Research Data Service Campus Annual Report

2016-2017

Mission Statement

"The Research Data Service (RDS) provides the Illinois research community with the expertise, tools, and infrastructure necessary to manage and steward research data."

2016-2017 In Review

2016-2017 Goals

- Improve our data management outreach and support across campus to be more systematic and consistent
- Establish the Illinois Data Bank, and the RDS's data sharing experts, as a campus asset
- Support campus data-related efforts and create clarity and consistency for researchers by connecting them to appropriate campus resources
- Continue to grow as national leaders who support improvement of the research ecosystem, specifically through data-related work and application of LIS skills

Major Activities and Accomplishments

Now in its third year, the RDS continues to advance as an active and functional program able to serve a wide swath of campus. As the proposal for an RDS was scoped and developed, several core areas were identified as mission-critical for campus, and over the past three years these core areas have grown, including data management plan reviews, data management training/consultation, and data sharing and archiving. With these nascent services established, the past year was devoted refining operations to ensure smooth and high-quality services for campus researchers.

This year also saw the RDS strengthen its interactions with other units within the Library and on campus. These collaborations cannot succeed without a committed and on-going effort to establish, nurture, and coordinate relationships. Although this is time-intensive, these relationships allow both the RDS and the collaborating unit to extend reach and scale in a more effective manner. For example, within the Library, Christie Wiley's position has changed to the Engineering Research Data Services Librarian, and she collaborates with the RDS to create tailored engineering seminars and consultations for engineering departments. This provides the best of both worlds: the RDS's broad expertise in research data management and Christie's familiarity with engineering departments. Similar joint outreach activities have also occurred with Life Sciences Data Services Librarian Sarah Williams and ACES as well as Prairie Research Institute Librarian Susan Braxton and PRI.

Additionally, relationships beyond the Library continue to grow and yield positive (and mutually beneficial) results. The Office of Proposal Development now regularly routes DMP reviews to the RDS. Dena Strong from Technology Services continues to spend a portion of her time with the RDS, a model that on several occasions has enabled researchers access to solutions that otherwise would have likely gone unrealized. Finally, the RDS also served as an initial conduit that connected Research IT efforts with the Scholarly Commons which resulted in pilot funding of a much-needed statistical consulting position. Such "silent" catalysis is easy to lose track of, but the impact of the ability to facilitate such networking should not be underestimated. All units that serve data-related needs must be connected.

In 2016-2017, the RDS's major accomplishments include:

- Through a combination of various marketing strategies, **increased campus interactions over 3-fold** from the previous year

- Increased the number of DMP reviews and **established a seamless review process** that effectively engages library subject specialists and other experts while maintaining confidentiality and speed of review

- **Expanded and refined the Illinois Data Bank** as a service, including comprehensive unit testing and technical documentation, addition of functionality for deposition of large and versioned datasets, and testing, revision, and documentation of curation procedures

- Continued to contribute to **campus-wide data-related efforts** including the Ad Hoc Data Management Task Force, the Illinois Data Science Initiative, and the Research IT Initiative

- One of six major universities to participate in a highly-productive, Sloan-funded planning grant to **establish a cross-institution network of shared data curation expertise**

- Published several **highly-accessed, peer-reviewed articles** describing the work in the Illinois RDS

- Selected by OCLC Research as 1 of 4 international programs, and the only from the US, to participate in **a review of exemplary services** dedicated to Research Data Management

Outreach

This year saw no shortage of interest in data. Phrases like "data science," "data analytics," and "big data" pepper the titles and descriptions of events and initiatives across Illinois' campus and beyond. Research data has always been a "raw material," but the reinvigorated interest stems from the potential to combine, mine, analyze, and visualize data at a scale and interdisciplinarity rarely seen before. The overarching aspiration is to enable research of greater depth and impact, but the results are only as strong as the completeness, the veracity, and the interoperability of the data itself. As data grows its use in ever-expanding ways, so must our emphasis on careful, diligent, and proactive research data management. What is the research potential for data with irregularly formatted fields? Or idiosyncratic vocabularies? Or undefined units? Or unrecorded transformations? Data management requires premeditation, and it is our mission to instigate and enable data management that allows for better research through better data management.

As a new unit on a large campus, we continue to put considerable effort into outreach. In 2016-2017, we presented or co-presented over 90 workshops and seminars across campus that were either wholly or partially dedicated to data management. Through these presentations, we've reached over 1250 students, faculty, administrators, IT professionals, and business office professionals. See Table 1 below for the breakdown of coverage of campus research units (not including Savvy Researcher workshops) and Figure 1 for a breakdown by service type.

