Research Data Service Annual Report

2020-2021

> *"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."*

2020-2021 in Review

2020-2021 Goals

- Expand on the 5 Year Review and implement key recommendations that surface
- Articulate replacement position and complete hiring process
- Engage with and build on existing diversity and inclusions efforts within unit, Library, and on campus
- Implement mission-critical feature developments for the Illinois Data Bank
- Continue to engage with organizations and communities of practice dedicated to advancing data stewardship, including the Data Curation Network
- Strengthen existing and develop new outreach mechanisms to promote the tools and services provided by the RDS
- Review and update service polices
- Begin 5-year review of early dataset deposits

Major Activities and Accomplishments

Overview

Despite the uncertainty during the pandemic and staffing constrains, FY21 was a successful year. While some activities were hampered while completing the hiring process for our replacement position and due to COVID-19, the RDS made progress on all 2020-2021 goals. Major activities and accomplishments included:

- **Mapped the process** for the RDS extended review
- **Strengthened staffing** by 1) successfully completing the hiring process for a new Research Data Librarian (Oct 16 2021 start), 2) updating the Assistant Director to Associate Director to accurately reflect needs and actual duties, and 3) onboarding a new graduate assistant who is staying on for two years
- **Held three DEIA conversations** with RDS core and affiliated staff to better understand ourselves and our campus resources
- **Implemented key features** in the Illinois Data Bank, including a private sharing link feature which allows depositors to share datasets with journal reviewers and/or external collaborators

during the draft (unpublished) stage; also completed implementations to better detect and fix bugs on the back end

- **Maintained Illinois' position as a national leader** in data services by 1) became a sustaining partner in the Data Curation Network, 2) published several peer-reviewed articles describing the work in the RDS, and 3) contributed to the APLU-AAU's initiative to accelerate data sharing
- **Continued broad campus outreach** for the Illinois Data Bank and RDS through maintaining professional relationships with library subject specialists, academic departments, and administrative units; for example, by participating for a third consecutive year in ISSS's international graduate student welcome and orientation
- **Expanded** collaborations with library subject specialists in engineering, medicine, GIS, education, humanities, and in the life and social sciences to provide domain knowledge during dataset curation and DMP reviews
- **Reviewed and updated** Illinois Data Bank policies; **created** new policies for RDS Service Standards and RDS Confidentiality

We postponed the 5-year review of dataset deposited in 2016 to prioritize other goals in FY21; likewise, while we established a roadmap for the RDS extended review, this work was not completed in FY21. We will report on these goals in FY22 annual report.

Illinois Data Bank

Going to its fifth year, the Illinois Data Bank is recognized as the most useful service offered by the RDS. As Illinois' institutional data repository, the Illinois Data Bank provides researchers with a free, stable, and convenient place to share finished research data with the world. FY21 saw an increase in the number of published datasets and downloads (Figure 1). Through June 2021, there are now 420 published datasets, including versions, and downloads doubled last year to 269,366. Furthermore, the use is broad across campus, with the top ten depositing departments (Figure 2) representing six different colleges and research institutes.

One important feature that was implemented during FY21 was the ability to generate a private sharing link. Over the years, we've received questions off and on from researchers about the ability to provide access to draft datasets during peer review. We saw increased interest in this feature this year, and the RDS Associate Director and RDS programmer completed comprehensive reviews to learn how other institutions and repositories apply this feature. After several rounds of discussion, in May 2021, the sharing link feature was implemented on the site. With this feature, researchers can generate a private link to share with their journal and/or external collaborator(s). Although new, several users have already adopted the feature, and we have received compliments on its usability.

As part of our on-going commitment to support the larger repository and curation communities, the RDS staff published two peer-review articles describing our curation workflows and technical challenges in the Illinois Data Bank. The first article describes curation in practice in response to the call from the Data Curation Network and appears in the *Journal of eScience Librarianship*. The second article describes our adoption of cloud services and was published in *Code4Lib*, which focuses on library technology. Despite only being out a few months, this latter paper has been cited already as a rare example in the literature that documents cloud-based infrastructure.

