

Finding common ground and identifying opportunities: Case study in data policy at an academic medical school

Erin D. Foster, University of California - Berkeley
Heather L. Coates, Indiana University Purdue University - Indianapolis

Acknowledgements

We would like to recognize Ariel Deardorff and Jonathan Petters for their thoughtful comments and constructive feedback on a draft of this report.

Introduction

We present this paper to our colleagues as a case study of a well supported effort by a large medical school to actively engage researchers in data management by requiring the completion of data management plans. We will outline the *people and roles* involved in this initiative, the efforts to *scope* and adjust *the mandate* over the past three years, the *communication* and work to change research *culture*, the *infrastructure* used to create these plans, and the *training and support* provided throughout to the research community. We hope this work offers insight and proves useful in advocating for and/or engaging with research data management work at your respective institutions.

In our first discussions about writing up this institutional effort to require data management plans, we debated how to frame the work and speak to the varied audiences who might be interested. For example, how could we synthesize the outcomes of this initiative to make a compelling case to academic health sciences leadership that involving libraries (and funding positions) can strengthen an institution's research enterprise? How might this work encourage library leadership to engage in these opportunities when possible, even if they take time away from other duties? Are researchers interested in hearing about how this type of mandate may assist in preparing them for future requirements? Finally, how could we share our knowledge with fellow librarians and data professional colleagues about our institution's approach to implementing a research data management plan mandate?

There were many potential audiences, and many potential lenses to this story, but we ultimately decided that our library and data professional colleagues might be the most interested in the arc of this initiative, particularly how this requirement evolved as the Draft and Final NIH Policy for Data Management and Sharing was released and implemented. We also choose to write this as a white paper - as opposed to a peer reviewed article - to allow us to present this work directly with as much detail as needed.

Background

The Indiana University School of Medicine (IUSM) is the largest medical school in the United States and supports a large research enterprise across nine campuses throughout the state of Indiana.¹ The Ruth Lilly Medical Library (RLML) serves all campuses of IUSM and is housed on the Indianapolis campus. The Library is situated within the Division of Education at IUSM and librarians are actively involved in the School's undergraduate curriculum.² The Library expanded support of the IUSM research community with the hiring of a Data Services Librarian in 2016 and the formation of the RLML Research Team a couple years later.

The scope of IUSM research has significantly grown over the past 5-7 years, particularly in its funding through the National Institutes of Health (NIH).³ With this growth, there was a desire from the IUSM Office of Research to establish stronger information management practices. These efforts were prompted, in part, by internal challenges the School experienced around the process of retaining research data and responding to questions regarding reproducibility of IUSM-produced research when they arose. Additionally, the School was dealing with growth-related challenges in the physical spaces

¹ <https://medicine.iu.edu/campuses>

² <https://ebm.bmj.com/content/26/5/249>

³ <https://medicine.iu.edu/news/2021/01/nih-funding-milestone>

allocated to research labs and management of various requirements for materials (e.g., instruments, equipment, paper lab notebooks) crucial for research endeavors.

In December 2017, the RLML was approached by external consultants. As part of the broader information management initiative, these consultants were brought onboard by the Associate Dean for Research Affairs to review and optimize the usage of physical lab spaces at the IU School of Medicine. The consultants reached out to RLML after seeing information about research data management support on the medical library's website as well as the research data services provided by the IUPUI University Library.⁴ Due to the presence of the Data Services Librarian and the collaborative relationship with the University Library's Digital Scholarship and Data Management Librarian, the medical library was well positioned to assist the IUSM Office of Research Affairs with this initiative.

A common space challenge faced by researchers is the physical storage of physical lab notebooks. Since electronic lab notebooks (ELNs) offered the potential of addressing both the physical space limitations and enabling better information management practices, it was determined that the school should pilot an ELN platform with library-based support for IUSM researchers.⁵ However, providing such a tool to researchers only partially addressed the goals related to information/data management landscape at the School. The goal for the IUSM Office of Research Affairs was, ultimately, to have a more complete picture of the research studies being conducted at the School. With that lens in mind, data management plans (DMP) were identified as a mechanism to gather information about the state of research data management and sharing practices across the school as well as a way to foster greater research integrity and encourage a culture of strong information/data management practice at the School. Presentations at the Research Data Access and Preservation (RDAP) Association⁶ and Medical Library Association⁷ conferences provide further details about this phase.

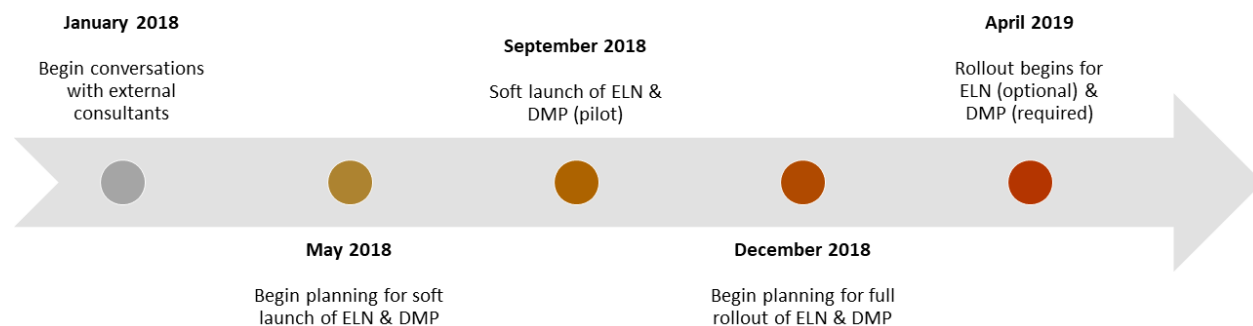


Figure 1: Timeline for the pilot phase

⁴ <https://cds.ulib.iupui.edu/digitalscholarship/datasupport>

⁵ <https://scholarworks.iupui.edu/handle/1805/23220>

⁶ <https://hdl.handle.net/1805/28938>

⁷ <https://hdl.handle.net/1805/23220>

DMP Rollout at IUSM (Year 1: 2019-2020)

