Working Group Recommendations for an Indiana University Research Data Commons

Final Report, June 30 2022

Kosali Simon (Chair)
Distinguished Professor, Herman B Wells Endowed Professor, O'Neill Chair, O'Neill School of Public and Environmental Affairs, Associate Vice Provost for Health Sciences

Diane Dallis-Comentale
Ruth Lilly Dean, University Libraries

Shiaofen Fang
Professor and Chair, Computer and Information Science, IUPUI School of Science

Kun Huang
Professor and Chair, Biostatistics and Health Data, IU School of Medicine and Fairbanks School of Public Health, Director for Data Science and Informatics, IU Precision Health Initiative, Associate Director of Data Science, IU Simon Comprehensive Cancer Center

Matt Link
Associate Vice President for Research Technologies, University Information Technology Services

Gerardo Maupomé
Professor, Global Health, Associate Dean of Research, IUPUI Fairbanks School of Public Health

Filippo Menczer
Luddy Distinguished Professor of Informatics and Computer Science; Luddy School of Informatics, Computing, and Engineering, and Director, Observatory on Social Media

Aaron Neal
Associate Vice President, Enterprise Systems, University Information Technology Services

Brea Perry
Professor, Sociology, College of Arts & Sciences, IUB, Interim Vice Provost for Research, IUB and Associate Vice Provost for Social Sciences, IUB

Beth Plale
Michael A. and Lauren Burns McRobbie Bicentennial Professor of Computer Engineering; Luddy School of Informatics, Computing, and Engineering, and Pervasive Technology Institute, Office of Vice President for Information Technology

Angie Raymond
Associate Professor of Business Law and Ethics, Kelley School of Business, Director, Program on Data Management and Information Governance, Ostrom Workshop

Sarah Wiehe
Associate Professor of Pediatrics and Associate Dean of Community & Translational Research, Co-Director, Indiana Clinical and Translational Science Institute

Contents

Introduction and Executive Summary

Starting on April 28, 2022, our Working Group set out to make recommendations for an Indiana University Research Data Commons (**IU-RDCom**), a strategy to identify and meet the growing needs associated with research data at our university. Developing a way to find, access, use and share research data is an iterative process that many peer universities are also currently pursuing. The process requires a university to identify researchers' needs, catalog services that currently exist, understand how they can be leveraged along with new investments to meet these needs, and to establish a sustainable governance structure for developing and evolving the IU-RDCom. A competitive research data infrastructure will pay for itself in many ways through new external funding while it increases our scholarly, educational and service missions.

The present report outlines our recommendations to VPR for practical steps IU should pursue in the near-, medium-, and long-term.[1] In brief, these **recommendations** are to:

1) Establish a governing body to coordinate a research data commons.
2) Task the governing body with implementing and building on our recommendations.
3) Encourage IU leadership to communicate and promote IU's strengths in research data.
4) Provide short-to-medium run financial support for building a foundation for the data commons.

As stated in the charge to the Working Group (WG), the broad mission of the IU-RDCom is multifaceted: to serve as a university-wide resource for discovering, sharing, and accessing data resources across the IU community; to build on our world-class strengths in centralized cyberinfrastructure and other areas to present researchers with easier and more integrated pathways to our data resources; to enable richer training opportunities for students; and to empower IU to better serve local organizations, our state, and other partners.

With exponential recent growth in the role of data in society and in scholarship, the need for universities to engage in strategic planning to strengthen research data infrastructure has been emphasized in new reports from the American Association of Universities (AAU), the Association of Public and Land-grant Universities (APLU), the National Science Foundation (NSF), and the National Academies of Science, Engineering, and Medicine (NASEM).[2] NASEM committees are also presently guiding the vision for federal research data infrastructure for the 21st century for similar reasons as for academia.[3]

From communications with research data leadership at peer institutions over the course of our work, it is amply clear that other universities are also prioritizing central-level strategies to meet these growing research data needs in academia.

Our Recommendations

1. IU should establish a coordinating structure, IU-RDCom, with an executive group staffed by VPR, Libraries, and VPIT, and an oversight committee of faculty selected by UFC.

---

[1] https://www.aau.edu/accelerating-public-access-research-data
[2] https://nap.nationalacademies.org/catalog/25116/open-science-by-design-realizing-a-vision-for-21st-century and https://www.nationalacademies.org/brdi/board-on-research-data-and-information
[3] https://www.nationalacademies.org/our-work/the-scope-components-and-key-characteristics-of-a-21st-century-data-infrastructure-a-workshop

2. The IU-RDCom will be tasked with implementing recommendations laid out in this report to provide a "front door" to enable discovering, sharing, and accessing data resources across the IU community. The creation of this infrastructure is iterative and requires a long-term governance structure to build in a phased manner while periodically collecting and responding to feedback.

3. IU leadership should develop internal and external communications strategies to promote IU's strengths in research data and its commitment to building the IU-RDCom as a vital resource for fulfilling our mission. IU has an opportunity to be a leader among peer institutions at this critical time of national attention to the academic research data enterprise.

4. IU should provide financial resources to the IU-RDCom groups to enable initial staffing and development within VPR, Libraries, and VPIT, with an expected transition to being 50% externally funded within five years.

Research Data Commons Working Group Committee Members

We appreciate the opportunity to participate in collecting feedback, deliberating and offering recommendations for the IU-Research Data Commons.

- Kosali Simon (Chair)
  - Distinguished Professor, Herman B Wells Endowed Professor, O'Neill Chair, O'Neill School of Public and Environmental Affairs, Associate Vice Provost for Health Sciences
- Diane Dallis-Comentale
  - Ruth Lilly Dean, University Libraries
- Shiaofen Fang
  - Professor and Chair, Computer and Information Science, IUPUI School of Science
- Kun Huang
  - Professor and Chair, Biostatistics and Health Data, IU School of Medicine and Fairbanks School of Public Health, Director for Data Science and Informatics, IU Precision Health Initiative, Associate Director of Data Science, IU Simon Comprehensive Cancer Center
- Matt Link
  - Associate Vice President for Research Technologies, University Information Technology Services
- Gerardo Maupomé
  - Professor, Global Health, Associate Dean of Research, IUPUI Fairbanks School of Public Health
- Fillipo Menczer
  - Luddy Distinguished Professor of Informatics and Computer Science; Luddy School of Informatics, Computing, and Engineering, and Director, Observatory on Social Media
- Aaron Neal
  - Associate Vice President, Enterprise Systems, University Information Technology Services
- Brea Perry
  - Professor, Sociology, College of Arts & Sciences, IUB, Interim Vice Provost for Research, IUB and Associate Vice Provost for Social Sciences, IUB
- Beth Plale
  - Michael A. and Lauren Burns McRobbie Bicentennial Professor of Computer Engineering; Luddy School of Informatics, Computing, and Engineering, and Pervasive Technology Institute, Office of Vice President for Information Technology

- Angie Raymond
    - Associate Professor of Business Law and Ethics, Kelley School of Business, Director, Program on Data Management and Information Governance, Ostrom Workshop
- Sarah Wiehe
    - Associate Professor of Pediatrics and Associate Dean of Community & Translational Research, Co-Director, Indiana Clinical and Translational Science Institute

Report Process Steps

The Research Data Commons Working Group used the following steps to research and write our report:

1. Solicited input from faculty, students, and other constituencies.
2. Identified useful models and best practices from other universities and institutions.
3. Developed a blueprint and timeline for implementing a Research Data Commons.
4. Addressed key issues such as scope, access, governance, funding, staffing, outreach, training, maintenance, and anything else considered important to the success of this initiative.

Providing outreach and gathering input from the broader university and research communities was a major priority for the WG and we solicited participation in many ways. We disseminated a university-wide announcement on May 16th, 2022 (Appendix 2), held a townhall webinar that was attended by over 50 people on May 23rd, 2022, and conducted a survey that was completed by 84 people as of June 1st, 2022 (results were summarized in an Appendix that is removed for this version of the report). Our WG members also conducted outreach individually by contacting numerous colleagues and groups, including those in peer institutions, government, and national organizations involved in academic research data, as well as research funders. We are grateful to our internal and external colleagues for providing the WG with the views on which our suggestions are based. **While our efforts during the past two months were extensive, we underline the need for ongoing collection of feedback and continued outreach given the dynamic nature of developing, implementing, and sustaining a well-functioning IU-RDCom**.

Report Construction

Our report consists of six sections:

Section 1: "Front Door" Infrastructure

We propose and examine infrastructure that provides researchers with a "front door" to research data and services at IU, both for consuming as well as producing data. We discuss the coordination needed for an enhanced experience for a variety of different personas at IU. Each persona includes a potential workflow of IU systems, which would be designed to minimize hurdles currently faced.

Section 2: Blueprint Proposal for a Website

We propose a potential website (a blueprint is provided) as one of several tools that will build infrastructure for a "front door". The mere act of planning a website will advance conversations; the website is not the end goal by itself. The website links out to the services and data available, with a vision of future, deeper integration and the development of new services. It will cover what is available to users seeking existing data, including data for users hoping to deploy primary data for secondary use, as well as for publishing and sharing data created. We build on the final report that was released by IUPUI's working group on public access to research data.

Section 3: Stakeholder Input Regarding Reduction of Bureaucracy

We summarize input collected on ways to reduce bureaucracy and waste. Please contact us for a summary of survey responses; it is removed in this version of the report for wider circulation. Respondents described areas with room for improvement, such as delays in accessing and licensing data, lack of knowledge about available datasets, data storage problems, and the need for resources, including staff to help with access, archiving, and sharing of data resources.

Section 4: A Sample of Exemplary Initiatives at Other Universities

We list exemplary initiatives at other universities**,** as well as consortia that might provide us with ideas for funding models.

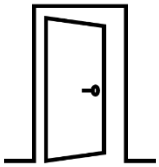
Section 5: Process Proposal for Short to Medium-Term Budget

We detail a short to medium-term budget for VPR, Libraries and VPIT. We constructed this budget after speaking with associated stakeholders at IU on the financial needs for providing coordination services we recommend(Budget details removed for public version)

Section 6: Possible Sources of Funding

We have identified possible sources of funding including NSF/NIH data hub grants, foundations, donor campaigns, indirect cost recovery, and IU internal funds—to support immediate actions as well as a longer run funding sustainability plan.

