

Research Data Service Annual Report

2022-2023

The Research Data Service (RDS) is a campus-wide program that provides the Illinois research community with the expertise, tools, and infrastructure necessary to manage and steward research data.

2022-2023 Goals

- Diversify outreach and communication strategies to provide alternative ways for the campus community to learn about data management best practices outside of traditional workshops and webinars.
- Successfully prepare for NIH's new Data Management and Sharing policy by providing proactive announcements, presentations, guidance, and plan reviews.
- Implement improvements for the Illinois Data Bank, including refining the versioning process, creating a dashboard of static and on-demand auditing reports to better monitor the service, and transitioning the issues workflow from Jira to GitHub.
- Actively engage with and support the Data Curation Network as new members join and the organization continues to mature.
- Initiate reviews of RDS operations and services and of Illinois Data Bank dataset deposits.

Major Activities and Accomplishments

FY23 marked a successful year as we fully accomplished or made substantial progress on all the goals set for the year and were fully staffed. Major activities and accomplishments included:

- **Improved and extended communications** through developing a comprehensive communications plan, designing a sticker promoting file backups using a squirrel aesthetic that plays off the popularity of campus critters, and publishing a self-guided workshop for researchers to learn about data management at their own pace.
- **Launched a broad awareness campaign** for NIH's new data management and sharing policy which included a talk tour, direct email contact with all current NIH grant holders, and customized communications materials.
- **Maintained and improved the Illinois Data Bank** by transitioning from Jira to Github for submission and tracking of issues, researching and designing interface and workflow improvements, and auditing all deposited datasets to assess file transfer success and prioritize future development.

- **Supported the Data Curation Network** through participation in several working groups, committees, research projects, and events.
- **Completed** a review of older datasets in the Illinois Data Bank and the first stage of an extended review of RDS operations, which entailed combining and unifying metrics collected in the last 5 years.
- Additionally, we continued to **strengthen collaborations** through our on-going collaboration with ACES to prepare for the 150th anniversary of the Morrow Plot in 2026, by participating in the planning committee to prepare for the upcoming 19th International Conference on Digital Preservation (iPres) to be hosted by the University Library in September 2023, and joining the OVCRI's new Research Integrity and Compliance Excellence (RICE) group.

Outreach and Engagement

With NIH's new data management and sharing policy and the addition of Sandi Caldron's communications finesse, FY23 saw many activities and improvements for RDS outreach and engagement.

Standard activities such as consultations, dataset curation, data management plan reviews, and outreach and training events like presentations, workshops, and guest lectures were strong. As in FY22, the RDS provided a mix of in-person, virtual, and hybrid opportunities for engagement that combined served 1,511 researchers (Figure 1). Overall, activities increased 35 percent from 179 in FY22 to 234 in FY23 with an especially significant increase in Data Management Plan (DMP) consultations. The NIH instituted a new Data Management and Sharing Policy that, for the first time, required a DMP with most grant applications. Consequently, DMP consultations jumped from 9 in FY22 to 30 in FY23 (Figure 2). (While the NIH uses the term Data Management and Sharing Plan, for simplicity's sake, we refer to all such plans as DMPs in this report.) We saw a similar increase in consultations in response to NSF's DMP requirement several years ago. Similarly, we expect this number slowly drop in the next few years as researchers become more familiar with the NIH's policy and make use of the resources we provide, such as sample DMP plans. DMP reviews were just one part of a coordinated outreach campaign to ensure our researchers and other research support units were well prepared for NIH's new policy. We also added an information page to our website, emailed researchers currently funded by NIH, provided communication materials to departments impacted by the change, and led several presentations in the months leading up to the policy launch in January 2023.