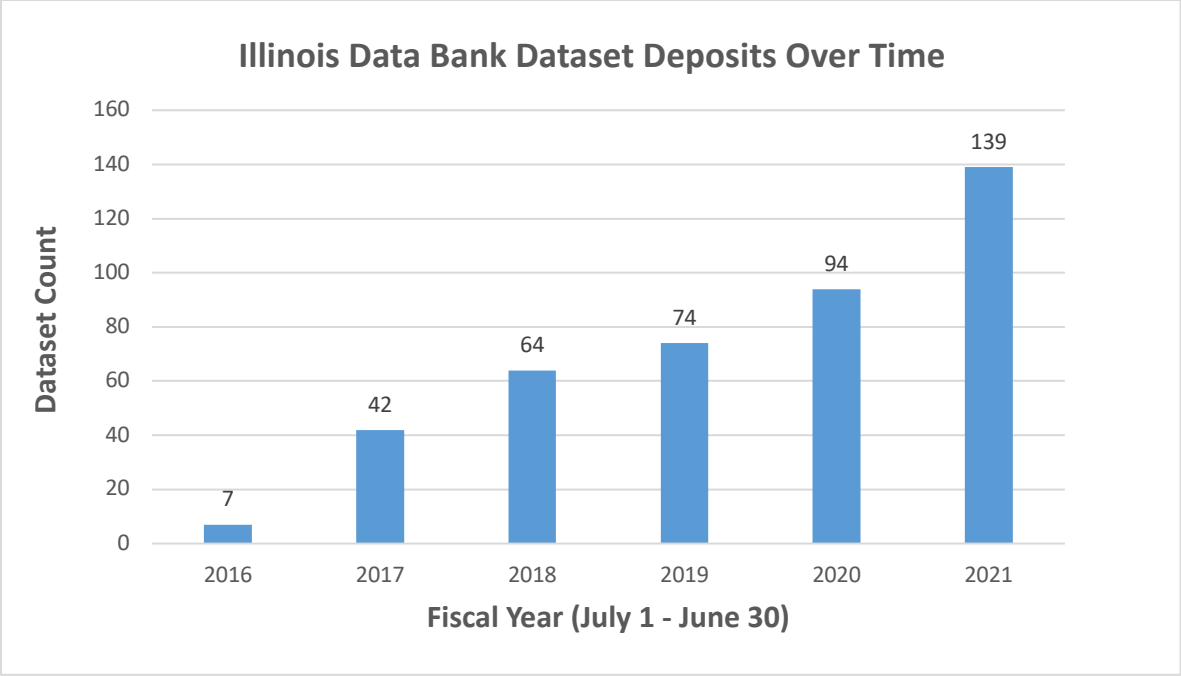**Illinois Data Bank Dataset Deposits Over Time**

Figure 1: Number of published datasets in the Illinois Data Bank grew steadily between May 2016 to July 2021, with users publishing a total of 420 datasets (including versions) to date.