Section 1: "Front Door" Infrastructure



IU researchers and staff who support researchers should be able to easily navigate the data assets, services, requirements, and procedures needed to carry out their work through a welcoming "front door". This "front door" is the Indiana University Research Data Commons (IU-RDCom). The need for a "front door" was vigorously accentuated throughout the input we received from the survey, town hall, and other means for providing voice.

IU is not alone in articulating the need to increase attention to research data, to make processes simpler, and to amplify institutional resources through better coordination. The AAU and APLU formed working groups in 2018 and 2021, respectively, producing a 2021 report to "help universities create robust systems for ensuring effective public access to high-quality research data." This does not mean that the "front door" experience should be simply a website or be owned by any one existing entity. Rather, this aim reflects a unified view echoed by a network of all entities involved. Whether through a brochure-type e-communication received by faculty and others being recruited to IU, through existing research workshops or programs, or when a researcher plans a new scholarly activity at IU, there should be a shared understanding of where resources exist, the steps to go through, and who serves as navigators (or point-persons) to help researchers through the processes.

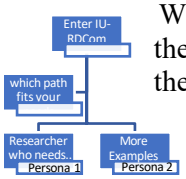
IU licenses, has expertise in, and produces vast amounts of data for research. A major goal of the proposed IU-RDCom network will be to enhance the IU community's knowledge of and ease of access to these resources. Research data also represent a resource to be stewarded with utmost care and respect for the individuals whose actions generate the information, while ensuring contributions to open and transparent science and scholarly creativity. A second goal is to increase the IU community's knowledge of and engagement with improving existing data policy. We recommend the IU-RDCom undertake a review of the extensive IU data infrastructure policy that currently exists. And once affirmed or amended, the IU-RDCom should be tasked with implementing actions to increase faculty understanding of these policies. If we do not, researchers may risk jeopardizing their research careers, possibly receiving fines or penalties from external bodies or not being competitive for research funds or scholarly outlets that require robust replication proof prior to publication.

IU researchers are both consumers and producers of research data. There are different criteria for a front door experience for **consumed** data relative to **produced** data. Some of the principles behind a research data strategy for the 21st century being espoused by national research bodies have to do with enabling the consumption of existing research data produced outside IU as a separate concept from enabling researchers within IU to engage in "open science" principles as producers of their own data. IU-RDCom aims to provide support for both **consuming** and **producing** types of scholarly activity.

Definitions:

Consumed data: *data used by IU researchers to carry our research activities. These data typically originate outside IU, but could also refer to an IU researcher consuming data produced within IU.*

Produced data*: data produced by IU researchers. One of the AAU/ APLU report's goals is "developing campus-specific strategies for making data from federally funded research publicly available."*



We envision people interacting with the IU-RDCom based on different needs. To best guide these people, we propose classifying each user into one of a few distinct personas based on their needs:

- Persona 1: researcher with new need for data that resides at IU
- Persona 2: research/finance administrator
- Persona 3: researcher with research question needing new data acquisition
- Persona 4: researcher who produces data that will be shared as a research output
- Persona 5: researcher whose trainees seek instruction in data techniques and workflow training
- Persona 6: researcher whose trainees seek instruction in data ethics and stewardship

For Consumption of Data

We recommend that IU-RDCom complete **a cataloging exercise —** an inventory of data resources that IU owns, licenses, or to which it could facilitate access — as well as to continually discover new opportunities for what can be gathered. There is a need for creation or procurement of a system in which to carry out this cataloging work and who/what type of staff would be doing the cataloging.

The cataloging exercise should focus on data resources that have most shared value. For example, in the survey, one respondent expressed concern that the specialized data they use are unlikely to be of use to any other researcher. If so, that type of data should have low priority for investments in cataloging. Instead, we should determine which data are of largest potential value through observed use or other means. There is no formula for determining this, but the IU-RDCom implementing these suggestions should consider a structure by which recommendations are made on which resources to prioritize for cataloging. For example, the Kelley School recently purchased a license to a large housing-related national database. By the nature of the topic area, there are likely many researchers who would benefit from that resource, as it could be used to study pollution, community building, material science, social mobility, housing markets, and so on.

A second noteworthy point about the cataloging exercise is that when data are currently housed by a group, those groups may incur high costs in making their data resources widely known. Groups may not have the resources to add other users, so it would not be desirable for such data to be listed in a catalog. Such factors must be understood to **build a culture of cooperation and sharing.** The IU-RDCom would develop strategies to identify which new data resources are of the highest value and to identify individuals/groups who wish to share in costs of data. For data resources already purchased, similar conversations could occur at the license renewal stage, or in thinking of current costs for scaling up use (for example, in creating security training or user guides specific to IU computing infrastructure). The front door should also include an 'announcement' ticker alerting people to funding that is available to purchase or license data.

There may be some data resources that IU researchers use, but not through a physical copy at IU, but through, for example, an agreement to access the resources on some platform outside IU. The same points above on data sharing apply here too—the knowledge shared in the catalogue relevant is about *how* to apply and use. Even if the data is a free resource, the fact that another researcher at IU has navigated the process will lower costs for other researchers to obtain access. The IU-RDCom will be tasked with deciding how such data knowledge resources should be "catalogued." This will occur together with many centers and institutes and other organizing structures that currently steward vast data riches at IU: examples are too many to mention, but include the Regenstrief Institute (information and innovation for better health), and the Polis Center (data on community).

When a new researcher who studies international non-profits joins IU, it may be valuable to know that there are others who have used a database of international non-profits that could be studied, and then to forge connections with those researchers. Thus, one design imperative is to **catalog what are data**

resources with which IU faculty have published research and/ or have familiarity with, rather than simply what data resources IU "owns."

An important aspect of data services is that many groups currently serve as **navigators for the specialized data resources** to which they enable access. For example, Regenstrief Institute is the avenue for access to Indiana's electronic health records, the Kelley School to data on people's use of material resources (such as purchasing of everyday goods in the Nielsen scanner data), and the Observatory on Social Media is the avenue for expertise and access to data from Twitter, Facebook and other social media platforms. It would be extremely useful if IU-RDCom can convene the data navigators at each group to be in contact to coordinate on building user tools, on managing user communities, on the appropriate workflows for data use agreements, on computing needs specific to a data resource, and other common opportunities. These activities also align with a major state initiative , the "Indiana Book of Data and Organizations" which will serve to help connect Indiana's data commons to state talent development and data infrastructure networks.

For Production of Data

Our front door first steps are designed to gather information via various means to identify researchers who share and may wish to share data. Thus, we must understand the researchers and their current data use, sharing, and storage. A useful example of what such services at other universities look like is University of Illinois: https://databank.illinois.edu/.

We will build on the work by the IUPUI working group that was called the "Public Access to Research Data Working Group (PARDWG)". It produced a report on April 2022 that summarizes their interviews with IUPUI researches and stakeholders on their current processes of making data publicly available on a website https://datasharing.iupui.edu/. The IUPUI report organized the campus and system-wide resources by the six stages of the research lifecycle described in the NASEM Report, *Open Science by Design: Realizing a Vision for 21st Century Research*.

When creating a pathway for personas producing research data from their own collection of primary data to make them publicly available, we should use the NASEM report's categories as a starting point for the IU wide version of a "front door" as well.

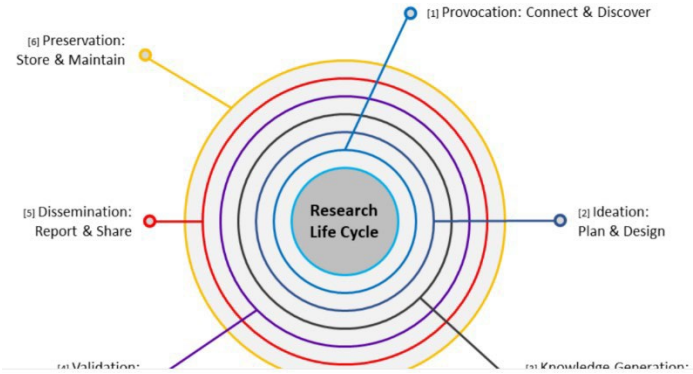
Besides the working group report, we are also learning from practice. For example, in the IU Precision Health Initiative, a large amount of genomic and clinical data have been generated and curated. An important lesson we learned is the need for a multi-level governance structure which involves numerous stakeholders, such as university/school/institute leaders, health systems, legal representatives, IT and security experts, domain/clinical experts, biomedical researchers, and data scientists/informaticians, to properly process and share the data. In addition, a university/school wide policy regarding data sharing is needed for promoting and incentivizing data sharing.

Section 2: Blueprint Proposal (Website)

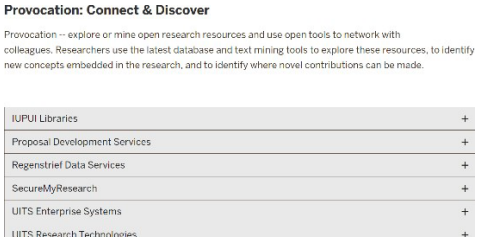
Blueprint proposal for a website that links out to the services and data available, as a precursor to future deeper integrations and development of new services. The blueprint will cover what's available to those seeking existing data as well as those hoping to deploy primary data for secondary use.

This Section follow naturally from Section 1, and is simply the recommendations for the **communications strategy of the IU-RDCom,** visual guidance to a website through which IU community will interact with the landscape, supports and structures created through Section 1. That is, this section proposes what a website would look like, once we have developed in the earlier section the scenarios of how we would guide actual people. We should think of this future website as **an interactive decision tree that guides people** to what they seek, because it is meant to service not just one type of question related to research data. Such a website may also be reflected through videos, brochures, welcome letters to new faculty, presentations to IU community; this website will be linked from many other places people may go to as their front door for that question, and in return, this new website will link back to them.
Thus, Sections 1 and 2 work in tandem.

A very important input to this initiative is the final report that was released by IUPUI's working group on public access to research data. https://datasharing.iupui.edu/
Again, while the June 30th report we are tasked with creating is to cover more than just avenues for researchers collecting primary data to make them publicly available, we may want to use this NASEM report's categories as a starting point for the IU wide version of a "front door" as well. Campus and system-wide resources could be organized by the 6 stages of the research lifecycle described in the NASEM Report "Open Science by Design: Realizing a Vision for 21st Century Research."