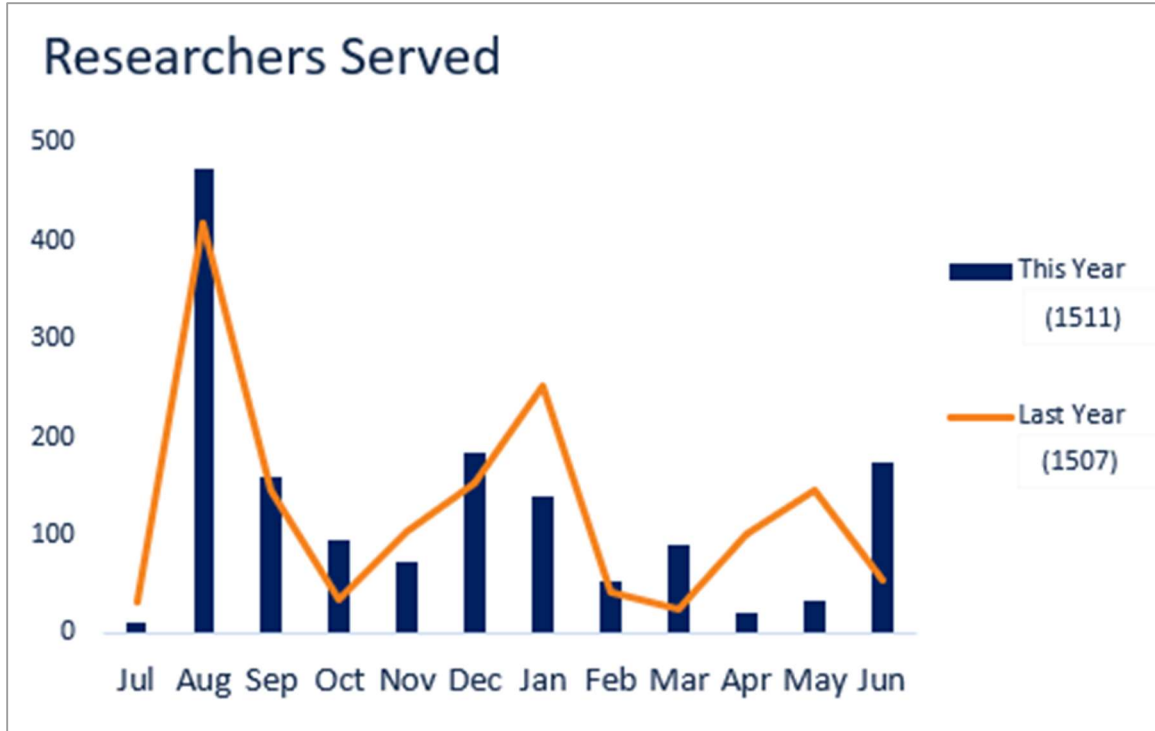


Figure 1. Researchers served remained stable compared to last year with a spike in August when the RDS participates in the orientation for international graduate students.

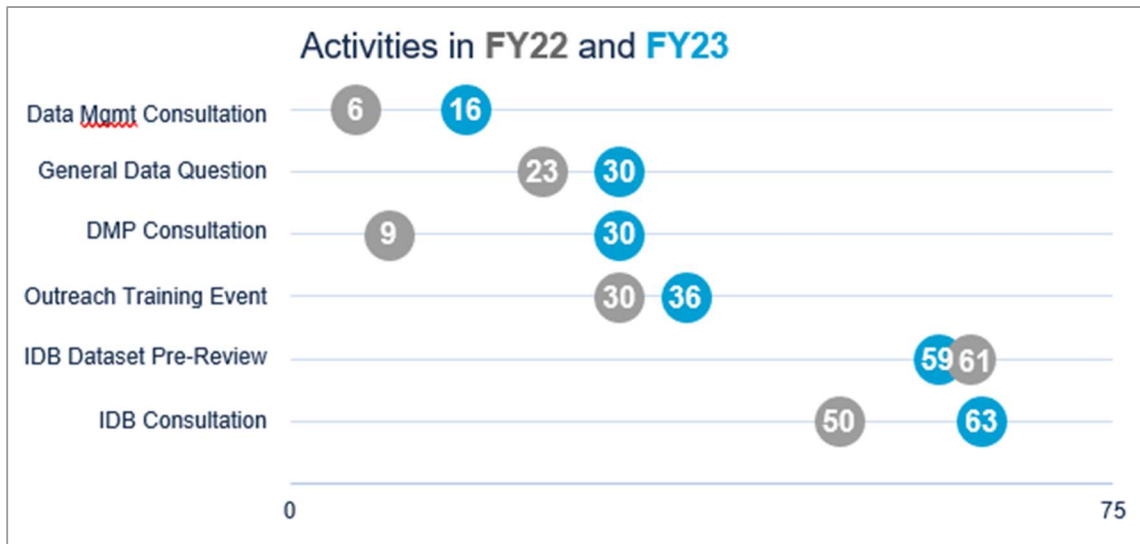


Figure 2. In FY23, the RDS saw an increase in DMP consultations, as expected given NIH's new Data Management and Sharing Policy. All other activities also increased except for pre-publication reviews, which decreased slightly.