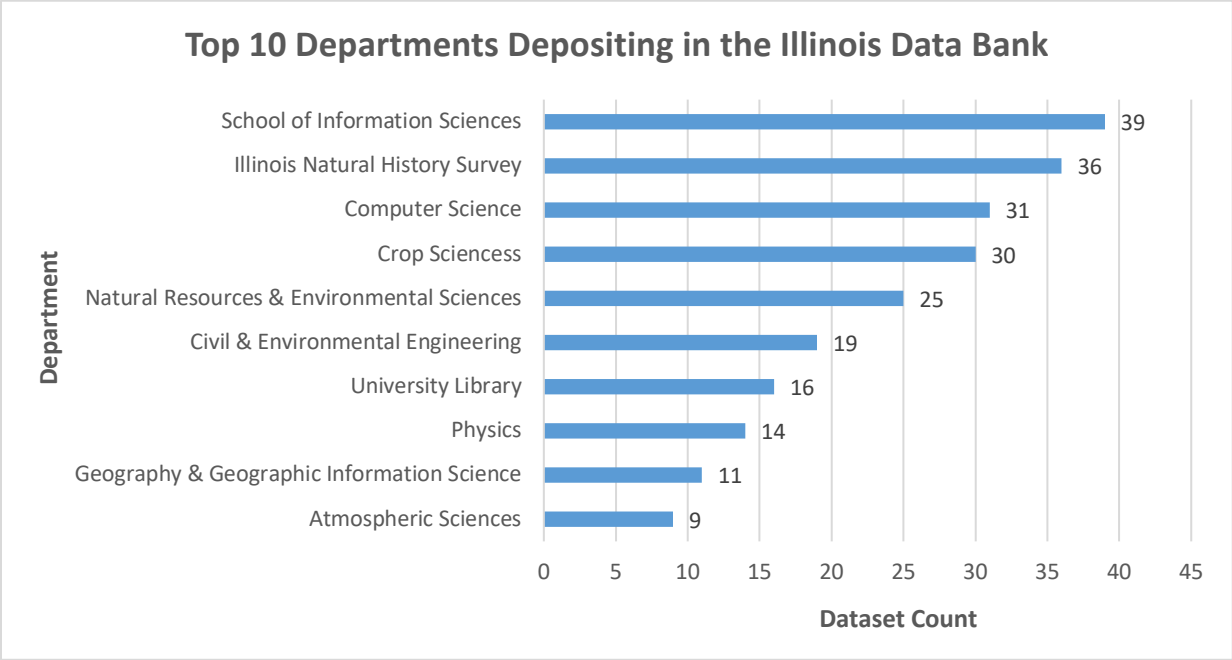**Top 10 Departments Depositing in the Illinois Data Bank**

Figure 2: To date, the top ten departments publishing datasets in the Illinois Data Bank are often STEM-focused and represent the following colleges and research institutes: Liberal Arts and Sciences, the Grainger College of Engineering, Agricultural Consumer and Environmental Sciences, the iSchool, the University Library, and the Prairie Research Institute.

RESEARCH
DATA SERVICE
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Data Curation Network (DCN)

The Data Curation Network (DCN) serves as a "human layer" in the data repository stack by connecting local datasets to expert data curators via a cross-institutional shared staffing model. This includes both curation of datasets across institutional boundaries and supporting the data curator community by providing professional development opportunities. The DCN was initially funded by the Alfred P. Sloan Foundation, beginning with a one year planning grant in 2016 and then a three year implementation grant in 2018. With the successful conclusion of those grants and extensive sustainability planning, this year the DCN officially transitioned from a grant funded project to a sustainable, member-driven organization. The RDS worked extensively with DCN on development of the sustainability plan and launch of the model, which has been well-received to date. Locally, once the model was in place, we worked with our administration and business office to become one of thirteen Sustaining Partners.

In terms of DCN operations, RDS staff continued meeting the goals for Illinois' participation by attending bi-weekly meetings, weekly curator stand-ups, submitting and curating datasets to/for the DCN. RDS members also participated in several DCN's interest and research groups. For example, in the Value of Data Curation group, RDS members along with several colleagues surveyed U.S. repository staff and user perceptions about the value of data curation. We designed a survey for all repositories based in the U.S. The preliminary results of this survey were presented at the 2021 International Digital Curation Conference (IDCC), and a publication will be submitted in a peer-reviewed journal. In another group, the End User Satisfaction Survey, RDS members collaborated on a study to determine how satisfied users are with data curation services provided by their respective institutions. The survey was sent to our Illinois Data Bank's users, and the results was presented at the DCN All Hand Meeting in June 2021. Finally, in Institutional Outreach and Communications group, we shared and discussed our outreach strategies for the RDS, including avenues such as Data Nudge, RDS snapshots, and Researcher Spotlights. While these activities are not the primary reasons we participate in the DCN, they have been extremely helpful for evaluating, benchmarking, and highlighting our current curation services and inspiring future directions.