This is an example of what the https://datasharing.iupui.edu/ site lists as resources



Section 3: Stakeholder Input Regarding Reduction of Bureaucracy

In this section, we summarize the input we received (via the survey and town hall) regarding ways to reduce bureaucracy and waste, and then we provide recommendations on how to do so. We collected ideas through the survey, the Town Hall, and through conversations with peers.

Summary of Survey Conducted by WG

To gather information from stakeholders regarding research data usage and access at IU, the committee fielded a confidential survey of Indiana University researchers. Between May 17 and June 1, 2022, a total of 85 respondents completed the survey. Eighty-eight percent were faculty members. Most respondents were users of secondary data for research (81%), and of those, over half (52%) accessed these data through IU. 85% accessed secondary research data from sources outside of IU. Of those, 41% paid fees to use data external to IU. Sources of funding included external grants and contracts (68%), internal funding programs (47%), faculty research or start-up accounts (58%), and personal out-of-pocket funds (21%). The survey identified notable barriers and frustrations to accessing or licensing secondary data at IU. Over half of respondents (55%) indicated they had experienced such barriers and frustrations.

The survey also assessed respondents' willingness and ability to share data resources with the IU researcher community. Of those who had used secondary data external to IU, about half (52%) currently held a data license that did not permit sharing with other IU researchers, 17% were permitted to share their secondary data, and 33% were unsure whether sharing is be permitted under their current license. In addition, 81% of respondents surveyed reported using primary data in their research. Of those, only 27% made any data openly available in a public repository and 60% made data available by request and/or with restrictions. About 46% of respondents reported holding primary data that they did not share, but 42% of those indicated that they were willing and able to make their data available to other IU researchers.

Sample quote: *"At IU we have very limited access to large health datasets, especially compared to other universities. We also have limited institutional knowledge (i.e., database management, contracts, purchasing, etc.) to facilitate such research. As such, our research has been greatly hampered, delayed, and stopped because of the limited access to and resources to facilitate secondary data analysis."*

Overall Themes that Emerged

- Hurdles faced in identifying, obtaining, and sharing data for research (and for teaching).
    - Lack of staff who can help researchers tap into existing equipment and data.
    - Waste and duplication of effort (e.g. buying data twice, did not know IU owned already)
    - Lack of services available to enable sharing of sharable data.
    - The need for funds for purchasing newly discovered data resources of wide need for use by many researchers (data licenses).
    - Difficulty identifying the "appropriate" channels to work through for researcher's needs.
    - The need for trainings in accessing data. Democratizing resources requires targeted education on methods of accessing and drawing insights from data, different approaches, and data media across disciplines
- Frustrations with research infrastructure strategy that could be solved through better communications.

- Need for communicating decisions that are made in computing and storage with more shared governance.
- Need for communications to increase understanding of what data sharing means and the principles of external forces towards encouraging open science, to see benefits to own research impact. To establish that "data sharing" requirements (e.g. of journals or of federal funders) do not mean breaking data use agreements or extra bureaucratic steps, but simply making research replication easier.

- All services offered by IU-RDCom is voluntary, no one will be compelled to change their workflow or share data.
- The need to expand currently narrow view among some that might exist of what constitutes "data" for research.

Recommendations Regarding the Reduction of Bureaucracy

1. Problem: How do we reduce wasteful re-purchasing of existing data and how do we reduce missed opportunities of collaborations (of using data potentially available to me through another unit, provided we are able to make it a win-win)?

   Recommendation: We recommend using the ORA contracts and purchasing systems as potential sources of information for consumers of data. We recommend that the IU internal database be searchable for signed data agreements between IU units/individuals and outside data suppliers. The IU-RDCom internal coordinating committee (described in Section 1) should be tasked with reaching out to units to explore to what extent their licenses are able to extend to others and how to incentivize doing so. For example, if someone at IU is purchasing a license for five users of a data set, can a sixth potential user from another unit be added at much lower cost than to forge a new data use agreement by him/herself?

2. Problem: How do we enabling IU researchers to share data and increase impact (in permissible and beneficial ways)?

   Recommendation: Our survey feedback identified many needs in easing the process of sharing data, by producers of data. The University of Illinois data bank (https://databank.illinois.edu/) provides a good example of how peer institutions are helping and encouraging sharing of data by researchers. All actions are voluntary. If someone currently purchasing data does not wish to collaborate or share with any other potential users, nothing in this vision will compel him/her to do so. Of course, in practice, a common question will be how to properly incentivize, reward, and recognize data-sharing while also ensuring that the data generating/acquisition group isgiven priority in accessing such data.

3. Problem: How do we reduce paperwork needed to procure a database?

   Recommendation: We received input from the survey that researchers experience lack of clarity in steps needed to procure a database, i.e. what forms are required, the lengths of time lags necessary, etc. We recommend that the IU-RDCom implement the "front door" process (Sections 1 and 2), which would dictate the processes for the personas described. Through interviews with the people or groups involved (considering the different personas), the IU RDCom would produce descriptions of scenarios and offer instructions regarding how to move forward.

   For example, one of the IURDCom committee member (the data provider) takes on persona 1, who has a specific example of a new data resource they wish to purchase. The data provider then

considers the options available to them. For example, perhaps the persona would like to seek internal funding. The data provider would then describe possible internal mechanisms for raising funds for data, such as partnering with a different school or applying for designated matching fund for research data procurement.

Once funding is identified, the persona may work with the data provider to find out what is needed. The data provider can provide the information needed, such as what data use agreements and forms are needed per the contracts office, what timelines to expect, and what to do when data security plans are needed for the data (SecureMyResearch).

4. Problem: How do we reduce hurdles to the effective use of databases?

   Recommendation: Here too, we recommend the committee implementing the "front door" strategy identified in Section 1, which will describe resources and services that exist at IU currently. We also recommend here (through the budget section) what IU could add to enhance these services. For many, it is not the lack of data available or new equipment needs, but the need for staff/leadership to ensure increased use of available computing and storage equipment and for effective use and sharing of existing data. Section 1 also addresses assessing whether site-wide licenses are available (and cost) when researchers seek to purchase new data so they can easily be scaled to university or campus (i.e., as opposed to seeking an individual license), which requires intervening with researcher prior to purchase/acquisition.

Section 4-A: Sample of Exemplary Initiatives at Other Universities (A)

We conducted in-depth interviews and communications with several universities and with leaders of national reports on related topics.

We are extremely grateful to Jon Dunn, Assistant Dean for Library Technologies, and Heather Coates, Digital Scholarship and Data Management Librarian, for connecting us with research leaders at these institutions. This section is summarized anonymously for the public report.

Institution 1: In 2019, started a tripartite collaboration, between Research, Libraries and IT, with support from the Provost.The collaboration provides researchers with services related to research data: infrastructure to make research data and services available to researchers and the public, to store, manage, curate research data, providing services throughout the life cycle of research and data, and work with researchers find solutions to new research data needs.

Funding: The Provost office funded several positions to enable this tripartite collaboration, including the director of the collaboration. The Provost also funded two research data management specialists who report directly to the director and one position in IT (data storage engineer) who also reports to the director (dashed line).

Governance structure: Joint governance between Research, Libraries, and IT with faculty advisory group.

Initial Stages: The University librarian, the lead administrator for Research, and the CIO worked together to create this initiative. They recognized the researcher's growing needs in data services and sketched out the possibility of positions to be funded to have a coordinated research data services group. There was recognition that the library would be its natural home, given the historic expertise in information management and the library's robust internal IT department, which includes many full-time
software developers. The lead administrator for Research reached out to all department chairs of all the departments to inquire about what they would like to see from such a new service, where were the pain points, and what were the opportunities.

The collaboration, and the library (within which the collaboration is housed) helps coordinate all the separate types of research data needs, including procurement and use of secondary data (which is done by Libraries), as well as the production and publication of primary data, no matter what the size or security complexity.

Faculty advisory roles: Centers play a large role in data procurement and services as well. There are also many individual research teams. All are included in relationships and services coordinated through the coordination unit, of which there are several examples. Data managers across centers are included in the building a community of practice for managers who sit in the actual research teams. There is campus-wide working group to develop research data framework and policies. The collaboration has a formal faculty advisory group that meets twice a year. The Library and IT jointly have a faculty advisory group convened by the CIO and Library leads. Lastly, there is a faculty advisory group for the new campus-wide data management and storage initiative. The faculty are very engaged. There are also periodic surveys and focus groups conducted before major decisions.

These are still ongoing discussions and building steps. For example, data management plans are evolving and also are new to several researchers. The services the unit provides include communicating the idea that there is long-term value of the data for the researchers themselves and for the university. Andthere is an internal active data storage and management system that also includes a long-term internal data repository, with the end goal of having all of the data be descriptively, at the
minimum, tagged so that it can be found and reused by university researchers. Some data has a public path

and some have a private path (internally, to a few or all at the university). Even with secondary data, researchers can publish created sub products of data if their DUA allows. And open science is often about the sharing of code and other details of data creation even if not the data themselves, if the data rests elsewhere and has its own stewardship path.

There is a robust library liaison program ; each liason meets with new faculty and ensures messaging and communication of these services. The liaisons are embedded in departments and have frequent contact with researchers, and can also describe IT solutions available. The Library and the coordination unit are independently and jointly involved.

There is also recently created a secure data infrastructure platform, launched about a year ago.

They are building a Data Commons ecosystem for supporting multiple kinds of research data. Ultimately, there will be a dashboard (a one-stop shop that points to best practices, contacts, policy documents, and personas routes) that people can log into and be directed, but people will always be needed to create the connections that underlie the dashboard and to be the points of contact.
The dashboard can only take you so far, can give you a visual picture of the journey that the human beings will direct you in.

Sustainability: All funded out of Provost office, although individual research groups of course bring in funding for new data purchases.

There has been a lot of buy in and communicating from the top admin levels. The Dean for research, CIO, the provost, and the University Librarian have been willing to invest time and capital.

Institution #2

We reached out to this university as they appearedhave a research data service that cuts across multiple organizational silos (Libraries, IT, and to some degree the Research office).