In addition to our NIH outreach campaign, we also continued to provide data management outreach and sustained relationships with many units on campus that we had worked with before.

RDS led workshops or presentations for many campus units:

- Beckman Institute
- Bioengineering
- Chemistry
- Environmental Engineering
- Institute for Genomic Biology
- Interdisciplinary Health Sciences Institute
- Library
 - Scholarly Commons
- Molecular and Cellular Biology
- Natural Resources and Environmental Sciences
- OVCR Administration
 - Campus Research Administrators working Group (CRAWG)
 - Office of Technology Management (OTM)
 - Sponsored Programs Administration (SPA)
- School of Integrative Biology
- Technology Services
 - IT Pro Forum
 - Senate IT Committee

RDS provided 6 guest lectures:

- Data Management, Curation, & Reproducibility (IS 477) – twice
- Professional Development for Psychology (PSYC 500) – three times
- Scholarly Communications (IS 511A)

Outside of campus, the RDS strengthened external connections with two presentations to the Data Curation Network about the Illinois Data Bank and as participants in a panel discussion on historical data for the Research Data Alliance as part of their Decade of Data webinar series. In collaboration with the iSchool and the Grainger Engineering Library, RDS staff also hosted several working meetings with scholars from Kyushu University and Chiba University in Japan. This collaboration resulted in a research project focused on data sharing and reuse among international collaborators. Additional reports about this project will follow in FY24.

Online engagement opportunities continued to perform well. Our broadest outreach effort, the Data Nudge monthly email hit nearly 650 subscribers at the time of writing this report, and the open rate remains stellar at about double the industry standard for education. As in FY22, deposit related activities (e.g., pre-publication review and consultation) contributed to more than half of the RDS activities performed in FY23 (52%) (Figure 3).

Finally, RDS staff have been working with College of Agricultural, Consumer and Environmental Sciences (ACES) to prepare for the 150th anniversary of the Morrow Plots crop rotation experiment. The Morrow Plots are the second-oldest ongoing agricultural experiment in the world. However, no longitudinal dataset has been published until now. In December 2022, the Morrow Plots Data Curation Working Group – led by ACES and consisting of members from the RDS, the Funk ACES Library, the University Archives, ACES IT, and Crop Sciences – published data from 1888-2021 in the Illinois Data Bank. As FY23 came to a close, the data had been downloaded almost 200 times and received media attention in many outlets. This dataset helps fulfill the University’s land grant mission by sharing our unique contributions to agricultural research with the world and also provides an example of well curated and thoroughly documented dataset. The Working Group published a conference paper and presented on the dataset at the 18th International Conference on Digital Preservation (iPres) in FY23 and are scheduled to present a follow up in FY24.

Illinois Data Bank

The Illinois Data Bank continues to provide Illinois researchers with a free, stable, and convenient place to share finished research data with the world. As in past years, Illinois Data Bank activities accounted for half of RDS activities overall (Figure 3). Dataset deposits continued to grow in FY23 (Figure 4), with more deposits in FY23 (111) than FY22 (97). As of June 2023, the Illinois Data Bank had published 632 datasets, including versions, and total downloads rose to 342,304. The top ten depositing departments last year continued to deposit more datasets, representing six different colleges and research institutes (Figure 5).