Outreach and Engagement

RDS staff worked remotely in FY21, and the unit quickly adapted to online teaching and hosted all workshops through Zoom. RDS staff served both individual users and groups of users through 216 distinct interactions, including consultations, dataset curation, data management reviews, and workshops (Figure 3). All activities were delivered online, and the number of users that RDS staff directly served effectively dropped in half compared with previous years, from over 1400 users in FY20 to around 630 users this year. This is not a necessarily a reflection of less participants at events, but mostly fewer events; in FY20 we held or participated in 47 distinct events while in FY21 that number was only 23. However, some events in the past which reached a few dozen users were also less well-attended this year. Anecdotally, this seems tied to several things. We observed more chaos in the planning and marketing of events as well as general fatigue in attending online sessions. It was also harder to get accurate tallies since some events were held asynchronously. On the other hand, activities that have always been entirely virtual by nature, namely Illinois Data Bank deposits and the Data Nudge open rates, were steady or even increased in FY21. For example, the Data Nudge hit 500 subscribers and maintained an average open rate double the industry standard for education in FY21. Dataset-related activities (e.g. pre-publication review and consultation) contributed to more than half of the RDS activities performed in FY21 (57%) (Figure 3).
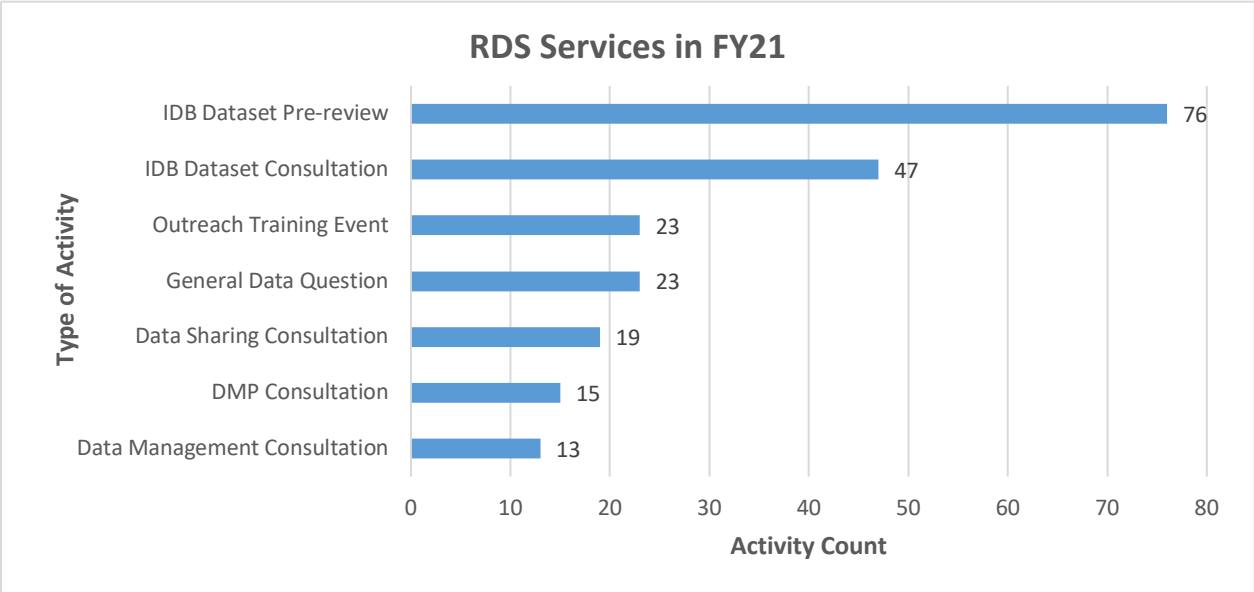
RDS Services in FY21

Type of Activity	Activity Count
IDB Dataset Pre-review	76
IDB Dataset Consultation	47
Outreach Training Event	23
General Data Question	23
Data Sharing Consultation	19
DMP Consultation	15
Data Management Consultation	13