Starting several years ago, this university envisioned a Research Data Services unit beginning as a grass-roots service with leadership coming from the Libraries, IT and Research.

At present, the team is heavily library-centric although there is participation from within IT and a few related campus services, and Research office engagement is growing: in the longer run, it is anticipated that the all three organizations are actively interested in addressing such needs in a collaborative fashion.

But currently, in actual governance, the services are essentially administered by the Libraries although with volunteers from many other organizations. However, there is no formal funding for the program and no formal agreements. In the longer run, for funding and sustainability, key campus administrators are engaging in the comprehensive research computing and data management framework to bolster/formalize the roles.

Institution #3

--This university is still developing that governance structure. They currently have a voluntary partnership of equal roles (a subcommittee involved members from the faculty, IT and the Libraries, plus a graduate student representative) to provide services in this research data space. This university completed a report to the Provost in 2021 with recommendations from a subcommittee with recommendations that alluded to establishing that governance structure, but there is no action yet;

Provost is working with the group on next steps and expect to hear more in the Fall.
--In terms of other data related initiatives (but not coordinated through a research data commons) this university has established a new central data collection and analysis group that is focused on mining data at the university for insights. This is spearheaded by the Provost's Office and the financial office; is mainly administrative data, not research data, and is similar to what IU's Decision Support initiative (https://dsi.iu.edu/) may potentially be.

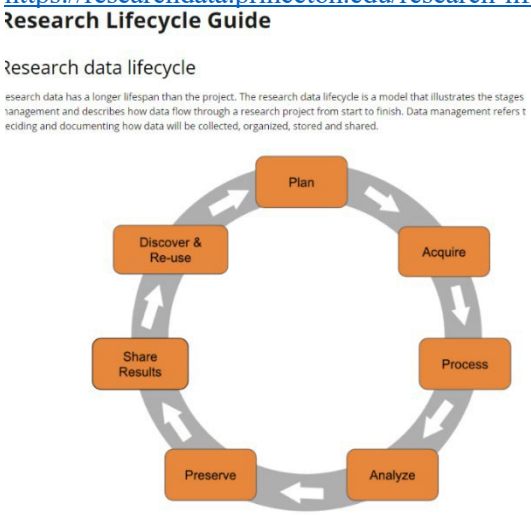
Institution #4

[4] https://provost.utexas.edu/initiatives/sustainable-open-scholarship-working-group/Their focus is on those who create data; researchers using 3rd party data work with their research office in negotiating and licensing data, and are attempting to increase faculty knowledge of these services offered. They are also drafting a university policy that addresses managing and sharing research data, on how best to acknowledge and address research data that comes to the university from an external source.
To date there is no university level governance structure in place but their efforts are evolving. Within libraries they have been actively engaged with the university administration since they started our program and launched it in 2016. Things accelerated when the university sent a group of university administrators, faculty and librarians to the AAU / APLU Public Access to Research Data Summit in 2019. The momentum from that convinced their Provost and V-P for Research to form a working group to better understand the current state of support for data sharing at this university and to make recommendations for improvement. This working group recruited members from across campus, including IT, statistical consulting, sponsored programs, the office of research, general counsel and important institutes. They produced a report in 2019.
However, since COVID hit soon after this report, the momentum stalled. The NIH's announcement that they would require data management and sharing plans for all submissions for funding revived interest in their work.
They are now working to implement some of the recommendations that they made in the 2019 report. The first order of business has been to raise awareness of data sharing requirements and to connect researchers to the many services offered at the university. They released a website for this purpose ; this initiative was introduced by their V-P for Research in a campus wide email. They are also working on developing a more focused policy on research data and developing a series of workshops.
A more durable governance structure has been elusive. The library continues to play a leadership role but they feel it needs to reside at the university level to be sustainable. They are making progress, but the discussions and decisions are still playing out.

Institution #4

This university is actively involved in developing research data infrastructure, some of it aimed at a very granular level of efforts for certain types of data needs (a plan for very sensitive data storage; they have been in touch with other universities' leaders in this area) and some effort are bigger than research data but include it within the envisioning of a new "front door" to all research services at this university. In our conversation, we agreed that communication strategies are important (for example, researchers have been affected by changes in cloud storage contacts), and that it is important to keep in mind the different "personas" and needs of researchers, to ensure a "user centric" approach to this building exercise. We also agreed that an important part of communications strategy is to help people see that a Research Data Commons is a very comprehensive approach, that there's many aspects to research data in which a university should be prepared to be of service for researchers, and that it is important for each persona to see where they fit in within this landscape (which can be visually explained using data lifecycles). That was of people seeing where everything belongs has a "front door" feel. An example to which they

pointed is at another university--
https://researchdata.princeton.edu/research-lifecycle-guide/research-lifecycle-guide



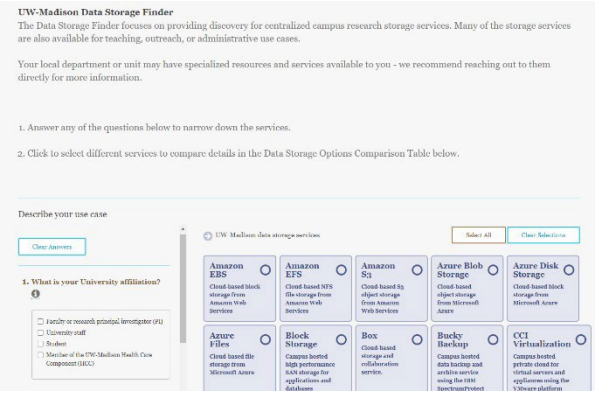
A "front door" website starts with something like this graphic and then allows people to click and take a "self guided your" which is not meant to be the final answer to what they seek (as a website can only go so far, can only be updated only so often) but helps in communication. People can choose personas and see the most common questions for that persona, and what areas of the university data landscape that persona might want to get to know in more depth through contacting an actual person there.
We also agreed that even though a website is hard to build without having the agreements between service units already worked out, that the *aim* of building a website is a useful planning activity, because it forces people from different service areas to be concrete, and that will lead to the agreements and understanding coming into place.

They also pointed to examples at

https://research.yale.edu/resources-researchers
They also pointed to nice example at University of Wisconsin of how researchers can figure out storage solutions. (Note: IU provides these services through SecureMyResearch, but these services are often not well known throughout IU. A communication strategy like Wisconsin's might be helpful:
https://storage.researchdata.wisc.edu/

(IU has

https://cacr.iu.edu/projects/SecureMyResearch/index.html)



We agreed that many universities are at the same stage we are at and devoting energy to learning how a university has fully realize the potential of research data and to be prepared for the fast-moving pace of data's role in society.
We discussed that there is often a misperception that we are building bureaucracy, so communicating that this is about navigating existing structures is key.

Other Notes

Institution #5

This university has a research data services unit that cuts across many organizations. Their governance structure includes Library, IT, . The coordination has allowed the group to coordinate understanding of

resources and services that reside in different units across the university, and in providing guidance to researchers to help with data management plans and encouraging best practices.

Institution #6

This university is about to undertake some major organizational and governance steps for research data management. The 2021 AAU/APLU/ report has been influential among leadership, and they are currently working through recommendations in that report. This university had already established a governance structure within the context of their larger IT governance model that covers both university researcher-created data as well as discovery and reuse of external data. Chairs are from Library and IT. This initiative to centralize and plan for the increasing importance of data in research occurred quite early, in 2009, through a report which was produced by the working group that was tasked by Research leadership with creating a blueprint for an integrated approach to research data sothe university could develop the university research infrastructure
This governance committee is comprised of campus leadership, but is only advisory to the central campus leaders. Within that committee, a research data services committee started as only staff who support research infrastructure, but over the last few years has transitioned to being 50/50 staff and faculty/researchers. This group is advisory, does not have a budget, but makes recommendations to campus leaders. In their next iteration of governance, the expectation is that there will be better integration of the outward facing expertise (of faculty, who are experts in data security, AI etc) broad vision into the current processes. In the text iteration, they hope to have budget and staffing oversight.

Institution #7

This institution we explored is in Canada. The information we gathered showed us that it has well-established structures, but these structures are very much tied to how data and high performance computing support is organized at the national level.

Section 4-B: A Sample of Exemplary Initiatives at National Organizations in the U.S.

Four recent efforts of national entities are noteworthy and relevant.

1. National Academies of Sciences, Engineering, and Medicine (NASEM) 2021. Developing a Toolkit for Fostering Open Science Practices: Proceedings of a Workshop. Washington, DC: The National Academies Press. https://doi.org/10.17226/26308.
2. NASEM, 2022. Open Scholarship Priorities and Next Steps: Proceedings of a Workshop–in Brief, a report from the NASEM Roundtable on Aligning Incentives for Open Science.
3. AAU/APLU series of meetings and reports on public access to data: https://www.aau.edu/accelerating-public-access-research-data
4. National Institutes of Health is going to enforce a new data sharing policy after January 25, 2023. This policy requires "researchers to prospectively plan for how scientific data will be preserved and shared through submission of a Data Management and Sharing (DMS) Plan." The DMS plan will be part of the proposal and will be reviewed.
5. NASEM workshop 2021-2022 on new data infrastructure for federal agencies "*The Scope, Components, and Key Characteristics of a 21st Century Data Infrastructure*" https://www.nationalacademies.org/our-work/toward-a-vision-for-a-new-data-infrastructure-for-federal-statistics-and-social-and-economic-research-in-the-21st-century

NASEM 's Roundtable on "Aligning Incentives for Open Science" convened institutional leaders and others through workshops to identify "barriers and incentives that further open science at universities".

The first meeting was held in February 2019. These meetings "explored the information and resource needs of researchers, research institutions, research funders, professional societies, and other stakeholders interested in fostering open science practices." They resulted in the 2020 publication "Advancing Open Science Practices: Stakeholder Perspectives on Incentives and Disincentives: Proceedings of a Workshop—in Brief".

A university's research data commons strategy connects to the open science movement at a national level. The definition of "open science" varies, but an earlier NASEM report defines that open science "aims to ensure the free availability and usability of scholarly publications, the data that result from scholarly research, and the methodologies, including code or algorithms that were used to generate those data" (NASEM, 2018).

