

Research Data Service Annual Report

2021-2022

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.

2021-2022 Goals

- Provide a welcoming and supportive on-boarding experience for the new Research Data Librarian (Oct 2021) and the acting director during sabbatical coverage (Feb-Aug 2022)
- Implement the 5 Year Extended Review based on the plan developed in FY21
- Re-design and update the RDS website
- Deliver formal workshops to train others at Illinois in data curation
- Maintain high-quality operations and circumspect development of the Illinois Data Bank
- Continue to advance research data stewardship through contributing to the Data Curation Network, conducting thoughtful research, and presenting our work nationally and internationally

Major Activities and Accomplishments

FY22 marked as a successful year as we accomplished all the goals set for the year and were back to full staff. Major activities and accomplishments included:

- **Successfully transitioned** individuals into new roles, including Sandi Caldron as Research Data Librarian (began Oct 2021), Chris Wiley as Acting Director (Feb-Aug 2022), and Amina Malik as RDS GA (FY23)
- **Improved and extended communications** through a comprehensive re-design of the RDS homepage, reorganizing and updating the Illinois Data Bank Help page, and creating checklists and guides for Data Transfer and Use Agreements and addressing data sharing during informed consent
- **Strengthened curation services** by 1) delivering a formal data curation workshop for UIUC Library subject specialists whose disciplines are likely to deposit datasets and 2) continuing our deep collaboration with the Data Curation Network (DCN) through exchange of curation expertise and a second year of support as a Sustaining Member
- **Continued** high-quality operations for the Illinois Data Bank, which in FY22 included maintaining a functional codebase, transitioning to a new ticketing system, and applying for designation as a trustworthy repository through CoreTrustSeal
- **Maintained** Illinois' position as a national and international leader in data services by 1) submitting several conference articles and posters describing RDS projects and 2) presenting at international symposiums, including the Beilstein Open Science

Symposium (Germany) and the Kyushu-Illinois Strategic Partnership Colloquia Series, and 3) hosting visitors and researchers interested in gaining an in-depth understanding of the RDS's activities and work culture via participation in meetings, shadowing, and extended conversation

One of the goals for last year was to implement a 5 Year Extended Review of the RDS. However, we postponed this goal due to the Director's sabbatical leave and to focus our efforts on outreach activities. We will shift this goal to FY23, and we will also revisit our intention to review datasets after 5 yrs. Likewise, our CoreTrustSeal application was still under review at the time of writing this report. Thus, we will report the result in FY23.

Staffing

In October 2021, Sandi (iSchool graduate of 2014) began at the RDS as the Research Data Librarian. Sandi came to us from Purdue University, Indiana, where she served as the outreach specialist for the Libraries Research Data team, curating datasets, creating outreach materials, leading webinars and workshops, and consulting with researchers on data management policies and best practices. As the new Research Data Librarian, she hit the ground running, offering consultations, updating data management workshops, contributing to the Morrow Plots data curation working group, and building relationships with other research support professionals around campus.

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. Dataset deposits continued to grow in FY23 (Figure 1), although the number of deposits in FY22 (97) was lower than FY21 (139). This is largely an artifact of how the fiscal year fell as the 2020 and 2021 calendar years were nearly equal at 122 and 126, respectively. Activities of the Illinois Data Bank still accounted for half of RDS activities overall (Figure 4). As of June 2022, the Illinois Data Bank had published 517, including versions, and total downloads rose to 299,180. The top ten depositing departments last year continued to deposit more datasets, representing six different colleges and research institutes (Figure 2).

Developer time in the Scholarly Communication and Repository Services group was necessarily focused on the migration of the IDEALS institutional repository in FY22 which also influenced Medusa, the core preservation system the Illinois Data Bank depends on. Necessary, but not urgent, improvements were instead noted and slated for future development in FY23. Regardless, while we remained circumspect and no new features were introduced for the Illinois Data Bank this year, the system still required steady effort from RDS Programmer Colleen to maintain the codebase, fix bugs, and manage unexpected service outages.