	DMP Consultation	Data Management Consultation	Data Sharing Consultation*	Informational Meeting/Seminar
ACES	✓	✓		✓
AHS		✓		✓
Business	✓	✓		
Education	✓			
Engineering	✓	✓	✓	✓
Fine + Applied Arts	✓			
Graduate College				✓
iSchool	✓	✓		✓
LAS	✓	✓	✓	✓
Veterinary Medicine	✓	✓	✓	
University Library	✓	✓		✓
Beckman Institute	✓	✓		✓
NCSA	✓	✓	✓	✓
PRI	✓	✓	✓	✓
IGB	✓	✓		✓
IHSI		✓		

*new this year

Table 1. Coverage by outreach type for colleges (blue) and research institutes (green). Abbreviations: ACES, Agriculture, Consumer, and Environmental Sciences; AHS, Applied Health Sciences; LAS, Liberal Arts and Sciences; NCSA, National Center for Supercomputing Applications; PRI, Prairie Research Institute; IGB, Carl R. Woese Institute for Genomic Biology. **Additional engagement** was also carried out with the Illinois Fire Service Institute, the Office of Sponsored Programs, the Office of Undergraduate Research, the Association of Data Science and Analytics (student organization), Business Process Improvement Shared Service, the Center for Prevention Research and Development, and the Division of Management Information.

DMP consultations continue to steadily increase and for the second consecutive year have more than doubled over the previous year. In 2016-2017 44 DMP reviews were conducted and feedback continues to be very strong (see pg. 6). Additionally, we support researchers in a variety of ways to recognize different working styles and maximize our utility. For example, the Grainger Engineering Library Information Center offers a DMP template for NSF on their home page, and it has been used in hundreds of proposals to date. We also continue to maintain customizations for DMPTool.org, and 66 new DMPs were completed in FY 2016-2017 within the tool.

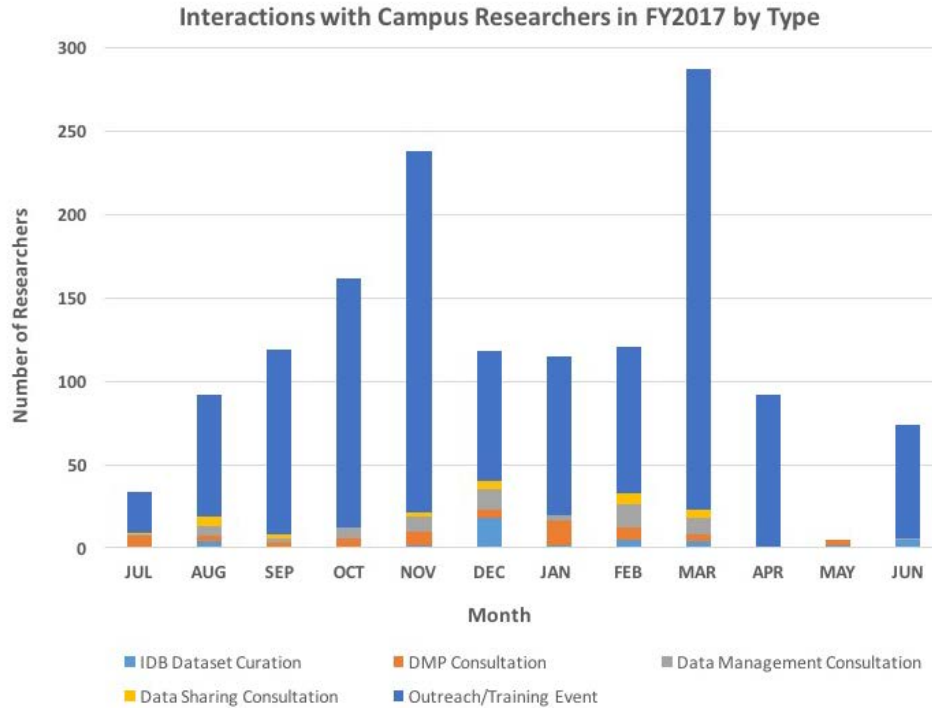Interactions with Campus Researchers in FY2017 by Type

Figure 1. Breakdown of direct interactions by service type for each month in the FY2016-2017 year. Indirect interactions, such as DMPTool or Grainger DMP template uses, are not included.