People & Roles

The people and roles involved in this work are an important indicator of the breadth and “level” of support that was crucial to the implementation and continuation of this initiative.

The Executive Associate Dean for Research Affairs, who oversees all research related activities at IUSM, sponsored the project. They delegated direct oversight to another executive position (i.e., Associate Dean for Research Affairs) who hired the external consultants to establish an information management strategy for the school. Once the Library was involved, an advisory board was established to provide guidance on this work and champion the initiative. The Information Management Advisory Board (IMAB) included several prominent PIs across multiple departments, representatives from the medical library (Library Director, Data Services Librarian) and the University Library (Digital Scholarship & Data Management Librarian), the IUSM Chief Information Officer, the HIPAA Security Officer, and the external consultants (2). The representative from the University Library also serves as the Research Data Steward for Indiana University, which is a role in the University’s data governance program.⁸

During this time period, IUSM also rolled out an electronic lab notebook (ELN). There was significant overlap between the rollout teams; with the IUSM medical library leading implementation, training, and outreach efforts for both efforts. Additionally, the medical library worked closely with an IT professional at IUSM supporting several research labs that participated as “early adopters” for the ELN system. In both cases, the Data Services Librarian led the project work for the medical library, and worked closely with other librarians to develop and deliver training for both the ELN and DMP initiatives.

Scoping the Mandate

In 2019-2020, the IUSM Office of Research Affairs set the expectation that all IUSM researchers receiving external funding should complete a data management plan annually. A data management plan describes how data is gathered, managed, analyzed, reported, archived, and disposed. DMPs are currently required (to varying degrees) by certain funding agencies⁹ and, as of 2023, National Institutes of Health (NIH) funded researchers are required to create DMPs as part of grant applications.¹⁰ Several issues presented themselves initially in scoping this requirement: 1) clarifying “external” funding, 2) accurately identifying principal investigators (PIs) on campus to whom this mandate applied, and 3) specifying the parameters of a DMP. The aforementioned Information Management Advisory Board (IMAB) was the primary body that grappled with these questions and guided the scoping and implementation of this mandate.

While the initial scope was defined as any IUSM researcher receiving more than \$50,000 per year in external funding, it quickly became clear that there was some confusion and uncertainty in determining who met this threshold. IUSM researchers frequently have affiliations with the nearby children’s hospital (i.e., Riley Children’s Hospital), Richard L. Roudebush Veterans’ Administration Medical Center, as well as the Regenstrief Institute, which is a research institute that specializes in informatics and health

⁸ <https://datamanagement.iu.edu/governance/data-stewards/index.html>

⁹ SPARC Data Sharing Requirements by Federal Agency: <http://researchsharing.sparcopen.org/data>

¹⁰ Final NIH Policy for Data Management and Sharing: <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html>

services research. Funding provided by these organizations - in addition to internal IUSM “start up” funds for new PIs - were excluded from the requirement; “external” then came to mean any organization (federal, academic, private, etc.) that was not one of these three entities/sources. However, since few researchers were solely funded by one of these “internal” sources, nearly all were required to comply with the requirement. In the interest of communicating this requirement broadly across departments and research groups, PIs were then told and/or led to understand that they should complete a DMP regardless of funding source. In particular, IUSM leadership and the Advisory Board felt this initiative was an important pathway to preparing researchers for the expected policy changes from the NIH.¹¹

Quantifying the total number of principal investigators in the school was an unexpectedly challenging aspect of scoping this mandate. The IUSM Office of Research Affairs provided the IUSM Data Services Librarian and the IUPUI Digital Scholarship & Data Management Librarian with a list of active awards from the most recent year (i.e., FY 2018), which gave a rough estimate of over 600 individual PIs this requirement applied to. As this number was such a rough estimate, the expectation was placed on the Department Chairs and/or Vice Chairs for Research¹² to either provide or confirm the number of expected PIs who should submit DMPs within their departments. This reflects the complexity of research award data as well as the differences between how investigators conceive of a project compared to how award data is stored across multiple systems. For this and other reasons, there is not always a one-to-one relationship between an award and a data management plan. The implementation team experienced first-hand the challenges of reconciling data from the award management, payroll, and Active Directory Systems to inform policy and practice. Additionally, in no system are the members of a research team or lab captured, partly because this differs across projects and over time. In particular, it became clear that records information management (RIM) data was most robust where regulation and compliance processes were established. As data management and sharing becomes a compliance issue under the Final NIH Policy for Data Management and Sharing (DMS), we anticipate that data about and from DMS Plans will become a part of the managed RIM data.