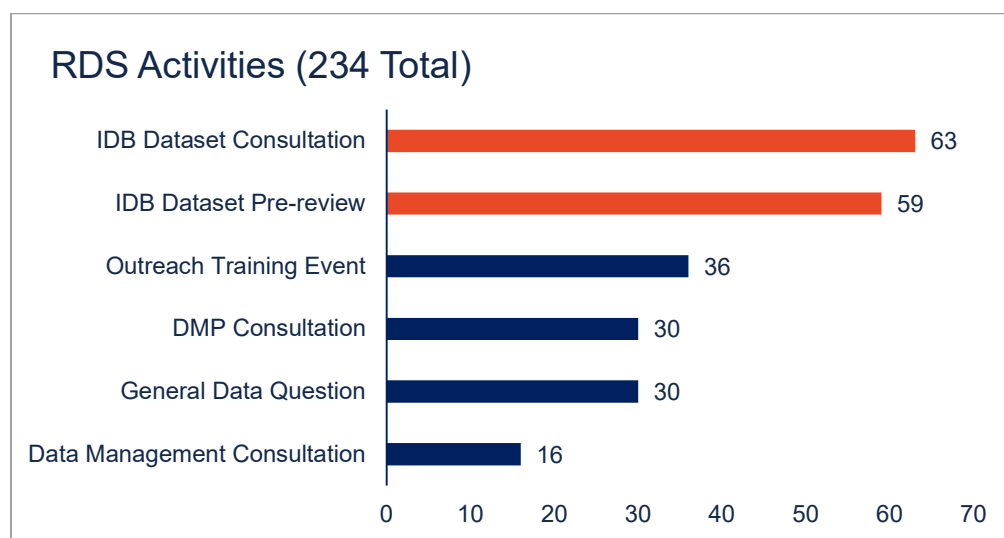


Figure 3. A breakdown of activities by type shows that virtual work related to the Illinois Data Bank (pre-publication review and consultation), as previous year, dominated RDS service offerings in FY23.

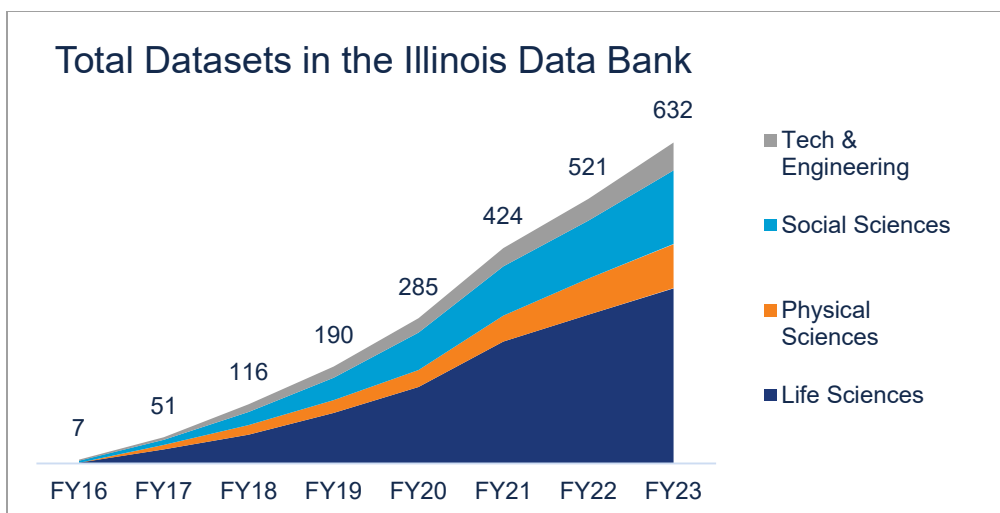


Figure 4. In FY23, the total number of datasets in the Illinois Data Bank reached 632, including versions.

In the fall FY23, we began to develop a new versioning workflow for the Illinois Data Bank. Per our policy, changes may only be made to published data files by creating a new version of the dataset. Files may need to be changed as a result of the peer review process for an accompanying article, to fix issues found during a post-publication curation review, or simply to publish updates. The process is currently completed manually by having depositors create a new submission, copy all of the metadata from the previous version over, and then upload new and old files. The RDS staff provides substantial oversight and assistance throughout the entire process. Despite our best efforts, the versioning process is confusing for users, and mistakes happen, especially for with complex datasets that have multiple authors and many files. In response to users' requests for improvements, RDS staff began a redesign process and started by interviewing representatives from other institutions within the Data Curation Network (DCN) to learn more about their processes and experiences with versioning. RDS staff also reached out to depositors who expressed an interest in a more robust versioning process to obtain their feedback. In April, feedback and suggestions were compiled and presented to the RDS team, which met weekly to design a new workflow. At the time of the writing of this report, the proposal is completed, and our Programmer is developing a prototype. We expect to test and implement the new versioning process in FY24.

After a long delay, we also received the first-round review of our CoreTrustSeal application to certify the Illinois Data Bank as a trustworthy data repository. A protracted process is consistent with the experience of our peers, and we expect that the revisions and subsequent reviews may take us well into FY24, if not beyond, to resolve. We will continue to report progress in subsequent annual reports.