Illinois Data Bank

In close collaboration with the Library Repository Development Team, the Illinois Data Bank was officially launched for broad campus use on Aug 30, 2016.[1] The interface has been continually refined in response to feedback, particularly regarding search and reporting. In response to requests from researchers, the Illinois Data Bank now offers an API, command-line guidance, and a sample client for uploading files up to 2TB in size. The new upload options have been used in many of the deposits, and the researchers report appreciating the ability to upload files from the command line on the systems they use during the project. To date, we are the only campus data repository that we are aware of using such technology or accommodating datasets of this size.

As of June 30, 2017, the Illinois Data Bank has ingested 56 datasets, 54 of which are published and publically available. The remaining two are under embargo until a later date. While dataset deposits have not been as numerous as expected, the available datasets have already shown substantial use. Download counts for datasets range from 8 – 1501 with a total of over 8,000 downloads across all datasets in just a little of a year of availability. Datasets are also coming in from a wide array of campus units, signaling broad utility of the platform (Figure 2). Integration of the Illinois Data Bank

---

[1] http://web.archive.org/web/20170810220146/https://news.illinois.edu/blog/view/6367/401234

with the Library's Medusa preservation repository allows tracking of datasets as a collection, including monitoring of total storage and file statics to help ensure a well-managed and reliable resource. To date Illinois Data Bank datasets consume just under 2TB of total storage (3 copies). A comprehensive review of the first year of the Illinois Data Bank will be undertaken in early fall of 2017.
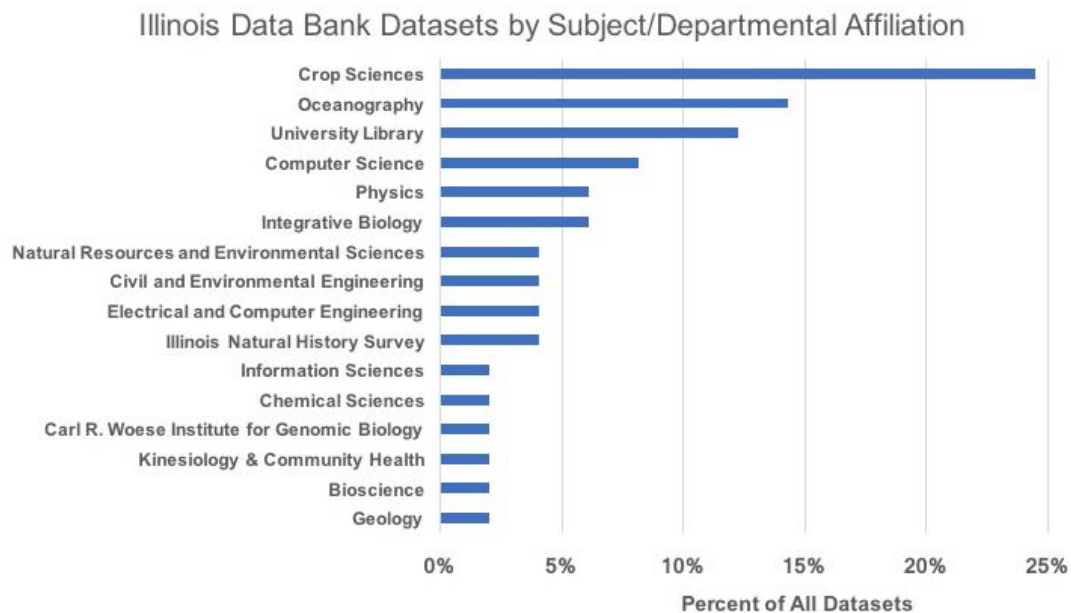
Illinois Data Bank Datasets by Subject/Departmental Affiliation

Subject/Department	Percent of All Datasets
Crop Sciences	~24.5%
Oceanography	~14%
University Library	~12%
Computer Science	~8%
Physics	~6%
Integrative Biology	~6%
Natural Resources and Environmental Sciences	~4%
Civil and Environmental Engineering	~4%
Electrical and Computer Engineering	~4%
Illinois Natural History Survey	~4%
Information Sciences	~2%
Chemical Sciences	~2%
Carl R. Woese Institute for Genomic Biology	~2%
Kinesiology & Community Health	~2%
Bioscience	~2%
Geology	~2%

Figure 2. Data deposits into the Illinois Data Bank show use by a variety of areas across campus.