A last step involved determining the level at which PIs were expected to complete a DMP - by study/project or by research group/lab? The conversation about this aspect of the mandate within the Advisory Board illuminated key areas of tension - School leadership wanted PIs to develop a detailed document (for themselves as well as administrators if needed) and yet did not want this to take up too much of the researchers' time. Due to this concern, PIs were asked to complete a DMP that *summarized* their research data management practices across all active research projects; effectively, one DMP per PI. While this reduced the burden on the number of DMPs to be completed per PI, it did raise a question as to the usability of these plans. If an allegation of research misconduct is received, can these generalized plans be useful? As the initiative matured, the scope of the DMP did change in order to more effectively capture the data management practices of researchers and align with external policies, such as the anticipated NIH policy.

¹¹ Ibid

¹² These positions serve as representatives across IUSM when it comes to research conducted in a given sub-disciplines

Communication & Culture

As with any mandate, communication is critical to the success of the effort. In this case, the fact that IUSM (research) leadership established this mandate and stayed consistent in their desire to achieve “100%” success, did help to spread the word and move this work forward considerably. The initial approach to communication was primarily “top down”, beginning with the Vice Chairs for Research at IUSM being appointed by the Executive Associate Dean for Research Affairs. Vice Chairs were expected to communicate information back to their respective areas and associated departments/centers. To accompany this communication, a general email blast was sent to research leadership (i.e., Vice Chairs as well as Department/Center Chairs) about the expectations, timeline, and resources available (see [Infrastructure](#) and [Training & Support](#) sections for more detail). Additionally, as the year progressed, the School newsletter included communication about the mandate in an effort to reach the general research community.

In these communications, the medical library was referenced as the primary contact and support for researchers. The Data Services Librarian devoted a significant portion of their time to this initiative and the related initiative (i.e., ELN pilot), which happened during the same time period. A key component to the support role was recognizing that this mandate required a shift in research culture/practice at IUSM. Many PIs had never been required to complete a DMP, especially those who conducted basic science research. In that case, much of the training offered to researchers involved explaining the *why* as much as the *how*. While the Office of Research Affairs championed “100% compliance” amongst IUSM PIs in completing DMPs, the librarians advocated for modest expectations. Due to the reality of this mandate requiring substantial cultural shift, even a subset of PIs completing a DMP furthered the goals of improving research integrity and data management practices at the School.

The initial communication strategy regarding this mandate was successful in informing leadership at IUSM about expectations; however, there were challenges in coordinating the expectations set by leadership in the Office of Research Affairs and the process of implementing the mandate, particularly around tracking completion rates, sending out reminders, and informing the IUSM research community about clarifications to the scope of the mandate. As the first implementation of this mandate, these types of challenges are not surprising. We attempted to streamline messaging from (and between) the Office of Research Affairs and the implementation team. Perhaps typical of a first year effort, the bigger picture for *why* this requirement was established was often lost in the *how* of implementation.

Additionally, while the medical library was excited and willing to be part of such an ambitious effort - particularly one that championed data services and the importance of research integrity - there was concern about the library being perceived as a “compliance monitor”. This was most notable in the library’s role of monitoring and reporting completion for each PI to the Office of Research Affairs. Most researchers understood the library’s role as offering support (rather than being an enforcing element of the mandate); however, it remains an aspect to be mindful of moving forward. Ideally, communications should clearly differentiate between who is holding them accountable to the requirement and who is supporting them in meeting the requirement.

Infrastructure

The initial question, logistically, was determining how researchers were going to complete this required DMP. The goal was to have researchers complete a standard template - something that was

generalizable (to a variety of research areas) as well as something that provided enough information to enable IUSM research administration in responding to scientific integrity issues (if any arose in connection to a particular PI). To that end, the DMP template used was adapted from one created by Heather Coates, member of the implementation team and IU Research Data Steward. This [particular DMP template](#) is designed to be operational and in the context of [IU Guidance for the Management of Research Data](#). This means that it is intended to be a “living document”, which is updated as a project progresses to reflect how research is actually conducted as well as to document changing project team members, responsibilities, etc. Unlike a funder DMP, an operational DMP captures sufficient detail to support the daily conduct of research.

One key change made to this original IU DMP template was shifting the scope from a project-specific DMP to a summary level DMP. Shifting the scope of the DMP involved discussion to determine how to phrase questions so that information captured reflected typical data management practices for a given PI. Doing so also raised the questions: 1) How useful would the information gathered be for the purposes of capturing a particular research group’s data management practices? And 2) If IUSM Research Affairs and the Research Integrity Office could rely upon these DMPs to answer questions about scientific integrity for a given PI and their research group?

The second major change to the original template was to convert many of the questions to structured responses (i.e., options to choose from), primarily to minimize the burden on PI. Questions that were good options for structured responses related to: roles within the team/lab, training topics covered within the team/lab, data types, and data storage needs and platforms used. Questions regarding data description, quality assurance and quality control processes, data retention periods, and data sharing were asked in an unstructured way (i.e., text box response), which elicited confusion from IUSM PIs about the scope and level of detail to include in their responses. For these key areas, as well as several others, the implementation team worked with early adopters and pilot teams to develop a range of sample responses. Other changes to the template related to modifying the prompts to reduce confusion about terminology. Perhaps unsurprisingly, the terms which seemed to be the least well-understood were data sharing, data retention, data archiving, data preservation, and data curation.