This applies to primary data researchers produce (such as collecting interview data in quantitative or qualitative ways, or images we scan from art works, administrative data we merge with other data), it also relates to secondary data scholars have learned to use in our research, helping others learn from the efforts of scholars on accessing those data. Part of this requires that we train researchers to use reproducibility framework so that others can learn the steps to make the work more transparent. This means a very wide framework of what is needed.

Stakeholders within a university:
Stakeholders around open science within the academic institution are frequently seen as the VPR office, the library, the IT organization, and researchers and educators who have a vested interest in producing, sharing, and using the products of their and other's research.
For example, Keith Yamamoto, co-chair of the Roundtable on Aligning Incentives for Open Science is the vice chancellor for science policy and strategy of the University of California, San Francisco.

Stakeholder groups outside of universities interested in incentivization of open science are broad, and extend to include publishers, funders (federal and others), data repositories, cloud providers, and others. Some quotes from these reports follow: "properly incentivize a more rigorous, transparent, and effective research culture."
"how best to identify and nurture an environment to support, encourage, recognize, and reward open science practices." "development of a toolkit" "advance the movement to make open science the norm." "professional societies in this issue, describing
the example of the Council of Graduate Departments of Psychology, which has been moving the field forward by creating resources for department chairs to encourage open science, such as research training methods and draft tenure and promotion language."

Sanjay Srivastava, professor and undergraduate education chair of the
University of Oregon Department of Psychology, began by noting that the movement toward openness in psychology was initially motivated by concerns about the inability to replicate research; lack of transparency in research methods made it difficult to fully evaluate the evidence.
Examples of national initiatives include:
NIH FAIR (Findability, Accessibility, Interoperability, and Reuse of digital assets) data-sharing policy and the policy on preprints.
Funders have interests here: "Ekemini Riley, managing director of Aligning Science Across Parkinson's (ASAP),"..." Funded researchers must agree to the following terms:• Share new results on the ASAP Hub (virtual platform)• Post experimental protocols to a protocol-sharing service, such as

protocols.io”

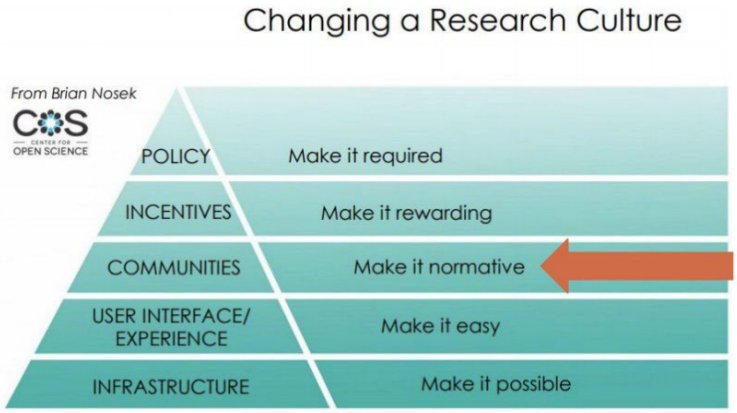


FIGURE 2-2 Changing a research culture.
SOURCE: Julia Stewart Lowndes, presentation, National Academies of Sciences, Engineering, and Medicine, November 5, 2020, based on Brian Nosek, Center for Open Science.

“In addition, Lowndes said, there is a need for salaried positions for open leaders to train and mentor research teams. Trainers, mentors, research software engineers, and community managers who participate in this work are often early-career academics supported by “soft money.””

Five priorities that emerged from discussions
“Ensure researchers have the guidance, training, and resources to fully participate in open science practices.2. Encourage deans, presidents, and provosts to signal that open science is a priority.3. Make the reporting of open activities and the accrual of credit both easy and normative.4. Build upon policies and processes already in place to support open science. 5. Identify and address the financial and human costs of open science”

The end of the document has a list of large projects like Open Genome Project that we could examine further for funders.
Association of American Universities and Association of Public and Land-grant Universities (2021). Guide to Accelerate Public Access to Research Data. Washington, DC. DOI: TKTKTKCC BY- NC-SA 2021 AAU APLU:
“Guide to Accelerate Public Access to Research Data builds on many prior efforts”
“In 2016 AAU and APLU formed a working group to examine issues relating to public access to the results from federally funded research. This working group examined how to improve public access to data resulting from federally funded research. In 2017, the group issued a reportwith a series of recommendations to universities on how to increase public access to research data on their own campuses and how they might work together to advance these efforts. The group also made recommendations concerning how federal agencies could help facilitate sharing of research data at universities.”
“In 2018, APLU and AAU hosted a National Science Foundation (NSF) funded workshop (NSF #1837847) that convened 30 cross-institutional teams with the goal of developing campus-specific strategies for making data from federally funded research publicly available. The two associations issued a report chronicling learning from the workshop. n As part of the NSF-funded (NSF # 1939279) Accelerating Public Access to Research Data Initiative, AAU and APLU reconvened representatives from the university teams at an Acceleration Conference in 2020, to share progress to date, successes and challenges”
This 2021 guide states “The Guide has been informed by 261 campus representatives from 111 institutions, representatives from several federal agencies, and other key stakeholders.”

“resource to help university administrators develop robust support systems to accelerate sharing of research data”

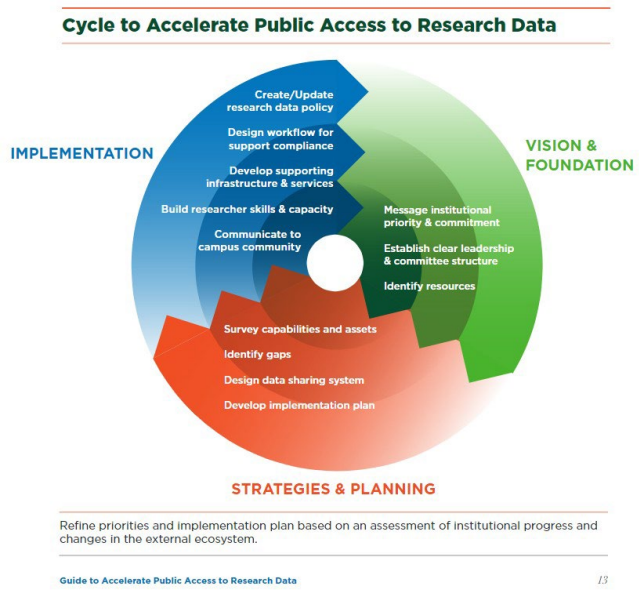
"By ensuring transparency and facilitating the reproducibility of research results, data access is also important to preserving research integrity and maintaining public trust in science."

"Greater coordination among campus stakeholders (including the university provost, senior research officer, and chief information officer as well as general counsels,
compliance, privacy, and security officers, librarians, faculty members, and students)"

"Research data stewardship refers to the activities required to plan, acquire, process, document, and package research data for sharing, as well as the acts of reviewing research data for potential restrictions and making the data publicly accessible."

"Federal agencies and other research sponsors have increasingly developed specific guidance about required behaviors for data sharing during proposal, sponsored project, and post-award phases, and compliance monitoring is expected to follow."

"Institutional administrators can help to send such signals through steps including: Making public pronouncements about the value and importance of public access to research data and articulating this goal as part of university priorities. Convening a cross-campus working group on public access to data consisting of key university officials from the research and sponsored programs office, library, office of the chief information officer, faculty, and others as deemed appropriate to each specific institutional context. Developing a clear and consistent campus-wide research data policy with clearly stated institutional expectations for data management and sharing. Providing existing and new resources to build the institutional support and infrastructure required to enable research faculty to easily make their data publicly accessible. Encouraging colleges, departments, and/or units to discuss how to account for and recognize public access to data in evaluation and assessment of faculty performance and how to make sure that data accessibility is recognized in the university's overall faculty rewards structure, including annual review, promotion, and tenure."
"



They stress that "

administrators at the highest level must clearly articulate the impor-tance to the institution of sharing research data and continually reinforce the institu-tion's priority and commitment to supporting data sharing. Visible and consistent top-level messaging will help ensure the success of researchers in sharing their data and campus efforts to establish a robust data sharing support system."
"Develop a message for the president and provost to articulate and regularly reaffirm that the institution

prioritizes and supports the open sharing of research results, including ensuring public access to research data."

"To guide the campus initiative in
accelerating public access to research data, senior leaders should establish and sup-
port an approach for coordinating planning, implementing, and communicating the activities of the initiative. This includes several elements: establishing who is responsi-ble for managing the initiative, a cross-campus structure or coordinating committeeto develop and manage the initiative, a clear charge for the struc-ture, and resources to support the initiative"

"Especially critical is clearly establishing who will manage ("own") the plan-ning, development, and implementation process. Another import-ant area to consider is the membership of the cross-campus coor-dinating structure to integrate perspectives in developing and implementing the initiative, which m ay take the form of one or more committees (e.g., advisory group and committee on public access to research data, task force with senior leader sponsors). Senior leaders should draw from administrative units that are key to understanding, developing, and/or expanding needed services and infrastruc-ture (e.g., representatives from the research office, research data and library services, information technology, sponsored funding and compliance units, policy and privacy"

"Review the institution's policies related to research data and explore research data policies of other institutions. n Discuss and draft content of the policy, including roles, responsibilities, and expected actions of individuals and organizations across the campus. n Working with the research data workf low developed by the coordinating committee, identify guidelines and procedures necessary to implement the data policy (including how compliance requirements will be monitored). n Seek feedback from campus stakeholders on drafts of policy, guidelines, and procedures. n Define a comprehensive strategy around awareness, compliance, and monitoring regarding the campus research data policy."

"Model Policy Documents: Florida State Universityhttps://regulations.fsu.edu/sites/g/files/upcbnu486/files/policies/research/FSU%20Policy%207A-26.pdfn Rice Universityhttps://policy.rice.edu/308n University of Minnesotahttps://policy.umn.edu/research/researchdatan University of New Hampshirehttps://www.usnh.edu/policy/unh/viii-research-policies/c-unh-policy-ownership-man-agement-and-sharing-research-datan University of North Georgiahttps://policy.ung.edu/policy/research-data-management"

"Utilize the high-level research data stewardship workflow to identify services, tech-nical infrastructure, and expertise needed to support researchers in data sharing. n Evaluate how other institutions have organized their services and infrastructure and what expertise they provide in promoting and implementing effective data stewardship practices. n Develop a multiyear plan to establish or update research data services. n Create a centralized web-based portal for data sharing that will link users to the various campus services and systems, as well as useful external resources. n Evaluate institutional memberships with services that support persistent identifiers, data citation, and other elements of the public access data system (e.g., ORCID, DataCite, Crossref). n Explore the potential to work with external partners to share technical resources and associated costs."