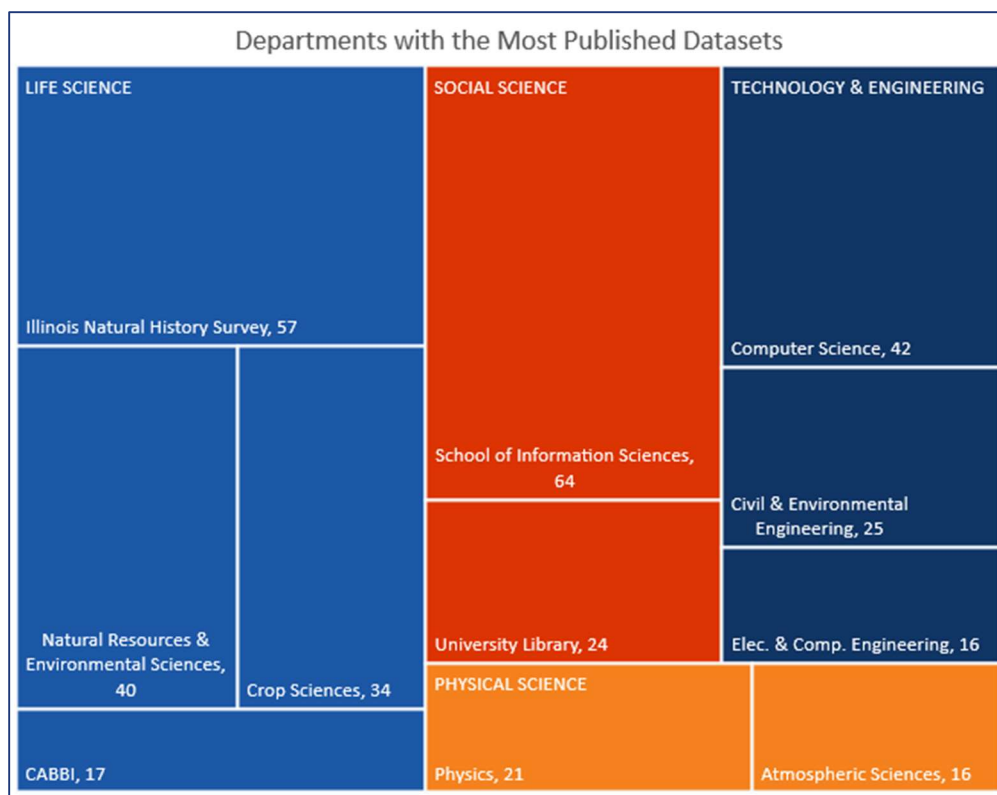


Figure 5. As in the previous year, the most prolific departments are largely STEM-focused and represent Liberal Arts and Sciences, the Grainger College of Engineering, Agricultural Consumer and Environmental Sciences, the iSchool, the University Library, and the Prairie Research Institute.

Data Curation Network

Illinois continues as one of over a dozen sustaining members (Figure 6) in the DCN. In FY23, RDS staff continued to meet the goals for Illinois’ participation in the DCN by attending meetings, curating datasets, and participating in several DCN Committees and projects. Sandi served on the Finance Committee and as a Curator at Large on the Governance Board. Heidi continued to serve as our institutional representative and on the membership committee, and Hoa served on the All-Hands Meeting (AHM) 2023 planning committee. The AHM 2023 was held in-person and virtually at Princeton University, providing opportunities for networking and project development. As the result of the AHM, Sandi and Hoa will lead the Institutional Outreach and Communication interest group for the upcoming year. The goal of this interest group is to keep everyone in the DCN informed of outreach and communication efforts by member institutions and to share resources, such as marketing language and flyer templates.



Figure 6. DCN sustaining institutions in FY23.

In FY23, the RDS staff helped the DCN produce resources and publications, such as *The Art, Science, and Magic of the Data Curation Network: A Retrospective on Cross-Institutional Collaboration* published in March 2023 by the University of Michigan Press. An open access version of the report is available online, and physical copies are available for purchase. This report provides insight into the formation of the DCN and how shared work has been instrumental in founding, building, and sustaining the organization.

RDS staff also co-authored a research paper accepted to *PLOS One* based on the results of a 2021 survey to determine how user satisfaction with the data curation services provided by their respective institutions. Several other outputs were produced by the Institutional Outreach and Communications group, such as a peer compare session on institutional data policies, a customizable flyer template with DCN info on one side and institutional info on the other for easy adaptation, and a researcher stories series. Additionally, RDS staff helped cover for the DCN's Director when she was out of office by checking and assigning submitted datasets for curation. Our engagement with the DCN continues to enable valuable connections with other members, which helps us at Illinois directly as well as the greater data curation community.