Following customization of the IUSM [DMP template](#), the next step was determining the tool to use to collect this information. While Indiana University does have access to the DMPTool, the implementation team decided to use the enterprise-level Qualtrics product. We determined that DMPTool was not fit for our purposes due to the need to customize the questions, use advanced logic, and IU information security requirements.

Even during the first year, we identified a number of limitations associated with using Qualtrics in this way. A key challenge was related to generating multiple versions of plans should PIs need or want to update them. A clarification made to the mandate was that PIs were only expected to submit one DMP annually so they did not have to update their existing DMP if new personnel joined their research groups (for example). However, in the spirit of encouraging DMPs to be a “living document”, the implementation team wanted to offer PIs the ability to update and version their DMPs as desired. Additionally, there was information collected in the Qualtrics DMP form that was repetitive (for example - people’s names, emails, etc.); almost immediately researchers questioned why information that was already collected by the School could not be pre populated into the DMP template. These limitations and feedback from the research community led the implementation team to work with University IT Services (UITS) to seek other solutions. We collaborated to identify 1) methods for harvesting existing

RIM data for integration with the DMP template (e.g., names/emails, grant information) and 2) a platform that was approved for integration with existing RIM and other systems.

The implementation team recognized from the outset that Qualtrics was not a long term solution for administering the DMP mandate; however, the decision to use it as the starting point was a reality of the constraints. We had to start with something readily available and could be implemented on a timeline that was already set.

Training & Support

As the primary support and training resource for this mandate, the medical library (specifically those on the implementation team) worked to build resources, deliver presentations, and offer one-on-one consultation as requested. In addition to developing the DMP Template in Qualtrics, library personnel developed a summary of the mandate that was heavily used in promotion, as well as an FAQ page. To help with dissemination and to centralize all these resources, the implementation team created an internal webpage where IUSM researchers could read about the mandate expectations, refer to FAQs, access the DMP template in Qualtrics, and get information about who to contact with questions.

The medical library reference email was the contact provided in case IUSM researchers had questions and/or training needs. Initially, all inquiries and training were handled by the Data Services Librarian within the medical library. The Data Services Librarian presented regularly as requested at department and center meetings and was frequently asked, especially as the deadline for the first year approached, to hold one-on-one consultations with PIs to answer questions as they completed their DMPs. This kind of engagement allowed for the identification of additional resources to help researchers comply with the mandate. Two of the most notable requests were: 1) example DMPs from different departments/disciplines to reference and 2) boilerplate language to add to different sections of the DMP based on disciplinary area. A number of PIs also noted that they were not sure that their data management practices were all that great and appreciated the boilerplate language not only to assist with completing the DMP, but as an education mechanism to learn about best practice/resources they were not using. Within the Qualtrics DMP template, the implementation team provided some sample language for use in different sections. While many researchers did use the sample language provided in the template, it was unclear to what degree that language reflected their practice(s) or if it was just a means to quickly submit the DMP. However, the communication around this effort and process of completing the IUSM DMP has contributed to greater awareness of issues, good practices, and the library's role in supporting research data management.

Beyond complying with the DMP mandate, discussions about the DMP requirement uncovered additional areas where research at IUSM required additional data governance practices and policies to be developed. A key example of this was research data retention expectations at IUSM. Similar to many institutions, Indiana University has records retention policies for administrative records but no such policy exists at the system level for research data. In part, this is due to widely varied regulatory requirements and disciplinary norms. Decisions about retention of research data are often left to the discretion of the PI unless set by the funder, in research agreements, or information-specific state or federal regulations. Practically, questions about research data retention commonly arise when researchers are leaving IUSM or IU and in the context of data transfer conversations. This is another area where the process is often left to the discretion of departments and research labs rather than standardized across the IUSM research enterprise. In this manner, the DMP mandate provided an

opportunity to highlight larger research data governance issues - and to do so in a way that demonstrates how policy and governance have a direct impact on (and support) building a culture and practice of strong information and data management.

Reflecting back as we look ahead (Year 2: 2020-2021)

People & Roles

During the second year, the initiative experienced significant changes in personnel. Both the Executive Associate Dean for Research Affairs and the RLML Data Services Librarian left IU for roles at other institutions. Additionally, several faculty champions on the Advisory Board and research teams participating in the pilot also left IU. The turnover of Advisory Board members resulted in revisiting several of the assumptions and approaches made during the first year, while demonstrating the value of discussing data management planning, processes, and policy beyond the context of an individual faculty member's research practices.

Scoping the Mandate

While the activities of the ELN rollout were relatively well-defined, the work associated with the DMP mandate was less clear.¹³ As the initial information was released, it became clear that researchers had varying levels of familiarity and comfort with data management plans. In the first year (2019-2020), the choice to require one DMP per Principal Investigator (PI) generated confusion as to how to report practices that differ between projects. Fortunately, the feedback received resulted in a change of scope for the mandate; for the second year (2020-2021), PI were asked to provide a project-level DMP for every active federally-funded research project. Support for this change was aided, in part, by the anticipated release of the [NIH Final Data Management and Sharing Policy](#).