Notes from the AAU/APLU report:
"In October 2017, Iowa State University established a Data Sharing Task Force (DSTF) to develop a framework for implementing services and policies necessary to support research data sharing. The sponsors of the Task Force were the Vice President for Re-search (VPR), University Library Dean, and

Chief Information Officer (CIO). It was co-chaired by the Associate VPR, Associate Dean of the University Library, and Associate
Director of Information Technology Services. Its membership included faculty from diverse fields; directors and/or associate directors from the Office of the CIO, Office of Intellectual Property a nd Technology Transfer, Office of Research Ethics, Office of Sponsored Programs Administration, Office of the VPR; associate dean of the Universi-ty Library; and research counsel from the University Counsel." "Full charge is here: https://www.research.iastate.edu/wp-content/uploads/2021/02/Iowa-State-University-Data-Sharing-Task-Force.pdf"

DOMAIN	PILOT GOALS	TASK FORCE MEMBERS
Policy	Research data, repository	Research counsel, IT systems lead, IP lead, AVPR for research ethics
Compliance	Awareness and prevention, monitoring, noncompliance	Research ethics and sponsored programs staff
Research practice	Understanding requirements, rigor and reproducibility, workflow and documentation	Faculty
Systems and services	Systems (repository, master record, Kuali, workflow)	Library, IT security lead, AVPR for research ethics

"Workflowrepresents the steps that need to be taken in research data stewardship as part of data sharing and compliance with institutional and sponsor requirements"

"Initial Actions: Define the scope of the research data system that needs to be considered in information gathering and planning. Conduct an initial scan of the data sharing landscape, including: an assessment of the institutional, funder, and other organizations' requirements relating to open data/data sharing; an inventory of existing research data policies, procedures, services, systems, and resources in place to meet those requirements; and a gap analysis of expertise, technical services and technology, and financial resources required to meet those requirements. Develop a set of goals and a high-level workflow for research data stewardship to serve as a guide for outlining an approach to developing or enhancing campus infrastructure for supporting public access to research data. Develop a strategic plan with implementation steps, an assessment approach, and an indication of human, technical, and financial resource needs for the project."

"Establish or Update Institutional Research Data Policy and Related Guidelines and ProceduresA campus research data policy establishes expected roles, rights, and responsibilities regarding data management and sharing for researchers and the institution. The pur-pose of the policy often refers to a commitment to data sharing and research integrity as a standard part of the research process, mandates from federal agencies and other organizations, and sometimes a philosophical or principle statement (such as a commitment to transparency and
open science). The body of the policy often includes statements relating to ownership, management, and sharing of research data; retention, archiving, and transfer of data; protecting sensitive infor-mation; and roles and responsibilities of campus community mem-bers (e.g., principal investigators, students, administrative offices). As with any policy, the narrative should be concise and easy to understand and should clearly define terminology. The narrative
should be mindful of the variations in research data stewardship practices across all disciplines and funder mandates and strive to minimize administrative burden on researchers. The development of the policy should also involve consultation with relevant stakeholders, including researchers and adminis-trative units relevant to implementing and supporting the policy."

“A number of resources and templates exist that could provide guidance to campuses developing research data stewardship cost models. Part of the solution to financial support lies in reducing costs. For example, institutions may develop fee-for-service structures for campus research data services that enable researchers to appro-priately budget and charge their sponsored awards for data management and sharing costs.”

“Initial Actions: Develop a multiyear budget for the staff and infrastructure necessary to implement effective research data stewardship practices across the institution. Encourage researchers to include research data curation and storage costs in their research budgets whenever allowed by research sponsors. Develop pilot efforts to better estimate the component costs of effective research data stewardship programs. n Encourage researchers to use domain repositories associated with fields of study or organizations to reduce institutional costs. n Explore potential consortia approaches to research data storage and management as ways to control costs.”

“Approaches may range from workshops that discuss research data stewardship for the full data life cycle to short courses on key software tools (e.g., GitHub, Jupyter notebooks) to writ-ten or video content”

“. In addition, awareness of support and training can be built through a communications program and training outreach that targets researchers of all types, including immediate touch points with new faculty, students, and staff, as well as other regular email or newsletter updates to campus researchers. In addition, external organizations frequently host workshops that could be made known to campus researchers.”

“Initial Actions: Develop consensus of research data stewardship training priorities and goals, and outline a plan with roles and responsibilities to support research data stewardship education and training. Identify specific opportunities to educate and inform faculty/researchers – especially Early Career scholars – of the research data policy and the importance of effective research data stewardship practices. Such opportunities could include the hiring/onboarding process, department-level seminars, workshops, information sessions, and other professional training opportunities. n Identify research data stewardship education and training opportunities for graduate and undergraduate students, such as Responsible Conduct of Research modules. Develop case studies, using campus-based examples, highlighting exemplary research data stewardship practices, including descriptions of the outcomes/benefits of the practices employed.

Cornell University–
Research Data Management Service Group
Comprehensive Data Management
Planning & Consultation Services

The Research Data Management Service Group (RDMSG) is a collaborative, campus-wide organization that links Cornell University faculty, staff, and students with data management services to meet their research needs. The RDMSG's broad range of science, policy, data, and information technology experts provide timely and professional assistance for the creation and implementation of data management plans (DMPs), and help researchers find specialized data management services they require at any stage of the research process.

As a virtual organization, the RDMSG is composed of a management group, a consultant group, and implementation teams. The management group includes a body of decision-making administrators from campus service providers such as the Center for Advanced Computing, the Cornell University Library, Cornell Center for Social Sciences, and central IT, other stakeholders (e.g., chief information officers from the Ithaca and Weill campuses), and a staff coordinator. The RDMSG consultant group consists of science, GIS, social science, digital humanities, metadata, and medical librarians, scholarly communication experts, a senior policy advisor, and other campus service provider staff data experts. Implementation teams conduct assessments, provide outreach and training, and initiate new projects. RDMSG is sponsored by the Cornell Office of the Vice President for Research and Innovation, the University Librarian, and the Cornell IT Chief Information Officer, and is guided by a faculty advisory board.

University of Illinois Urbana-Champaign
Data Nudge: A Monthly Reminder
to Manage Your Data

The Data Nudge is a succinct, topic-focused monthly email to help “nudge” researchers towards better data practices. It is created by the Research Data Service (RDS) in the University Library but taps experts across campus to ensure accurate and high-quality content. For example, while creating Data Nudge content, the RDS staff will contact IT, IRB, or other professionals on campus to vet materials to ensure accuracy and consistency with university policies and best practices.

Begun in 2017, the open rates average 52% per month and has never dipped below 40%, which is more than 2 times the industry standard for higher education. The current 400+ subscribers include individuals from all career stages and domains--faculty, staff, postdocs, graduate students, and undergrads representing 13 of the University's 16 colleges--as well as others from other U of I research institutions and administrative units. The Data Nudge often receives spontaneous positive feedback about its content, clarity, and relevance, even from outside the U of I, from both other universities and even government organizations.

Some topics are U of I-specific (e.g., local storage options) but others are universal, such as data analysis, data cleaning, data destruction, data loss, data sharing, and data visualization. Although the target audience is researchers at the U of I, anyone can sign up, and the content is freely open to anyone to use directly or adapt for their campuses under a CC-BY license. View past Data Nudges and subscribe to the Data Nudge: http://go.illinois.edu/nudge

“Create a research data stewardship communications and marketing plan with materials to support that plan. Develop a narrative to explain how data stewardship services and systems can be leveraged to enhance researcher success and reduce researcher burdens. Identify research data stewardship “champions” to share “success stories” that support institutional data

sharing goals and would have a positive impact on faculty/researchers. Champions could be on-campus or off-campus leaders in specific disciplines, or meet other criteria. Identify or create opportunities to recognize and celebrate exemplary research data stewardship practices newly adopted by faculty and other researchers, such as awards.
Encourage colleges, departments, and/or units to discuss how to account for and recognize public access to data in evaluation and assessment of faculty performance and ensure data accessibility is recognized in the university's overall faculty rewards structure, including annual review, promotion, and tenure"

Notes from the NASEM report for new federal data infrastructure

"The National Academies of Sciences, Engineering, and Medicine will appoint an ad hoc committee to produce three complementary reports on topics that will help guide the development of a vision for a new data infrastructure for federal statistics and social and economic research in the 21st century. The topics the committee will explore include the following:

Report 1: The components and key characteristics of a 21st Century Data Infrastructure including:
The challenges and opportunities related to data infrastructure governance;
The skills, capabilities, techniques, and methods required by the new data infrastructure; and
Issues related to sharing non-traditional data assets, including state and local government, institutional, private sector, and sensor data;
Report 2: The implications of using multiple data sources for major survey programs, including:
Addressing changes in measurement with new data sources;
Approaches for linking alternative data sources to universe frames to assess and enhance representativeness; and
Implications of new data sources for population subgroup coverage, and life course longitudinal data;
Report 3: The technology, tools, and capabilities needed for data sharing, use, and analysis, including:
Alternative approaches and techniques for protecting privacy and confidentiality;
Alternative sustainable organizational models for data sharing; and
Approaches to ensure transparency of the datasets, the use of the data, and the resulting products."

Section 5: Proposal for Short to Medium-Term Budget & Timeline

Timeline: For the first phase of implementing recommendations identified through this report, we suggested the following, based on input gathered on needs from respondents, sharing those with organizations involved (Research, Libraries and IT), and from our working group deliberations. As time would be needed to plan and recruit for new personnel or re-assign duties, we do not anticipate new spending could be successfully deployed immediately, thus we target Year 1 of meaningful budget implementation to be Fall 2023, assuming that the governance structure establishment starts in late Fall 2022.