Figure 3: A breakdown of activities by type shows that work related to the Illinois Data Bank, which is an exclusively virtual service, dominated RDS service offerings in FY21.
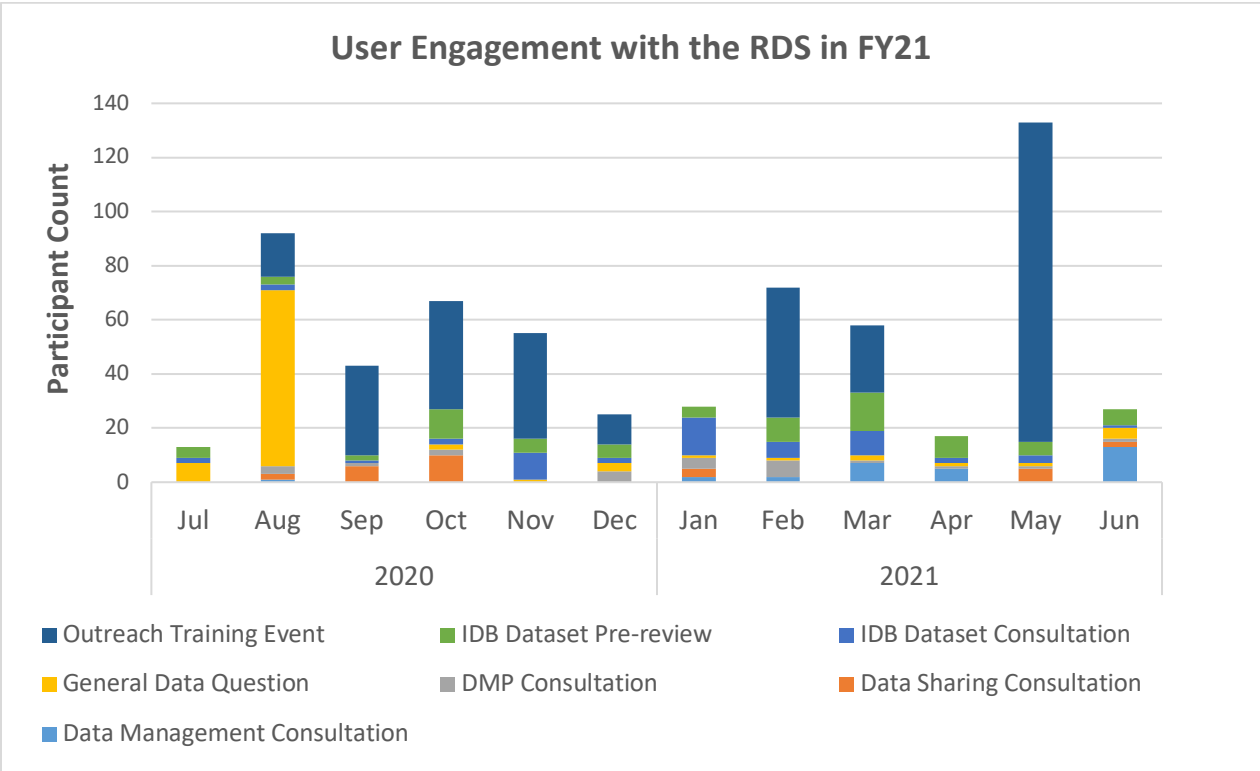
User Engagement with the RDS in FY21

Legend:
- Outreach Training Event
- IDB Dataset Pre-review
- IDB Dataset Consultation
- General Data Question
- DMP Consultation
- Data Sharing Consultation
- Data Management Consultation

Figure 4: User engagement in FY21 was more equally distributed across RDS activities than in previous years, although, not surprisingly, outreach still remains the most expedient way to engage with a greater number of users.

To adjust to the different on-campus dynamics during COVID-19, we pivoted in a number of ways, which strengthened the RDS overall and ultimately helps to pave the way for on richer outreach in the future. For example, with more datasets being deposited in the Illinois Data Bank, in FY21 we collaborated with more subject specialists in the Library for curation, specifically those with expertise in in medicine, GIS, languages, and the digital humanities. In this process, colleagues were invited to assist us with reviewing the datasets from those disciplines. This work is mutually beneficial; it provides another way for librarians to connect with departments and gives them a unique view into research activities. For the RDS, it helps us build another level of coverage for when curation expertise is not found directly in the RDS (the other being the DCN). One goal we have for FY22 is to host a formal data curation workshop to introduce and engage others on campus, and especially subject specialists, in data curation work.

Feedback for 2020-2021 Services

Illinois Data Bank

From the department of **Center for Advanced Bioenergy and Bioproducts Innovation** (**CABBI**):

"Thanks for pushing this through and on the weekend. I really appreciate you taking your personal time to do it. Thank you again. I really appreciate your work on this!"

From the **Cline Center**:

"Thanks for all of your help! As always, we really appreciate what you and your group does. You all provide such a great service."

From the **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."