The work to scope this mandate in the first year emphasized the reality that research data management and sharing are complex and dynamic issues that vary across institutions, fields of research, and research teams. For all the high-level conversations led by organizations such as the National Academies of Science, Engineering, and Medicine, National Institutes of Health, National Science Foundation, and others, there remains enormous variation in the levels of awareness and understanding of these issues and effective practices. The lack of consistent terminology, or common terms used in many ways, further hampers discussion across these boundaries. Even within IUSM, the range of experiences and perspectives is wide-ranging. Given the broader initiative within which this effort was happening, there was not enough time to conduct an in-depth conversation to co-create a shared understanding of these issues across the school. Instead, the implementation team did our best to bring common terminology, definitions, and approaches to the immediate work of the Advisory Board. A related challenge was the strong conviction by many researchers that information being asked for in the DMP template was already provided in their IRB or IACUC protocols, grant proposals, and other research documentation. While it is true that some elements of data management are captured in these documents, the full picture of research data management is not captured elsewhere. We believe this stems from a lack of clear and common terminology as well as a lack of understanding or value placed on research data management as a central aspect of good research practices. An interesting example

¹³ <https://scholarworks.iupui.edu/handle/1805/23220>

of this was the struggle to describe quality assurance and control practices (within the DMP). For some researchers, QA/QC practices apply primarily to the research protocols rather than to the data itself. Additionally, there was some confusion about the distinction between research data and the materials and samples from which they are derived. While the NIH has distinct policies, researchers may not see the differences. In short, there were often fundamental conversations around research practice and training that were sidelined due to the urgency around implementing the mandate.

Communication & Culture

There were often differing expectations regarding how the work of the Advisory Board (AB) would get done. The implementation team sought to foster conversations with a looser and more participatory format, while many IUSM departmental leaders expected a more structured format in which they were briefed on discrete options and made clear choices between them. Additionally, the decision-making authority of the AB within the context of the school was not always clear. While we recognized that the expertise and leadership was invaluable to making this endeavor viable, we struggled to develop clear and defined decision-making processes for the group. During the rollout period, timelines and goals shifted rapidly and developing rapport within the AB group was not prioritized. Although we would have liked to discuss how issues such as training, data sharing, and data retention varied widely across research communities and teams, practical issues took precedence. In a 1-2 hour monthly meeting, it was a struggle to introduce key information and engage them in critical thinking around complex issues. The implementation team often did not have sufficient time to provide necessary context to frame the question or issue; thus, we adopted a satisficing approach in which we brought issues and recommended solutions to the AB, rather than involving them in the time consuming process of brainstorming potential solutions. In several cases, they proposed potential solutions that we had not considered; however, having concrete options to react to seemed more effective in stimulating conversation and decision-making. Finally, the membership of the AB was not always representative of IUSM faculty, with fewer early career researchers than initially planned. Thus, the Implementation Team intentionally sought input from early career researchers and those managing smaller teams to incorporate a broader range of experiences and perspectives into the decision-making process. Over time, the role of the Board in making recommendations to the Executive Associate Dean for Research Affairs solidified, despite significant changes in Board membership.

Several scenarios arose that exemplify the issues faced in changing the culture of research data management and sharing.

Scenario 1: Department/Vice Chairs relied on their own experience as researchers during the conversations. Their experiences often tended to reflect research models involving large, relatively complex teams with members at several levels - research scientists, post-docs, full-time staff, graduate students, undergraduate students. However, there are many PIs in the School of Medicine whose research is conducted at a smaller scale, with many teams having fewer than 5 members. Additionally, there were few early-career faculty represented. When such researchers were not represented within the AB, we felt it was our responsibility to advocate for their needs, as far as we were aware of the differences in needs. This sometimes put us in the difficult position of challenging the assumptions and expertise of senior leadership represented on the AB. While we saw ourselves as advisors and advocates for researchers, more direct representation would have been beneficial in broadening the conversation.

Scenario 2: Initially, the Implementation Team assumed general familiarity with emerging practices and standards relating to data sharing, data reuse, Quality Assurance/Quality Control, data retention, and data repositories. For example, some did not like the term data sharing because it has a negative connotation within their field; reusing another teams' data is not viewed as a positive thing. The conversations about data sharing often revolved around fundamental issues and deeply rooted in personal experiences of the board members, rather than the broader issues of research integrity and open science. Exposure and awareness of these topics varied widely, so specific examples and issues resonated with some, but not all, of the Advisory Board members. It was difficult to reach consensus, in part due to the lack of shared understanding of what these issues entail across the wide spectrum of biomedical research. Additionally, because we entered into these conversations with little knowledge of what the AB members knew, early conversations were not as productive as they might have been. Over time, we realized that AB discussions worked best when we provided a quick introduction to the issue at hand, framed the question or decision as clearly as possible, and proposed a few options.

Scenario 3: Advisory Board members and PIs repeatedly advocated for highly structured responses to the DMP and requested boilerplate language for the free text questions. This is reflective of their understanding of a DMP as a product that is required for compliance, rather than recognizing data management planning as a valuable part of the research process that is integral to producing high-quality research. We sought to balance the approach taken so that it emphasized both the planning process and the required DMP, without placing undue burden on PI. Maintaining this balance was an ongoing struggle because we did not yet have data to demonstrate the value of the information captured within the DMP and balance that against the cost of implementing the DMP mandate.