Preservation Review of Datasets and RDS Extended Review

In FY23, we also reached milestones in two reviews. The first is a preservation review of datasets published in the Illinois Data Bank in 2016. This review is outlined in our policies (https://databank.illinois.edu/policies#preservation_review), where we indicate that we will do a 5-year review of datasets to assess on-going viability of the dataset. This review is an important part of our commitment to robust and thoughtful management of datasets deposited in the Illinois Data Bank. Initiation was delayed due to COVID, but was begun this year and lead by Hoa Luong, in her leadership role for the Illinois Data Bank. At the time of review, there were 49 datasets that had reached the 5-year mark since publication. Due to the small sample, the review was done manually based on four critical metrics: download count as proxy for interest and potential reuse; format score as a value assigned to each dataset based on the type of file formats contained; citation count as determined by the number of related materials associated with the dataset; and total size of the dataset in bytes. Among those 4 metrics, format score is manually recorded by the curator, where:

- Score of 1 is preferred format, meaning the file formats are open (non-proprietary) and recommended for long-term preservation, such as CSV, TXT, R.

- Score of 2 is of moderate concern. Datasets received score of “2” when they contained a mix of open and proprietary or less common formats. For example, a dataset that contains both CSV and Excel is scored as “2”.
- Score of 3 indicates a major concern, meaning the file formats are proprietary which requires special software to open and use. For example, MatLab is proprietary format found in some datasets and HDF5 is non-proprietary but encountered less frequently.

This review serves as an exploratory to assess the process, and as the result of the review, no datasets warranted a deeper review for possible deaccession. We learned that download times and format score play an important role in determining which datasets will need a deep review, which helps us prioritize features in the Illinois Data Bank that will assist in future management of the collection.

The second effort is an “Extended Review” of RDS services. This review is intended to be a comprehensive look at RDS activities in the last 5 years, as well as partnerships with other units, and an environmental scan to get a sense of the data services landscape externally, with the goal of helping us anticipate future needs. In FY23, we completed harmonization and analysis of service metrics and received feedback on the report structure in from communications experts and administrators. We’ll continue with the next parts of the review in FY24, and the final report will be shared with OVCR and published in IDEALS.

Feedback for 2022-2023 Services

Illinois Data Bank

From the department of **Aerospace Engineering**

I also want to again thank you for the useful suggestions and sharing the pre-filled README. This has made the data more usable for the readers and I will follow these suggestions in any subsequent submissions I make as well.

From the **Illinois Natural History Survey**

[Subject Specialist's name] has provided many helpful suggestions on my previous submissions so I guess my metadata creation skills are gradually improving!

From the **School of Information Science**

Thanks so much! We really appreciate your careful reviews!... I really appreciate being able to register a DOI ahead in this way!

Data Nudge

From the **Privacy Office**

I really like the comic book approach and especially the endangered animal example. So many great insights from that alone.

Data Management Workshop

Graduate students from the **Grainger College of Engineering**

Thank you so much for doing this workshop with us! A number of us already told me that they learned a lot today and will be applying those strategies to managing their research data.

DMP Review

From the **Siebel Center for Design**

You are amazing! Thank you so so much for your fast reply.

Changes

The RDS did not directly encounter any major challenges in FY23 within the unit itself, but there are several major changes within the Library that impact the RDS. First, the arrival of our new Library Dean challenges us to ensure that we can accurately communicate the value and impact of the RDS to the Library and campus as whole, as well as communicate the benefits of key memberships (e.g., the Data Curation Network and DataCite). Second, staffing changes in other units impact our collaborative network, including the several staffing changes in the Scholarly Commons, which historically provided a hub of support for GIS, text mining, data discovery, and other data-intensive research projects. Additionally, the PRI Librarian, who retired in spring of 2023, was a very close collaborator by virtue of PRI's highly data-centric operations. She provided front-line support for all PRI data consultations and DMP reviews, as well as curation of PRI datasets. She also participated in the Data Curation Network as a UIUC curator. The search for the PRI Librarian replacement is well underway but staffing and/or scoping changes in the Scholarly Commons seems likely to take longer to resolve.