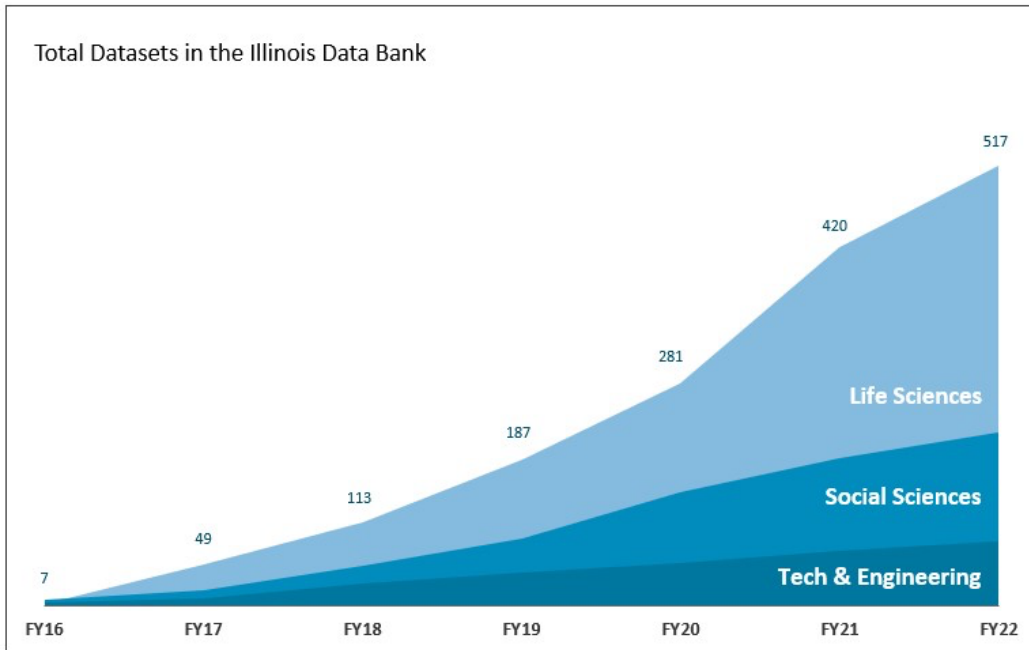


Figure 1. In FY22, the total number of datasets in the Illinois Data Bank reached 517, including versions.

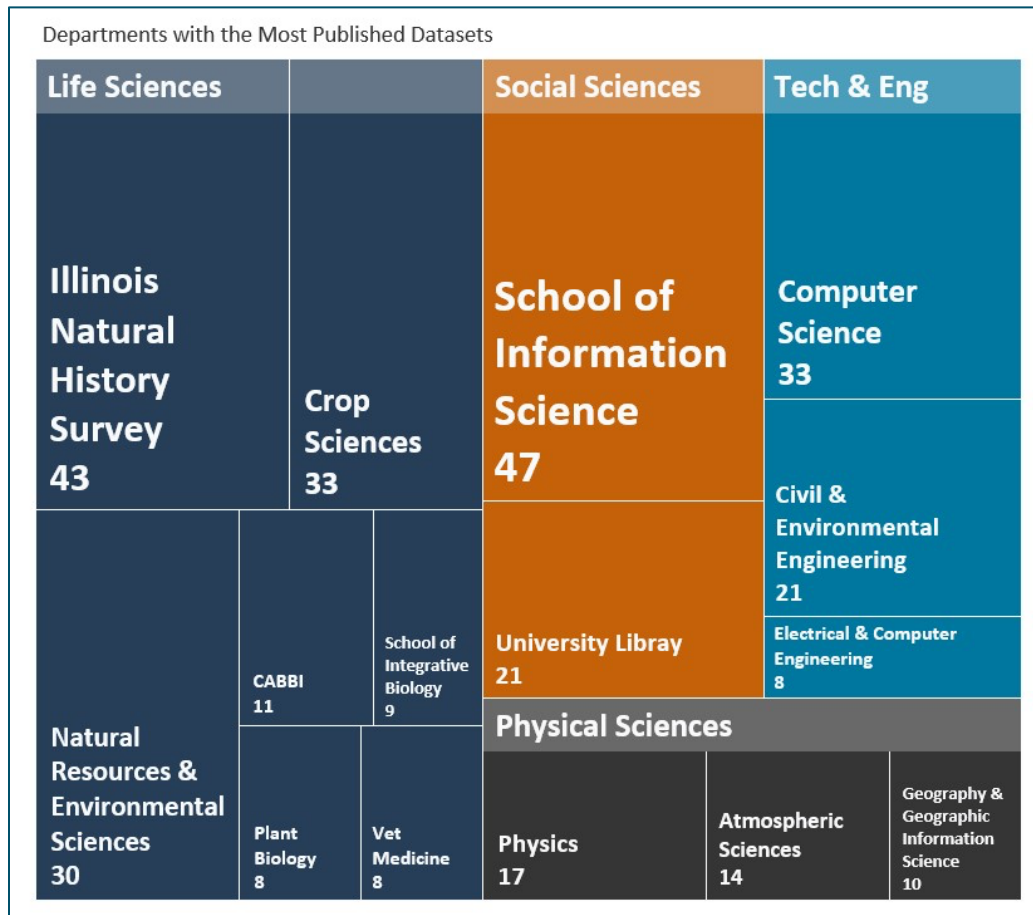


Figure 2. As in the previous year, the most prolific departments are largely STEM-focused and represent Liberal Arts and Sciences, the Grainger College of Engineering, Agricultural Consumer and Environmental Sciences, the iSchool, the University Library, and the Prairie Research Institute.