- We expect Year 2 (Fall 2024-Spring 2025) would be the same as Year 1 (Fall 2023-Spring 2024).
- For the medium term, Year 3-5 (Fall 2025-Spring 2028), we anticipate the needs rising, depending on response and funding success of the initial investments. New funding needs will be identified, as the IU-RDCom builds, that leverage connections to increase grant writing services at IU, and to increases connections with recruiting, with teaching and service missions.

We request resources to help service providers (UITS, Libraries, and Research) to establish their and faculty participation in IU-RDCom. This will serve to enable coordination (e.g. cover the time of individuals from each service organization for coordination activities, and to liaison with the faculty group from UFC). Items requested include FTE for leadership and coordination, and admin staff.

Libraries

Funds for equipment (data licenses and other related, for competitive requests administered by IURDCom, requiring matching unit funds)Funds are also requested in the following categories: research data librarians, data repository managers, software engineers. Funds are also requested for IT leadership and coordination, workflow and services development and research data engineering.
Consideration of budget needs past initial Years 1-2:

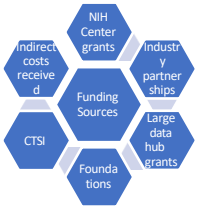
- Actual costs as people start to use more services
  - Storage, Computing, Personnel: Data engineers, data navigators (help create user friendly and sharable internal resources)
  - Adding to data navigators and teams who assist in guiding users-additional resources for existing navigators to scale up to extend to more users and to coordinate across other navigators.
- Administrative lead to organize and coordinate with data navigators
- Activities include:
  - (survey of services units to understand what their barriers are to reducing hurdles e.g. time for review of agreements, or creation of cataloging of data)
  - Assess scenarios of funding for services that will be built
    - Will make it possible for grants to incorporate costs of these activities
    - Funds to increase trainings in workflow at all levels of scholars

- Funds to assist preparing collected data for dissemination and sharing
  - (funding can be based on expected number of users, impact).
  - Money for applications for data purchases
  - Funding for infrastructure needs for technical delivery (options) (Link)
    - Access infrastructure, data storage, archival, meta-data management, etc.

(Timeframe now-5 years to 10 years: Year 0 Fall 2022—in-kind donations from organizations for planning phase, and recruiting for positions to be established by Year 1 (Fall 2023). Year 1-2—initial building of RDComm, Year 3-5 medium term, years 6-10 is longer term.)

Section 6: Possible Sources of Funding

Introduction: Initial investment on research data infrastructure will enable IU to attract *large infrastructure grants* that will help fund its continued evolution. The data services provided by the IU-RDCom will also enable many researchers to pursue new external funding opportunities, through which IU will receive *indirect cost recovery*. In the section below, we provide details on these two types of ways that the IU-RDCom central investments would be recouped and grow with new external funding.



Infrastructure Funding Opportunities enabled by IU-RDCom
The IU-RDCom investment comes at a key time for IU researchers' ability to comply with new rules starting January 25th 2023 which require NIH applications to contain a detailed data sharing plan.[5] IU will also be prepared to meet similar requirements from other federal research funders as efforts accelerate to align requirements with open science principles. Along with these increased requirements come opportunities for funding infrastructure building. Establishing a robust infrastructure showing evidence of integrated groups at IU, and evidence of IU's willingness to contribute, would position us to respond to any such solicitation from the NSF, NIH or other sources. We anticipate, for example, that NIH will soon issue calls for funding infrastructure building at institutions whose research data ecosystems cannot currently support requirements for its new Data Sharing requirement. An IU-RDCom also allows us to be on par with other institutions creating similar models to pursue joint grants from IMLS, NSF, etc. who wish to fund many institutions collectively. As one example, we have strengths through CTSI to engage in such activities around research data infrastructure to improve health and wellbeing in Indiana.
Other examples:

- Lilly Endowment's "Charting the Future for Indiana's Colleges and Universities" which has approved more than 138 million for initiatives including "better use of data to improve student outcomes": https://lillyendowment.org/our-work/education/higher-education/charting-the-future-for-indianas-colleges-and-universities/

- The National Leadership Grants for Libraries Program (NLG-L) which offers implementation grants of up to $1Million requiring 1-1 cost-sharing from an institution's own funding. https://www.imls.gov/grants/available/national-leadership-grants-libraries. Next proposal deadlines anticipated in future years.

IU-RDCom can encourage and provide supports to IU researchers to seek more center grants (e.g. NIH "P" grants) where a "data core" budget would help bring in resources. These data cores can support the purchase of costly data, and could also fund staff positions for specialized data managers and analysts. A closely related type of funding model NIH offers is the "data coordinating centers". Specifically, the U24 mechanisms are often used for such Data Management and Coordinating Centers. Most of these RFAs are associated with a specific consortium, clinical trial, or disease. However, some of them are more general. For example, a specific Biomedical Data Depository grant is being solicited by multiple NIH institutes (https://grants.nih.gov/grants/guide/pa-files/PAR-20-089.html). In addition, there are NIH grant mechanisms for developing data sharing technologies, often under the informatics or Big Data programs such as the NCI Informatics Technology for Cancer Research (ITCR) program. The Department of Biostatistics and Health Data Science in SOM/FSPH is currently establishing an initiative for pursuing more such grants.

Research Funding Opportunities enabled by IU-RDCom

There is direct funding benefit from the data services provided. The IU-RDCom will enable researchers to pursue new grant funding which they earlier could not pursue due to a lack of research data infrastructure. As one example, a participant we spoke with described the recent NSF "Bridge2AI" funding opportunity they had not been able to pursue, but would have mounted an effort, if services we recommend in this report had existed.
We expect researchers could also tap into vast ongoing NIH opportunities for direct support, as well as other sources, so that the associated indirect receipt by IU would far surpass the investment in IU-RDCom. Having broader access will allow for more research groups that currently pursue federal funding to consider more areas of interest and be more successful in pursuing grants. There is also associated benefits in research compliance of funding. Having access to institutional resources (including knowledgeable staff that can point researchers to either an IU or national resource) will allow researchers to focus on novel research ideas and worry less about compliance with grant requirements. It can significantly lower the friction in the proposal process and enable the University to more efficiently capture grant dollars.

Additional Resources:
Another resource for us in thinking of possible external funders is to consult curated lists of funders for large data initiatives from places that are collecting these details. For example, https://midwestbigdatahub.org/ (which IU is part of, through the Indiana University Network Institute (IUNI)) collects this type of information, so we will be well served to use the list of the grant opportunities they have developed: https://midwestbigdatahub.org/get-involved/funding-opportunities/.
Many funding opportunities are specific to single areas of research.
Example: the BRAIN Initiative is an R24 opportunity at NIH to create a data archive. " A goal of this program is to advance research by creating a community resource data archive with appropriate standards and summary information that is broadly available and accessible to the research community for furthering research."[i]
There are also models of industry partnerships for creating university research data consortia.
In an example at University of Houston, a faculty member who created a partnership with energy companies is quoted as saying, "Data from industry are generally superior to anything that academia can collect because the industry has the resources and infrastructure to develop and support the highest level of subsurface imaging of the deep sedimentary basins ..", and a student in the consortium comments that "Students can work directly with critical industry through our interpretations and analysis which benefits the company that provided it."
Another example of a consortium is the Linguistics Data Consortium https://www.ldc.upenn.edu/about . This consortium is created by universities alone.
At Indiana, the AnalytixIN initiative is in the process of establishing a data lake for multiple types of data covering biomedicine, agriculture, and manufacturing. For biomedicine, a cloud-based system is being developed using Google Terra platform. The AnalytixIN platform could be used as enabling infrastructure for competing for external funding.

Summary: Our cursory investigation yielded many examples of potential sources of infrastructure grants. We will be well positioned to pursue there, especially ones that require matching dollars, which other institutions may not be in a position to provide if they have not undertaken these initial steps. We also identify ways of funding to IU in the form of indirect recovery from new research this initiative will enable. We recommend a larger effort to discover the external funders, for example through working with IU Foundations relations, in parallel with the building of the IU-RDCom.
Page Break

Appendix 1: Charge Email to Working Group

Prior to the Charge meeting on April 28th, 2022, IU Vice President for Research Fred Cate wrote:

Dear XX,

I am writing to ask you to serve as a member of the Research Data Commons Working **Group** to be chaired by Kosali Simon, Distinguished Professor, Herman B Wells Endowed Professor, and Paul O'Neill Chair in the O'Neill School of Public Health and Environmental Affairs.

The broad mission of the Research Data Commons is to serve as a university-wide resource for discovering, sharing, and accessing data resources across the IU community; build on our world-class strengths in centralized cyberinfrastructure and other areas to present researchers with more integrated and easy-to-access resources; enable richer training opportunities for students and faculty; and empower IU to better serve local organizations, our state, and other partners.

It is my hope that this Working Group will create the vision for the Research Data Commons and the practical steps for achieving it in the near-, medium-, and long-term. Your work should include:

1. Soliciting input from faculty, students, and other constituencies;
2. Identifying useful models and best practices from other universities and other institutions;
3. Developing a blueprint and timeline for implementing a Research Data Commons; and
4. Addressing key issues such as scope, access, governance, funding, staffing, outreach, training, maintenance, and anything else you consider important to the success of this initiative.

You may assume that there will be some new funding to support the Research Data Commons, but I encourage you to focus less on a wish list of what could be done with new money and more on how to take advantage of existing investments and grant funding, as well as external funding sources such as NSF/NIH data hub grants, foundations, corporations, and philanthropic donors. It is important that whatever you propose be financially sustainable in the long run.

I know it is ambitious—frankly, everything about this is ambitious—but I would be grateful if you could produce at least a preliminary report by June 30, 2022. Fortunately, Kosali and her colleagues have already done considerable preliminary work. My office will provide whatever staffing or other support the Working Group needs.

I hope very much you will agree to participate and I deeply appreciate the expertise, insight, and commitment I know you will bring to this important task. Please do not hesitate to let Kosali or me know if you have any questions.

