDM) practice. 6th World Conference on Research Integrity. Hong Kong.
- Bishop, B. W., & Hank, C. (2020). Curation, digital. In Audrey Kobayashi (Ed.), *International Encyclopedia of Human Geography*, 2e. Elsevier. <https://doi.org/10.1016/B978-0-08-102295-5.10531-1>
- Bonito, A., Titus, S., & Wright, D. (2011). Assessing the preparedness of Research Integrity Officers (RIOs) to appropriately handle possible research misconduct cases. *Science and Engineering Ethics*, 18. <https://doi.org/10.1007/s11948-011-9274-2>
- Campos-Varela, I., & Ruano-Raviña, A. (2019). Misconduct as the main cause for retraction. A descriptive study of retracted publications and their authors. *Gaceta Sanitaria*, 33(4), 356-360. <https://doi.org/10.1016/j.gaceta.2018.01.009>
- Frewer, L., Hunt, S., Brennan, M., Kuznesof, S., Ness, M., & Ritson, C. (2003). The views of scientific experts on how the public conceptualize uncertainty. *Journal of Risk Research*, 6(1), 75-85.
- HHS: 42 C.F.R. Parts 50 and 93 2005. Public Health Service Policies on Research Misconduct; Final Rule.
- Holdren, J. P. (2013). Memorandum for the Heads of Access to the results of federally funded scientific research. (Executive Office of the President: Office of Science and Technology Policy, February 22, 2013).
- Jaguszewski, J., & Williams, K. (2013). New roles for new times: Transforming liaison roles in research libraries. <https://conservancy.umn.edu/handle/11299/169867>

Koumoulos, E. P., Sebastiani, M., Romanos, N., Kalogerini, M., & Charitidis, C. (2019). Data Management Plan template for H2020 projects (Version v01.100419). Zenodo. <http://doi.org/10.5281/zenodo.2635768>

Mayer, T., & Steneck, N. H. (Eds.). (2011). Promoting research integrity in a global environment. World Scientific.

National Academies of Sciences, Engineering, and Medicine. (2016). Science literacy: Concepts, contexts, and consequences. The National Academies Press.

National Academies of Sciences, Engineering, and Medicine. (2019). Reproducibility and replicability in science. The National Academies Press. <https://doi.org/10.17226/25303>

National Center for Science and Engineering Statistics. (2017). Rankings by total R&D expenditures. <https://ncesdata.nsf.gov/profiles/site?method=rankingBySource&ds=herd>

Office of Research Integrity (n.d.). RIO Bootcamp. <https://ori.hhs.gov/rio-boot-camp>
PHS: 42 C.F.R. 50 1989.

PLOS. (n.d.) Data availability. <http://journals.plos.org/plosone/s/data-availability>

Resnik, D. B., & Shamoo, A. E. (2011). The Singapore statement on research integrity. *Accountability in Research*, 18(2), 71-75. <https://doi.org/10.1080/08989621.2011.557296>

The Royal Society. (2017). Data sharing and mining. <https://royalsociety.org/journals/ethics-policies/data-sharing-mining/>

Titus, S. L., Wells, J. A., & Rhoades, L. J. (2008). Repairing research integrity. *Nature*, 453(7198), 980-982. <https://doi.org/10.1038/453980a>

Smale, N., Unsworth, K., Denyer, G., & Barr, D. (2018). The history, advocacy and efficacy of data management plans. *bioRxiv* (pre-print). <https://doi.org/10.1101/443499>

Steneck, N. (2007). ORI introduction to the responsible conduct of research (Rev. ed.). Dept. of Health and Human Services.

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. <https://doi.org/10.1177/0340035216682041>