Next-Generation Training in the Profession

Part of RDS's funding allocation continues to support a 0.5 FTE graduate assistant (GA) from the iSchool. This provides hands-on experience in research data services, which is a relatively new career path for Library and Information Science students, and at the same time provides support for the RDS. In FY23, Amina Malik started her first year as the RDS GA and helped implement Google Analytics for the Illinois Data Bank, contributed to the content development and preservation of Data Nudges, and assisted in Love Data week event workshop instruction. Amina is continuing in the RDS for a second year. In spring 2023, the RDS also hosted Uyen Nguyen as an iSchool practicum student who was trained and introduced to research data management while earning two credits toward her degree. Additionally, we hosted an iSchool PhD student, Chenyue Jiao, for a practicum in the summer of 2023, during which she did an

excellent landscape analysis on how data repositories link datasets to related resources and a bibliometric analysis of articles that mention use of Morrow Plots data. Practicums are atypical for iSchool PhD students, so Chenyue does not earn credit, but the experience was especially beneficial to her since she is interested in perusing a career as a data librarian after she completes her doctorate. As such, she requested another practicum in the fall of 2023, and we are delighted to work with her further.

Leadership in the Profession

External Presentations

- Imker, H.J., Schackart, K.E., Istrate, A., & Cook, C. "Discovering Biodata Resources from Across the Globe: an Open Data and Machine Learning Enabled Inventory" European Bioinformatics Institute, Hinxton, UK. July 6, 2022
- Caldron, S. "Historic Yields: Making Old Agricultural Data New Again." USAIN Agricultural Research Services Interest Group Meeting, online. September 28, 2022
- Caldron, S. "Making Infographics More Accessible." Midwest Data Librarian Symposium, online. October 7, 2022
- Anderson, B., Caldron, S., Farrell, S., Fenlon, K.S., Hendrickson, L., Kaplan, N. "Historic agricultural and environmental data: Facilitating discovery of, access to, and use of analog data. Research Data Alliance: Decade of Data Series, online. June 21, 2023.

Written Works

- Carlson, J., Narlock, M., Blake, M., Herndon, J., Imker, H., Johnston, L., Kozlowski, W., Xuying Xin, C., Lafferty-Hess, S., Luong, H., Marsolek, W., Moore, J., Scott, D., Hudson Vitale, C., Ezray Wham, B., & Wright, S. (2023). *The Art, Science, and Magic of the Data Curation Network: A Retrospective on Cross-Institutional Collaboration*. Michigan Publishing Services. <https://doi.org/10.3998/mpub.12782791>
- Marsolek W., Wright S. J., Luong H., Braxton S. M., Carlson J., & Lafferty-Hess S. (2023) Understanding the value of curation: A survey of researcher perspectives of data curation services from six US institutions. *PLOS One*. Accepted manuscript.
- Anderson, B. G., Caldron, S. L., Henry, J., Imker, H. J., Luong, H., Trei, K., & Williams, S. C. (2022, October 20). *Cultivating the Scientific Data of the Morrow Plots: Visualization and Data Curation for a Long-term Agricultural Experiment*. International Conference on Digital Preservation (iPres), Glasgow, Scotland. <https://doi.org/10.17605/OSF.IO/X8YMS>