Data Curation Network

FY22 officially concluded the "Launching the Data Curation Network (DCN)" grant from the Alfred P. Sloan Foundation. Illinois continues as one of the seventeen (17) sustainable members (Figure 3) in the DCN. Sandi was introduced to the DCN as a curator and was recently invited to serve as a Curator at Large, starting in August 2022. In this role, Sandi will represent all DCN Curators in governance meetings and elections. Heidi continued to serve as our institutional representative, and Hoa filled in from February to August 2022 while Heidi was on sabbatical leave.



Figure 2. DCN sustaining institutions

In March 2022, RDS staff participated in the DCN project retrospective meeting to reflect on the Sloan grant. The project team examined the project's success, identified several key enabling structures, discussed the challenges facing the DCN, and reflected on the future of the DCN and other potential collaborative endeavors. A report was produced as the result of the meeting and will be published by the University of Michigan Press as a Maize Book in March 2023.

RDS staff continued to meet the goals for Illinois' participation in DCN by attending meetings, exchanging dataset curation, and participating in several DCN interest groups and projects. For example, RDS members collaborated with others in the DCN to present the results of the End User Satisfaction Survey, a study to determine how satisfied users are with data curation services provided by their respective institutions, in poster format, at the 17th International Conference on Open Repositories. The poster won second place and received many compliments from peers. An article based on these results will be submitted to a peer-reviewed journal. Colleen was invited to speak to the DCN Big Data interest group about her experience implementing Globus for the Illinois Data Bank. In the Institutional Outreach and Communications group, we developed a set of interview questions that can be easily adapted by other research units to understand researchers' motivations and thoughts on data sharing and curation. RDS staff volunteered to pilot the interview tool and shared the resulting story in the DCN newsletter. We shared the story as well as successes and challenges using this format with the group. Finally, RDS staff helped plan and lead the annual DCN All Hand Meeting. Collectively, these activities have been extremely helpful for building connections with other DCN members as well as evaluating, benchmarking, and highlighting our current services here and inspiring future directions.

Outreach and Engagement

Maintaining a flexible work environment, RDS staff worked mainly onsite with some remote days in FY22. Practicing caution with COVID19, even though staff were back on-site, all activities were delivered through Zoom for the safety and convenience of our researchers (although we note that in person requests have now picked back up for FY23). In FY22, we served both individual users and groups of users through 179 distinct interactions, including consultations, dataset curation, data management plan reviews, and workshops (Figure 4). Although distinct events were down 17 percent from last year, the number of users directly served increased significantly compared with previous year and even exceed the year before (from over 1,400 users in FY20, down to around 630 in FY 21, and now over 1,500 users this year). This was because most campus events (e.g., ISSS's student orientation) were back to normal. In addition, RDS staff were invited to give a guest lecture for the IS 537 Theory and Practice of Data Cleaning class at the iSchool in Fall 2021.



Figure 4. 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 FY22.

With the new Research Data Librarian on board, we were able to increase our capacity for workshops. Additionally, FY22 saw new relationships with units on campus that we had not worked with before. For instance, we collaborated with the Office of Privacy and Cybersecurity to deliver a new “Storage, Privacy, and Security” workshop to researchers at Gies College of Business. The session was recorded to provide access to others who could not attend and, according to the coordinator at Gies, it was well-received. We were also invited to deliver a workshop for students and faculty at Center for Power Optimization of Electro-Thermal System (POETS). By establishing new connections, we were able to reach out to more colleges and departments and even with colleagues in Japan. RDS staff hosted a half-day meeting with a visiting scholar from Kyushu University, Japan, who wanted to learn more about the unit history,

services, workflows, and infrastructure, and he also participated in several RDS meetings. In collaboration with the iSchool and the Grainger Engineering Library, we participated in Kyushu-Illinois strategic partnership webinar series, Data Curation and Information Management in the Age of Digital Transformation. We also are in the planning phase with Research IT to provide more data management workshops to IT professionals and attend more IT events. As researchers often reach out IT professionals for their data needs, through a “training the trainer” model, we hope to strengthen this avenue of data management support for researchers.