Scenario 4: Early on, there was little awareness of the gaps in funder, publisher, institutional, and school level policies. In particular, many members of the AB believed that the institution and school had clear policies regarding ownership, management, retention, and sharing of research data. In demonstrating the most significant gaps, particularly related to retention and sharing, we assembled a list of all policies relating to research data. As we discussed the gray areas, gaps, and support for researchers in navigating somewhat complex decisions about retention and sharing, the AB recognized that additional policies were necessary to provide clearer guidance for PI. In 2021, the School adopted a [data ownership and retention policy](#) that sets a minimum retention period of 7 years after the last publication related to the data in question. Late revisions to this policy from IUSM personnel not involved in the AB introduced conflicts with IU system-wide policy, which are in the process of being resolved.

Scenario 5: As we discussed the support services available, it became apparent that levels of trust in those services varied widely. Generally, institution-wide services were sometimes seen as unresponsive, while services provided within the School and departments were viewed as more supportive and relevant. Greater exposure to a particular unit, whether directly or indirectly through colleagues, was often associated with greater trust and use in their services. However, this continues to be a challenge.

Our ability to work together in implementing this requirement relied on trust between the AB members and belief that it was a worthwhile process for the School, despite the lack of incentives. Developing buy-in and trust takes time and may be best developed by doing the work together. This emerged over time, as we shared information, expertise, and began to recognize that we were building a network of support for this effort and champions for the role of information management in the research enterprise.

Infrastructure

While Qualtrics was a widely available tool that was known to the Implementation Team, several limitations quickly became apparent. Feedback during the 2019-2020 completion cycle indicated that the team/lab level DMP introduced significant confusion. Investigators tend to think about data management practices as associated with specific projects. Even as we were supporting PIs in completing the DMP v.1, we were working to improve the experience. The Fireform system, an electronic data collection and workflow tool is secured and approved for integration with other IU systems. While the content of the DMP remained largely the same across both years, particularly the free text responses, shifting the [DMP template to Fireform](#) allowed us to pull in existing award, personnel, and other data to reduce the burden of completing a project-level DMP. This allowed us to be responsive to the feedback provided, while ensuring that the unique information captured by the DMP was not lost, specifically:

- Data collection procedures
- Quality Assurance & Quality Control procedures
- Data retention periods
- Data sharing practices
- Access permissions when team members leave

We also asked respondents to identify relevant funder data management or sharing requirements, provide citations for data being reused, and identify applicable standards for data and metadata so that we could gauge awareness of these issues and prevalence of such norms. These fields were free-text responses so that we did not limit the options available.

Given the timeframe set by leadership, we sought to create functional processes without over-engineering them. Thus, the initial guidance we provided was relatively general, leaving the investigators to gauge the appropriate level of granularity for the free text responses. However, the questions received provided valuable guidance for developing the FAQ and expanding the sample responses. A particular topic of confusion was whether the mandate applied to various award types (i.e., R01, P, K, KL2, etc.).

We continue to advocate for the data management planning process and DMP mandate as valuable for facilitating research integrity and documenting research data. There are many types of DMP, not all of which are developed to support the work of PI and their research teams. As we looked ahead to the implementation of the NIH Final Data Management and Sharing Policy, we reflected on the lessons learned from this experience to prepare us for and potentially complement to funder requirements.

Training & Support

The guiding principle for the Implementation Team in this effort was to advance the data management and data sharing practices of IUSM researchers. From the beginning, we strived to minimize the burden on researchers, while continually centering their perspectives and experiences in decision-making. In-depth consultations helped us to raise visibility of certain concerns, particularly the burden of navigating multiple services, systems, and processes that are siloed. It is often left to the researchers to make sense of incomplete guidance and fill in the gaps. As this effort continues, a key question is how research support and research data governance matures along with it.

We have used information submitted in the DMP to inform support and enhance the guidance for developing NIH DMS Plans. This return is key for maintaining buy-in and demonstrates clear benefits both to PI and departments of the time invested in completing these plans. By making this information accessible to the departments, we are supporting their strategic planning, compliance with annual departmental reviews, and furthering conversations related to data management and sharing. It is our hope that the availability of these DMP will facilitate conversations about research data support and cyberinfrastructure across all levels of IUSM. We are in the process of identifying whether and how departments are making use of this information. The AB was provided with a [synthesis of the quantitative responses for 2020-2021](#) as well as a qualitative analysis of a subset of 2021 free text responses¹⁴. AB members were encouraged to distribute these reports to their departments, alongside other related communications.

Due to the enormous variance in the research context influencing data management and sharing practices, it is incredibly difficult to offer broad prescriptive recommendations that work for all types of research, disciplines, and data. While there are foundational practices¹⁵, strategies can be implemented in many different ways and using a range of technologies. At the school-level, a key first step is to identify and describe influential contextual elements and offer support for researchers in prioritizing the many decisions to be made. The key contextual elements we have identified include the following:

- funder policies,
- publisher data availability/access requirements,
- contractual agreements,
- institutional policies and procedures related to data retention,
- transfer/sharing data, and security for high-risk data,
- potential for commercialization and/or patents, and
- community norms related to reproducibility, collaborative data collection, and data sharing for reuse.

Recognizing these characteristics or dimensions of data that affect data sharing can support researchers in differentiating between different streams or types of data, data classifications and risk levels, and data repository options. Where prescriptive requirements from funders or programs, publishers, and research communities exist, we direct researchers to those relatively defined pathways. Where no such requirements exist, we encourage consideration of disciplinary or community norms and the goals of the research team as guiding factors in decision-making.