2023-2024 for the RDS

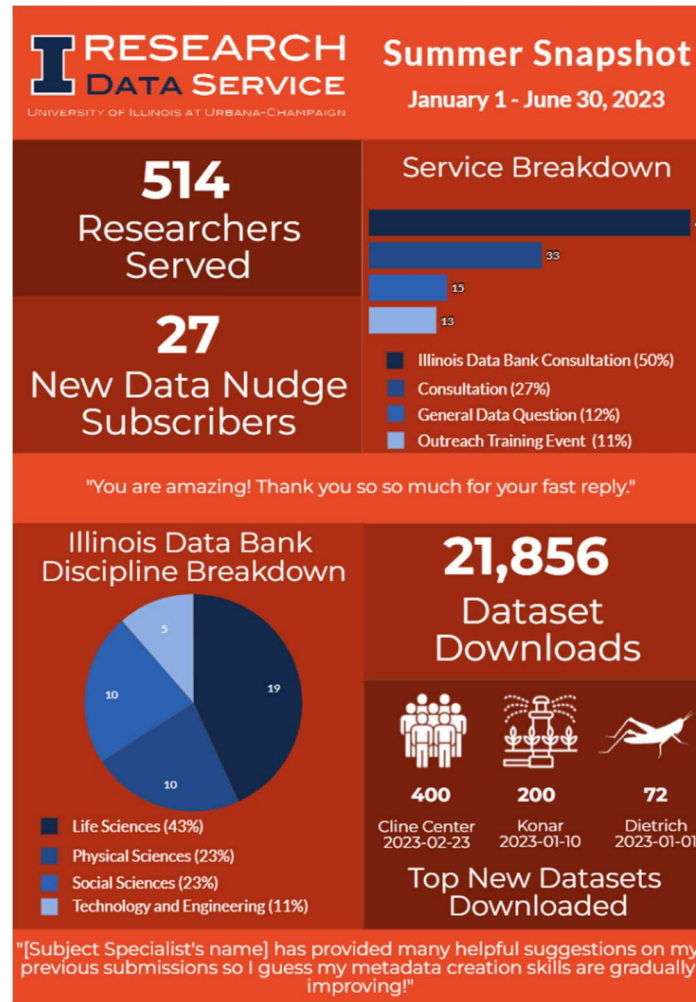
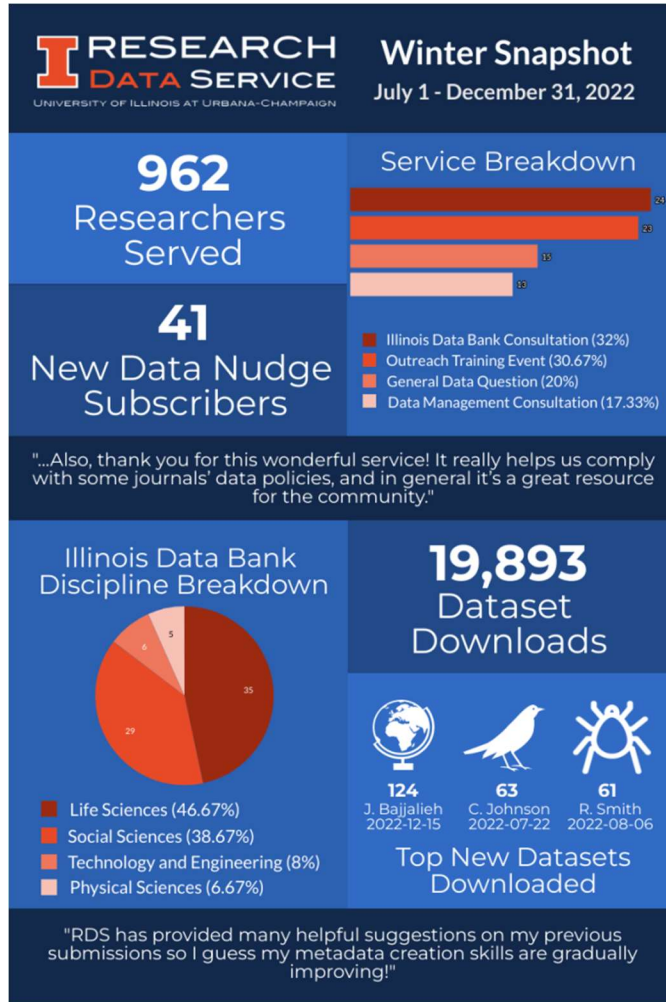
Core Staffing

- Heidi Imker, Director (1 FTE)
- Hoa Luong, Associate Director (1 FTE)
- Sandi Caldrone, Research Data Librarian (1 FTE)
- Colleen Fallaw, Research Programmer (1 FTE)
- Amina Malik, Graduate Assistant (0.5 FTE) – expected graduation May 2024
- Dena Strong, Senior Information Design Specialist (0.1 FTE volunteer time from Tech Services)

Annual Goals for FY24

- Complete the next phase of the RDS Extended Review and submit a revision of the Core Trust Seal application for repository certification.
- Implement current RDS communications plan to create a robust and cohesive campaign that delivers our key messages to our core audiences.
- Recruit new researchers to be featured on the Researcher Spotlight to promote the value of research data sharing at Illinois and the Illinois Data Bank as a resource.
- Continue in-depth engagement in Data Curation Network, for example, by leading the Data Curation Network's Outreach and Communication Interest Group with the production of more useful marketing and outreach materials and knowledge
- Implement the new, robust workflow for versioning process and implement effective automated tests to make future feature development and refactoring smoother
- Research and refine how related materials are displayed in the Illinois Data Bank

Appendix





Backup sticker for outreach