Activities that have always been entirely virtual by nature, namely Illinois Data Bank deposits and the Data Nudge open rates, were steady in FY22. We celebrated [5 years of the Data Nudge](#) in December 2021 with a special issue which summarized Data Nudges successes in the past 5 years. Subscriptions continued to increase and hit nearly 600 subscribers at the time of writing this report, and in FY22 the average open rate continues to be double the industry standard for education. As in FY21, deposit related activities (e.g. pre-publication review and consultation) contributed to more than half of the RDS activities performed in FY22 (57%) (Figure 4).

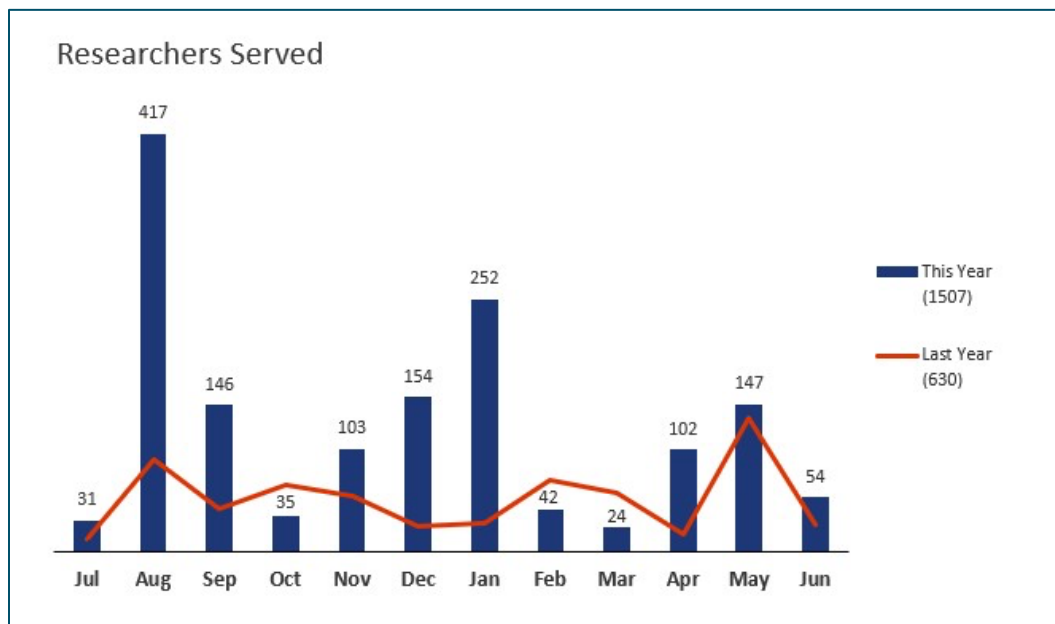


Figure 5. User engagement in FY22 increased significantly through outreach and training (e.g., ISSS Student orientation in August) as activities returned to pre-pandemic levels.

In FY22, we continued to partner with subject specialists on curation. We delivered a formal data curation workshop to subject specialists that we have been working with or expect to work with. This workshop helped them to understand more about the Illinois Data Bank, our workflow, how we communicate with researchers, and the subject specialists’ role in data curation. The workshop was well-attended and received positive feedback from subject specialists. We also created a checklist to assist subject specialists during curatorial review. For the future, we may consider offering this workshop upon request.

Finally, with the assistance of the Library Web Team, we launched a new RDS website last August. The website was organized around the services that the RDS offers and provides more definitions, guides, and templates for researchers to access at the point of need. The website also offers us greater flexibility to more easily add new content. We added a “Recent News” feature as a place to announce data-related information from both the RDS (e.g. newly published datasets) or others on campus (e.g. OPRS’s Guidance for the Use of Zoom in Data Collection, Technology Services’ Guide to Cloud Storage Essentials, etc.).

Feedback for 2021-2022 Services

Illinois Data Bank

From the department of **Natural Resources and Environmental Sciences**

You are amazing to work with. Thank you! I appreciate all of your suggestions.

From **Mathematics**

Thank you very much for reviewing our files. Your past comments on our other files have been quite helpful to us.

From **Physics**

Thank you for making sure our dataset is understandable to others.

Data Nudge

From an **external subscriber**

The nudge that has had the largest impact on my computing habits has been "Love Your File Directory." This nudge moved me to think through the file structure I use and to find one that works on the four different computing platforms that I work on.