The library offers unique expertise and perspective in these decisions which is not present in IT, information security, or compliance approaches. Library-based support tends to consider issues of provenance, preservation, and curation through a discipline agnostic lens. This position is valuable for identifying trends and needs across schools and departments, so that we can improve our own services while advocating for broader solutions that address previously unrecognized needs. As we discussed data management and sharing with research teams, we learned about particular points of friction, service gaps, and the invisible burden of managing research projects and teams. As research has grown more complex and collaborative, the burden of navigating the associated bureaucracy has fallen

¹⁴ Dolan, L., Whipple, E., & Coates, H. L. (2022). Copy and Pasting vs. Customization: Using Qualitative Analysis of NIH Grantees' Data Management Plans to Shape Future DMP Support. Poster presented at the Research Data Access & Preservation (RDAP) Summit. Available at <https://hdl.handle.net/1805/28526>.

¹⁵ Briney, K., Coates, H., & Goben, A. (2020). Foundational Practices of Research Data Management. *Research Ideas and Outcomes*, 6, e56508. <https://doi.org/10.3897/rio.6.e56508>

largely on researchers. While this is not a problem that the library can solve in isolation, we can build on existing relationships and collaborations to develop a more cohesive network of research support that minimizes the obstacles experienced by researchers in managing and sharing their data.

Many institutions are adopting technological solutions, such as research portals and storage selection tools. Developing and sustaining communities of people is more difficult but vital. We believe that data management and sharing practices will only be sustainable when the culture of the school, departments, research teams, and the disciplines within which they are embedded values this work. While this effort addressed the top level of the research culture change by making it required, we are still working towards making DMP rewarding, normative, easy, and possible (Figure 1).

Changing a Research Culture

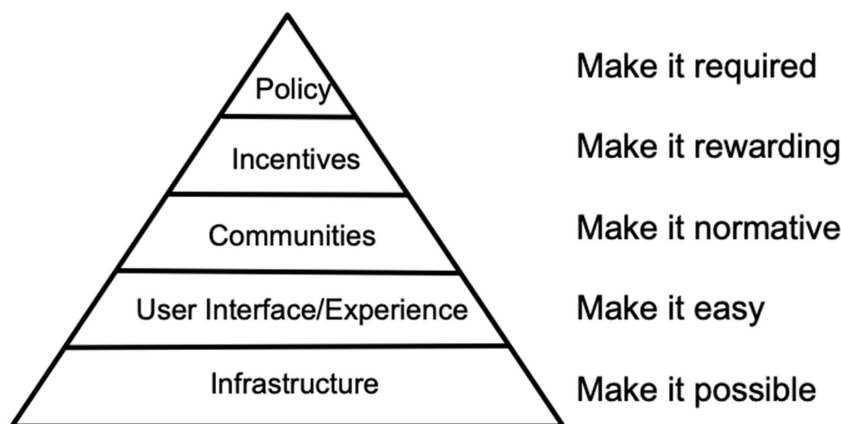


Figure 1: How to change a research culture¹⁶

Year 3 (2021-2022) & Preparing for the NIH DMS Policy

In the third year, training and support began to shift from addressing logistical concerns to more substantive data management and sharing questions. Input and feedback from researchers completing the DMP as well as the information submitted within them have offered valuable information and insights into the current understanding¹⁷ and needs of IUSM researchers. In particular, the decision-making related to sharing data is complex, contextual, and an issue often not considered until publication. The RLML continues to support researchers by providing presentations, answering questions, and updating guidance for completing the IUSM DMP.

However, as we prepared for the Final NIH Data Management and Sharing Policy, which went into effect in January 2023, the conversation shifted again towards logistical and compliance questions. It was only in the last months of 2022 that the details of the NIH implementation were released, so much remains to be determined. Understandably, PI struggled to reconcile the new NIH policy with the IUSM DMP mandate. A key challenge was shifting from describing established practices for an ongoing

¹⁶ Mellor, D. (2021). Improving norms in research culture to incentivize transparency and rigor. *Educational Psychologist*, 56(2), 122–131. <https://doi.org/10.1080/00461520.2021.1902329>

¹⁷ Dolan L., Whipple E. C., Coates H. L. (2022). Copy and Pasting vs. Customization: Using Qualitative Analysis of NIH Grantees' Data Management Plans to Shape Future DMP Support. RDAP conference (virtual). <https://hdl.handle.net/1805/28526>

project to describing a tentative plan for a proposal not yet funded. Project proposals are not operationalized until project start-up, so developing a plan during the proposal development phase requires PI to consider the pertinent issues and sufficiently address them while remaining appropriately vague and brief. This is not impossible, but it is particularly challenging the first time. For example, use of a particular system for data capture may no longer be viable if the institution discontinues the license or a better tool becomes available. Another example is the constantly shifting access to cloud storage platforms such as Box, Google Drive, and Microsoft 365. Fortunately, the areas of commonality - describing data types, tools, and metadata - helped us to focus on the more challenging decisions related to data sharing.

The data available from previously submitted IUSM DMP was invaluable in identifying common data types, workflows, etc. Several were rich in description, enabling us to develop sample responses for the NIH DMS Plan that were based on local research. Free-text responses indicated the relative familiarity with various issues such as standards, metadata, storage options, sharing, which has informed our approach to supporting researchers in developing NIH DMS plans. Specifically, the responses about data sharing were illuminating, primarily by describing the variety of ways in which data sharing is understood across biomedical research communities. These responses further underscore the need to develop a shared understanding of the spectrum of data sharing practices as well as institutional policies and processes related to managing and sharing data.