Data Nudge

From the **Gies College of Business:**

"I am working on my dissertation right now and have really appreciated your tips about data accessibility and backups. Thanks to your data backup nudge, I started saving things to Box AND Google Drive AND emailing myself the documents whenever I make substantial progress with writing or editing. Cloud services could go down at any time and it's been great for my peace of mind to do this more intentionally. "

Data Management Workshop

From the department of **Student Affairs Assessment Committee**:

"Thanks so much for coming and speaking for us! It was a great session."

DMP Review

From a subject specialist at the **Library**:

"Thank you for sharing this DMP. I always learn a bit more about the process used by the Research Data Service as well as the research of faculty in the College of [name]"

Major challenges

Like FY20, lack of staff and the COVID-19 pandemic were two major challenges for the unit. As discussed above, the overall number of researchers served dropped due to a variety of COVID-19-related reasons. Also, due to staffing, we did not as aggressively market the services as we have done in past years. As in past staffing shortages, we relied on our colleagues in the Library and in Research IT for support. For example, Susan Braxton, PRI Librarian, very ably took on curation responsibilities within the DCN to cover Illinois' commitment to the network (as well as cover our formal cost share contributions to Sloan). With the new librarian joining us in FY22 and the resumption of more in-person activities on campus in FY22, we expect to see more interactions next year and renewed stability for the unit.

Next-Generation Training in the Profession

Since the RDS was established in 2014, every year part of the 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 FY21, Lauren Phegley joined us as the RDS GA and contributed to several RDS projects, including early planning for a major website resign, creating content for Data Nudges, and participating in workshop instruction. In Spring 2021, she was nominated to be one of two GA representatives on the Library's the DEIA Task Force and since then has been deeply involved in the Assessment Sub-Team. Lauren is continuing with the RDS for another year and plans to pursue a position in data librarianship after graduation. She will be the second RDS GA to do so.

Leadership in the Profession

Activities with the APLU and AAU

The Accelerating Public Access to Research Data Initiative was launched by the Association of Public and Land-grant Universities (APLU) and Association of American Universities (AAU) in 2017 with the goal of identifying and promoting practices to improve public access to research data. The RDS Director served on the Steering Committee for this initiative and contributed to the [Guide to Accelerate Public Access to Research Data](#). The guide was published in spring 2021 and highlights both the Data Curation Network and Illinois' Data Nudge. Additionally, APLU and AAU hosted a four-part webinar series to accompany the guide, and the RDS Director presented and moderated *Building Capacity and Buy-in* session. Coming from the APLU and AAU, these activities are especially useful for promoting cohesive practices across US institutions, which is a major challenge, and it has been a helpful vehicle for sharing strategies.

External Presentations

- Kollen, Chris, Rio, Fernando, and Heidi J. Imker. "Developing and Sustaining RDM Services." OCLC Works in Progress Webinar Series, February 9, 2021.
- Henry, Josh, Bethany Anderson, Heidi J. Imker, Hoa Luong, Kelli Trei, and Sarah C. Williams. "Digging into the data of the Morrow Plots: an interdisciplinary approach". E-poster and video overview at: American Association for the Advancement of Science (AAAS) Annual Meeting, February 17-20, 2021.
- Imker, Heidi and Lisa Johnston. "Grants Are Temporary but Data Stewardship Is Not: Sustainability of the Data Curation Network" Coalition for Networked Information (CNI) Spring Meeting, March 15th, 2021.

- Curty, Renata, Lisa R Johnston, Sophia Lafferty-Hess, Hannah Hadley, Hoa Luong, Susan Braxton, Jonathan Petters, Jake Carlson, and Wendy Anne Kozlowski. "Understanding the Value of Curation: Preliminary survey results of data curation practice and perception." International Digital Curation Conference, April 19, 2021.
- Imker, Heidi J., Hoa Luong, William H. Mischo, Mary C. Schlembach, and Chris Wiley. "An examination of data reuse practices within highly cited articles of faculty at a research university." International Digital Curation Conference, April 19, 2021.
- Orlowska, Daria, Colleen Fallaw, Yali Feng, Livia Garza, Ashley Hetrick, Heidi Imker, and Hoa Luong. "Better Data Management, One Nudge at a Time." International Association for Social Science Information Service and Technology (IASSIST), May 19, 2021.