Sincerely,
Fred H. Cate
Vice President for Research
Distinguished Professor and C. Ben Dutton Professor of Law

Page Break

Appendix 2: IU Impact Email Blast

link https://research.impact.iu.edu/more-iu-research/stories/research-data-commons-working-group.html
May 16th 2022

research.impact.iu.edu/more-iu-research/stories/research-data-commons-working-group.html

Research Impact

ABOUT | KEY AREAS | MORE IU RESEARCH | FOR FACULTY | STUDENT RESEARCH | IU RESEARCH

Home / More IU Research / Stories

IU's Research Data Commons Working Group

Monday, May 16, 2022

A group from across the university has been tasked with providing recommendations for establishing an Indiana University Research Data Commons (IU-RDC). This objective embodies a collective recognition of the importance of a university-wide resource for discovering, sharing, and accessing

IU's Research Data Commons Working Group
Monday, May 16, 2022

A group from across the university has been tasked with providing recommendations for establishing an Indiana University Research Data Commons (IU-RDC). This objective embodies a collective recognition of the importance of a university-wide resource for discovering, sharing, and accessing data resources across the IU community. Such an initiative would build on our world-class strengths in centralized cyberinfrastructure and other areas to present researchers with more integrated and easy-to-access resources; enable richer training opportunities for students and faculty; and empower IU to better serve local organizations, our state, and other partners. This is an ambitious goal, but a lot of work has been undertaken by working groups at individual campuses and units already.
As a first step, this working group is tasked with collecting input from faculty, students and others on a vision for the Research Data Commons and the practical steps for achieving it in the near-, medium- and long-term. The working group will be soliciting your feedback through an open town hall webinar on May 23 at 1p.m. Please register to attend. {link was provided to https://iu.zoom.us/webinar/register/WN_l758-hoSSZuLSrSl179rNg } We will also administer a brief survey {link was provided to https://iu.co1.qualtrics.com/jfe/form/SV_0v91F85i6bJOjqK} to learn how a more coordinated approach to research data could help researchers in their day-to-day activities (for example, what processes could be simplified). The group is also tasked with identifying useful models and best practices from other universities, and possible funding sources through, for example, NSF data hub grants, so that IU's activities in this area will be financially sustainable in the long run. The working group also appreciates feedback on scope, access, governance, funding, staffing, outreach, training, maintenance, and anything else we consider important towards a research data commons. From feedback received, the working group will develop a preliminary report and submit to VPR by June 30, 2022.

The group consists of 12 members; please feel free to reach out to us individually at any time, and/or provide input through the survey or townhall meeting.

Appendix 3: Medium-Term Strategies for Scaffolding a Successful Data Commons

This section contains a non-exhaustive set of points we encourage the IU-RDCom considers for medium-term strategies and scaffolding for a data commons. Possible key components examined include outreach, policy, governance, training, communication, infrastructure, purchasing, and provenance/replicability. Once an ideal framework is envisioned, we recognize the need to evaluate where IU excels and where there are gaps. Included in this evaluation would be an examination of the tensions between centralization and customization, the importance of trust and outreach, the discussion of potential solutions to governance questions (e.g., deciding on new data purchases), the need for a rubric for evaluating new services provided, and how growth and sustainability of the efforts will be supported.

Communications Strategies

The importance of research data topics will be communicated. For example, communication is needed to help researchers see that the IU-RDCom activities will relate to *consumed* and *produced* data. We note that the scaffolding needed for accessing and managing research data *produced* from IU research has unique features relative to the scaffolding needed for accessing and managing data *consumed* as a part of IU research.

Communication will also be needed to establish that IU-RDCom is not creating a new organization; it is a governance structure, a network of existing entities. Specifically, it does not aim to tear apart any existing structures or to take away the data governance rules that currently exists within groups, but simply to provide a structure and "umbrella" services and supports. The IU-RDCom itself will not obligate a researcher to share data that they produce. Rules governing the sharing of data originate from elsewhere, such as funders, journal policy, and data use agreements. The IU-RDCom structure simply enables researchers to have easier ways to securely extend the reach of their research by sharing data, as encouraged by the "open science" movement. It provides an important research resource that rivals that of our peer institutions.

Using the results of Sections 1& 2, IU-RDCom will identify existing units/groups that procure data at IU, explore the opportunity for potential collaborative data sharing hubs. That is, the IU-RDCom will, to the extent possible, network between separately managed collaborative data sharing hubs where the organizational entity can be a recognized academic unit or a project. In the medium term, IU-RDComm requires 3 main focuses: 1) TECHNOLOGY- this is the infrastructure portion. 3)COMMUNITY BUILDING for COLLABORATIVE Data sharing, and 3) POLICY review/updating/creation.

Background information for 1) Technology relates to principles from
https://support.dataverse.harvard.edu/harvard-dataverse-general-terms-use

Information for 2) COMMUNITY[5] Possible design principles are:

Clearly-Defined Boundaries- Individuals who have rights to appropriate resources must be clearly defined, as must the boundaries of the resource itself.

Researchers' data commons is bounded by a well-defined purpose, a set of values to prioritize, a well-scoped mission. Together, these determine who can contribute, access, and use the data resource or make decisions about it. It also helps determine the shape and context of the data resource itself.

Appropriate Rules- Rules are appropriately related to local conditions (including both regarding the appropriation of common resources — restricting time, place, technology, quantity, etc.; and rules related to provision of resources — requiring labor, materials, money, etc.)

- The various resources the commons stewards, such as data, people's time, funding, as well as the organization itself, have appropriate rules to describe how they can be used and under what

[5] https://grants.nih.gov/grants/guide/rfa-files/RFA-MH-20-600.html

conditions. In general, the rules should ensure that those who contribute resources benefit from their contribution and that harms from the use of resources are curtailed.

Rule-making processes

- Collective-choice arrangements allow most resource appropriators to participate in the decision-making process. In short: those who are affected by decisions and rules that govern the resource or the community itself should have a way of influencing those decisions.

Monitoring- Effective monitoring by monitors who are part of, or accountable to, the appropriators.

- This means that compliance with the rules established is monitored and that users of the commons have an active role in monitoring compliance. With regard to data commons, this includes monitoring of data production processes — ongoing validation of data integrity, verification of data quality, — as well as monitoring data access and use.

Sanctions- There is a scale of graduated sanctions for resource appropriators who violate community rules.

- This principle refers to the set of accountability measures that should be in place to guarantee rules are enforced. However, the focus on *graduated* sanctions implies that not every violation of a rule is treated the same and, for instance, intent and harm are taken into account when applying sanctions.

Conflict resolution mechanisms- Appropriators and their officials have rapid access to low-cost local arenas to resolve conflicts among appropriators or between appropriators and officials.

- When conflict arises in a data commons, there needs to be an effective, inexpensive, and otherwise accessible way to handle that conflict. In addition, a data commons needs to decide and make clear which conflicts will be handled internally and which ones should be resolved externally, for instance by going to court.

Right to self-governance- The rights of a community to devise and govern its own institutions is recognized by external authorities.

- In the context of data commons, this principle encourages us to understand how far the decisions we make about the collection and use of data are in line with, for instance, data protection regulations.

Nestedness/Interoperability- Appropriation, provision, monitoring, enforcement, conflict resolution, and governance activities are organized in multiple layers of nested enterprises.

- In relation to data commons, this principle can refer to a possible need for one data commons to interoperate with another, or to break-up one large commons into smaller, nested commons that interoperate with one another. Doing so would allow each smaller commons to make decisions that better reflect their circumstance and match a narrowly defined purpose.

Key to creating an IU-RDCom environment is the commitment to flexibility for the researchers to determine how data sharing is appropriate for their desired purpose and task. As such, the IU-RDCom has an overarching structure- based in technology- that serves as a front door to access. (The "front door" is described further in Sections 1 and 2.) Of course, assisting researchers in becoming more focused on data sharing is no easy task, especially since every discipline and issue often has differing cultures and necessary tasks. Thus, IU-RDCom is designed with an eye toward a 'hub structure. Researchers will use a front door to access individually controlled 'Community Research Hubs.' As such, there will be many Hubs- with different foci and set of rules for each.

At its most basic, a 'Community' is a group of researchers who have self-organized and seek support to prioritize data sharing. There can be- and in fact should be- many communities- that each elect to organize around a purpose(s), data set (s) and/or data topic. This will allow researchers to build a sharing environment that reflects the needs of the community. For example, imagine a group of researchers interested in climate change. The researchers may agree to share the data they have purchased, but of course- the sharing is limited based on the licenses under which they obtained the data. In this situation, the Community may decide to have a very limited group of community members- who all agree to share

the data *only* within those proscribed limits. In contrast, a group of researchers may want to share–in an open environment particular data sets. This community can design their sharing environment in exactly that way. Researchers make a commitment to use the research space based on the overarching Rules of the Research Data Commons; in exchange IU provides supports and allows those currently owning/managing hubs to establish the rules of their community.

Ongoing steps should include

- Immediate Outreach: after inventory/survey has identified some potential participants- work to assist them in understanding the value of this activity
- Longer term outreach: monitor the procurement process to find, encourage and connect faculty who may not know of the opportunity the IU-RDCom presents.
- Ongoing publicity: as some of these communities move into later stages- and/or succeed in their goals
- Educate: potentially build into various portions of IU curriculum this mechanism, support for technology, data management study, etc. Workforce development—training students to enter the workforce being skilled at data related aspects of society's work, is an especially important aspect of this work.

Adjustments, amendments and other alterations shall be decided by newly established governance structure.

For shared data storage: If the space is not being used (as determined by the governance structure)- notice will be given – assistance will be given in migrating any data out of the space– and the space will be reapportioned. The space and the project are managed BY the community members- with limited exceptions, such as critical security. This exception does NOT include- cost, staffing, migration demands, decisions about 'active use,' - the community explicitly governs these decisions. Access to communication- via IU newsletters, etc.- to advertise the Hubs.

Users should consult with IU experts on data storage (IU's SecuremyResearch) rather than decide on their own what storage is used. There are many reasons that these decisions should not be individual—especially to protect us collectively. OpenICPSR is a service that should be explored. Low barriers to hosting a data collection, a simpler version than Scholarworks and for data instead of publications. This could be implemented through IU hosting instance of Harvard's Dataverse. But an IU group could use Harvard's dataverse to store their data collection, and dataverse has "knobs" that allow one to control access, so a community could choose to host and manage their collection using the dataverse platform. Whether through IU or elsewhere, a conversation with SecureMyResearch would be advisable.