Data Management Workshop

From the **Gies College of Business**

Thank you all so very much for taking time out of your busy schedules to talk with us today. The information you provided will certainly assist our researchers in their data collection process. While the in-person group was small, I have already had requests for the recording which is great! Thank you for being an amazing resource for the research community.

DMP Review

From the **Illinois State Water Survey**

Wow! Thank you...for your thorough feedback.

Major challenges

Relative to the staffing and pandemic disruptions of past years, we did not encounter any major challenges in FY22. The Research Data Librarian quickly adapted to the unit, the Library, and campus. Serendipitously, workshop requests increased shortly after the Research Data Librarian came on board as the campus as a whole moved towards operations that more closely resembled those prior to the pandemic. With the Director on sabbatical leave, we postponed the 5-Year Extended Review. The review planned is extended and extensive so in FY23 we will stagger activities so that we can also take on the 5-year review of datasets deposited in the Illinois Data Bank. Both reviews will help us carefully evaluate RDS operations and services and plan for the future.

Next-Generation Training in the Profession

Every year part of the RDS's funding allocation has gone towards supporting a 0.5 FTE graduate assistant (GA) from the iSchool. This provides support for the RDS, but more practically (and importantly), it also provides much-needed hands-on experience in research data services, which is a relatively new career path for Library and Information Science students. In FY22, Lauren Phegley continued her second year as the RDS GA and contributed to several RDS projects, including helping with the website design, creating content for Data Nudges, assisting with workshop instruction, and participating in the Library DEIA Task Force. Lauren graduated in May 2022 and successfully launched a career as Research Data Engineer in the Library at UPenn. In this role, Lauren is leveraging her experiences in the RDS to assist their researchers with managing research data throughout the research data lifecycle. We successfully found a replacement for Lauren who will start in August 2022.

Leadership in the Profession

External Presentations

- Imker, Heidi (2021). "Making Research Data Ready for Open Science: Universities Responding to Challenges in a Heterogeneous Environment" Beilstein Open Science Symposium. Virtual. October 6, 2021
- Imker, Heidi, Hoa Luong, Bill Mischo, Mary Schlembach, and Chris Wiley. (2021). "Self-Reported Data Reuse Associated with Highly Cited Journal Articles". Kyushu-Illinois Strategic Partnership Colloquia Series: Data Curation and Information Management in the Age of Digital Transformation. Virtual. December 8, 2021
- Imker, Heidi, Schackart, Kenneth, Istrate, Ana-Maria, and Cook, Chuck (2022) "Using Open Data and Reproducible Methodologies to Create a Global Inventory of Biodata Resources" International Society of Biocurators 1st UK-Local Biocuration Conference, Hinxton, UK. May 5, 2022
- Wright, Sarah, Wanda Marsolek, Hoa Luong, Susan Braxton, Sophia Lafferty-Hess, and Jake Carlson. (2022). "Researcher Approved: a Multi-institutional Survey of Depositors

to Six Academic Data Repositories”. Presented at the International Conference on Open Repositories, Denver, CO. 6th - 9th June 2022

Written Works

- Anderson, Bethany G., Sandi L. Caldron, Joshua Henry, Heidi Imker, Hoa Luong, Kelly Trei, and Sarah C. Williams. (2022). “Cultivating the Scientific Value of the Morrow Plots: Visualization and Data Curation for a Long-Term Agricultural Experiment”. Accepted a short paper to present at the International Conference on Digital Preservation 2022 (iPres), in Glasgow, Scotland
- Narlock, Mikala, Hoa Luong, and Jon Peters. (2022). “We’re all doing the best we can with what we’ve got: Preservation practices of Data Curation Network members”. Accepted a long paper to present at the International Conference on Digital Preservation 2022 (iPres), in Glasgow, Scotland

2021-2022 for the RDS

Core Staffing

- Heidi Imker, Director (1 FTE)
- Hoa Luong, Associate Director (1 FTE)
- Sandi Caldron, Research Data Librarian (1 FTE)
- Colleen Fallaw, Research Programmer (1 FTE)
- Lauren Phegley, Graduate Assistant (0.5 FTE) – graduated in May 2022 and currently is Research Data Engineer at the University of Pennsylvania
- Dena Strong, Senior Information Design Specialist (0.1 FTE volunteer time from Tech Services)

Annual Goals for FY23

- 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
- Prepare for NIH’s implementation their new Data Management and Sharing Policy in January of 2023 by providing proactive announcements, presentations, guidance, and Plan reviews
- Implement improvements for the Illinois Data Bank, including refinement of 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 Illinois Data Bank dataset deposits

Appendix