Written Works

- Imker, Heidi J., Hoa Luong, William H. Mischo, Mary C. Schlembach, and Chris Wiley. 2021. "An examination of data reuse practices within highly cited articles of faculty at a research university". The *Journal of Academic Librarianship* 47, 102369. https://doi.org/10.1016/j.acalib.2021.102369
- Luong, Hoa, Colleen Fallaw, Genevieve Schmitt, Susan Braxton, and Heidi Imker. 2021. "Responding to Reality: Evolving Curation Practices and Infrastructure at the University of Illinois at Urbana-Champaign". *Journal of eScience Librarianship* 10, e1202. https://doi.org/10.7191/jeslib.2021.1202
- Fallaw, Colleen, Genevieve Schmitt, Hoa Luong, Jason Colwell, and Jason Strutz. 2021. "Institutional Data Repository Development, a Moving Target." *Code4Lib Journal*, Issue 51, 2021-06-14. https://journal.code4lib.org/articles/15821
- Orlowska, Daria, Colleen Fallaw, Yali Feng, Livia Garza, Ashley Hetrick, Heidi Imker, and Hoa Luong. 2021. "Better Data Management, One Nudge at a Time." *IASSIST Quarterly* (in press)

2020-2021 for the RDS

Core Staffing
- Heidi Imker, Director (1 FTE)
- Hoa Luong, Associate Director (1 FTE)
- Sandi Caldrone, Research Data Librarian (1 FTE) – starts 2021-10-16
- Colleen Fallaw, Research Programmer (1 FTE)
- Lauren Phegley, Graduate Assistant (0.5 FTE)
- Dena Strong, Senior Information Design Specialist (0.1 FTE volunteer time from Tech Services)

Annual Goals for FY22
- 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 FY20
- 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

Appendix

Winter Snapshot
July 1 - December 31, 2020

RESEARCH DATA SERVICE
University of Illinois at Urbana-Champaign

294
Researchers Served

52
New Data Nudge Subscribers

Service Breakdown
- 48
- 23
- 16
- 14

- Illinois Data Bank Curation (47.5%)
- Consultation (22.8%)
- General Data Question (15.8%)
- Outreach Training Event (13.9%)

"Thanks to your data backup nudge, I started saving things to Box, Google Drive, and emailing myself the documents whenever I make substantial progress with writing or editing."

Illinois Data Bank Discipline Breakdown

- 44
- 10
- 5
- 4
- 1

- Life Sciences (68.8%)
- Social Sciences (15.6%)
- Physical Sciences (6.2%)
- Technology and Engineering (7.8%)
- Arts and Humanities (1.6%)

55,266
Dataset Downloads

421
Horna Munoz
2020-08-01

226
Peyton
2020-11-16

213
Han
2020-08-21

Top New Datasets Downloaded

"I also want to again thank you for the useful suggestions . . . this has made the data more usable for the readers."

Summer Snapshot
January 1 - June 30, 2021

RESEARCH DATA SERVICE
University of Illinois at Urbana-Champaign

335
Researchers Served

62
New Data Nudge Subscribers

Service Breakdown
- 74
- 24
- 9
- 7

- Illinois Data Bank Consultation (64.9%)
- Consultation (21%)
- Outreach Training Event (7.9%)
- General Data Question (6.1%)

"I am working on my dissertation right now and have really appreciated your tips about data accessibility and backups. Thanks to your data backup nudge [...]"

Illinois Data Bank Discipline Breakdown

- 42
- 13
- 12
- 4

- Life Sciences (59.2%)
- Physical Sciences (18.3%)
- Social Sciences (16.9%)
- Technology and Engineering (5.6%)

60,631
Dataset Downloads

747
Teck Yian
2021-03-06

662
Stodola
2021-03-15

383
Cline Center
2021-01-08

Top New Datasets Downloaded

"Thank you so much [...] - I really appreciate your incredible help and tips. I've made the changes and the link appears to be working now. I will go ahead and resubmit my manuscript."

RESEARCH DATA SERVICE
University of Illinois at Urbana-Champaign