Though awareness of data management and sharing issues has increased, there is still room for improvement in the areas of policy and processes within the school. The main institutional policy related to data sharing is undergoing revision to clarify processes for research data. Additionally, institutional documentation is being developed for data transfer and data retention processes when individuals leave the institution. Key processes related to onboarding & exit procedures, data transfer, and data retention still need to be developed and/or formalized at the school, department, and team levels. While it is reasonable to expect each research team to have documented team and/or study procedures, it may not make sense for a department to dictate specific standards. The implementation of research data policy, process, and procedures requires a layered approach that allows customization where necessary.

The IUSM DMP mandate will continue in 2023, although we anticipate many PI will submit DMS Plans in lieu of completing the more detailed DMP. In 2023, at least, the NIH DMS Plan will be accepted as meeting the IUSM requirement. However, we know that plans submitted at time of proposal are not operational or fully formed. Much work will need to be done during the project start-up period to operationalize plans for storage, documentation, and workflows. The timing of the IUSM DMP mandate revealed several challenges of a top-down approach while also suggesting potential opportunities for continued engagement with research teams. Since the mandate did not align with the funding application calendar, it did not enable us to offer support in data management planning prior to submission of funding applications. The proposal development period is an important one for integrating data management activities into the conduct of the study, but only if PI are willing and allow sufficient time to engage with the planning process. The project start-up period may prove to be a more fruitful time to support research teams with integration of data management and sharing practices. With the NIH DMS Policy implementation, we have the opportunity to try this approach. More often, researchers approach the library about data sharing further downstream in the research process, typically when manuscripts are being submitted for publication to journals with data access policies or when closing out a project. While it is possible to support data sharing and the requisite curation at

these points in time, it is far more time-consuming and likely to be unfunded. By connecting outreach to defined points in the NIH award management process, we hope to have more opportunities for connecting with research teams during the pre-proposal and active project phases.

Conclusions

No single research support unit can address all the data management and sharing needs of a research project or team. Thus, effectively supporting and training researchers is a collaborative and interdisciplinary effort undertaken at multiple levels. Often, the most pressing issues faced by researchers fall between the scope of responsibilities of units such as the Human Research Protection Program, research administration, research compliance, information technology, information security, data governance programs, and libraries. As institutions are challenged to support broader data sharing, we will need to develop a more connected and collaborative network of support for researchers. In some cases, our researchers are experts themselves on the infrastructure, policy, and people necessary for effective data stewardship. Where such experts exist, institutional support networks must include them.

Reflecting back over the past five years, we have identified a few things that we might approach differently. Primarily, we would have advocated more strongly for setting aside time during the pilot to engage the leadership in more meaningful and in-depth conversations about the value of data management and sharing plans within the School, Departments, and teams/labs themselves. Once the implementation timeline began, it was very difficult to have these conversations. However, grounding the initiative with specific and measurable outcomes may have resulted in greater use of the information gathered via DMP beyond the librarians and Advisory Board. Additionally, we would have preferred to devote more time and effort on education and training related to documentation and metadata, standards, retention, and options for sharing data. The many competing demands for PI time make it difficult to propose further training. Finally, there are unresolved issues of trust between faculty and research administration, as well as other centralized research support services. In part, this is due to the context in which this initiative began. It was very much a top-down initiative, led by the IUSM Executive Associate Dean for Research Affairs. Additionally, the DMP initiative was just one part of a broader effort to improve efficiency of lab space and information management in order to increase research awards received. In this context, risk management and compliance were prioritized over cultural change. In part, this may be due to the industry background of the consultants and the Associate Dean for Research Affairs, at that time. However, it is important to recognize that sponsorship by IUSM executive leadership enabled the initiative to proceed further and at a faster pace than a more grassroots-based approach.

Many questions remain, some of which are listed below.

- How can the Advisory Board effectively analyze and distribute information gleaned from submitted plans thus far?
- How can departments and schools make use of the information provided in DMS Plans?
- Where in the existing research processes (proposal development, internal review by the Office of Research Administration, ethical review, project startup, project closeout, etc.) can we integrate support?
- Are these requirements, both the IUSM DMP mandate and the NIH DMS Policy, actually improving research data management within IUSM?

- How can we shift practice towards machine-actionable DMS Plans?
- How can institutions and/or schools create space to have conversations about emergent issues that span silos?
- How can institutions, schools, and departments make effective use of the researchers who have expertise in data management, sharing, and stewardship?
- How can institutions, schools, and departments reward good data management and sharing practices?

Supplemental materials

We have created an [OSF space](#) to share the following:

- DMP templates
 - [2019 Qualtrics version](#)
 - [2020-2022 Fireform version](#)
- Training materials
 - Early adopter pilot - [Lab X consult presentation](#)
 - Early adopter pilot - [Lab Y consult presentation](#)
 - [Briefing slides on ELN + DMP](#)
- Quantitative analysis of structured responses (unpublished report)
 - <https://osf.io/6q84u/files/osfstorage/6489d2d884f52c0249fe5b53>
- Qualitative analysis of free-text responses (RDAP 2022 poster)
 - <https://hdl.handle.net/1805/28526>