Contribution to Campus Data Initiatives

The RDS continued to participate in various campus-wide data-related initiatives in 2016-2017. For example, Data Curation Specialist Elizabeth Wickes and CLIR Fellow Qian Zhang contributed instruction to Software Carpentry and Hack-a-thon events. Director Heidi Imker also continued to participate in the Ad Hoc Data Management Task Force, convened by the Provost, VCR, and CIO, by participating in monthly meetings, planning, and focus groups. The Task Force's report was submitted to the Provost in late 2016 and subsequently a standing committee was formed to continue the task force's charge. The Director also participated in early planning for the Illinois Data Science Initiative (iDSI).

Temporary responsibilities at the Grainger Engineering Library Information Center for RDS Director beginning in spring 2017 and loss of a Data Curation Specialist at the end of March left the RDS short-staffed the last quarter of the year. Additionally, the second Data Curation Specialist left in June and the CLIR fellow is done in August. As such, although we reached our goals for 2016-2017, productivity and engagement dropped in all areas for the unit as a whole. Our goal during the staff shortage is to maintain key areas of responsibility; namely review of DMPs, Illinois Data Bank operation and curation, as well as prompt turnaround on any questions or consultations requests that come in.

Feedback from 2016-2017 Services

DMP Consultation

From College of Education:
"Thank you thank you thank you. You're amazing at what you do."

From Department of Anthropology:
"The help of the Data Management Plan review service was great. I rewrote the DMP for a National Science foundation proposal, the assistance was quick and helpful."

From Department of Electrical and Computer Engineering:
"On behalf of my team, please accept my sincerest gratitude for your efforts in helping us with our data management plan with your attention to detail. I have incorporated all the changes you suggested."

From College of Education:
"Outstanding piece of work Elise. Very much appreciated. I have incorporated and massaged your edits/comments. Now just waiting on final review from one of the project team members on the project description. Will let you know if there are other questions."

From Office of Proposal Development:
"Thanks very much for reading this. I'm so glad for your comments. When reading them I again realize how important it is to have you and the other members of RDS as the subject experts for data management! Please pass on my thanks to Susan as well."

Data Management Training and Consultation

From Department of Natural Resources and Environmental Sciences:
"On April 13th, Heidi Imker, Sara Williams, and Christie Wiley presented a workshop on data management to NRES 502 Research Methods in Natural Resources and Environmental Sciences. It was a tremendously valuable addition to the course, and I will definitely continue inviting them in future years. I appreciated the advanced planning they did with me to develop content that would best suit the needs of these particular students, who are all pursuing the degree online and don't get the data management training that (hopefully) comes from working in a faculty member's lab. The feedback from the students was very positive, several of them saying that they had never thought about these issues before and that the information would be extremely helpful to them when they work on their capstone projects to complete the degree."

From Interdisciplinary Health Science Initiative:
"Thank you so much for taking the time today to come to IHSI to train and educate us on data flow. Your training provided us with a clear picture of the project for all members of the team, along with a common language to communicate with each other to more effectively and successfully work together. You are fantastic and thank you again!"

From Department of Molecular and Cell Biology:
"My experience working with the Research Data Service has been most positive. The response to my initial inquiry was prompt and thoughtful, and I have been very impressed with the staff's expertise and willingness to help. The explosion of electronic data in biology, combined with increasing scrutiny by scientific journals and funding agencies, makes it imperative for us to rethink our methods of storing and archiving data. The advice offered by the Research Data Service is practical and customized to my lab's specific needs. I am confident that my lab is now in a better position for scientific accountability."

Illinois Data Bank

From the US Department of Energy:
"It's great to hear that you have a good, designated repository where your researchers' datasets can reside and that those datasets are getting DOIs."

From the Prairie Research Institute:
"This is the first time I have gone through this process and it was really informative. I really liked the website. Everything was intuitive and easy to follow along."

From College of ACES:
"We liked that we could put essentially any information in there that we wanted, including the original handwritten field notes. Having that in the same repository as the genetic data, the phenotypes, the R scripts for data analysis - the paper explains everything from field observations to how to sequence the DNA. We have a broad range of data types all in one repository, which is great. I also really like the ability to link to other datasets; the raw DNA sequence data is many gigabytes and that's stored in NCBI, but we can put a link from this dataset back to NCBI as well."

See also forthcoming *Illinois Data Bank Year in Review*.

2016-2017 Leadership in the Profession

External Presentations

- Imker, H.J. "From Burden to Benefit: Opportunities in New Data Management and Sharing Policies" Cellular Molecular Pharmacology Seminar, UCSF. September 8, 2016

- Anderson, B., Braxton, S., Dunham, E., Imker, H., Rimkus. K. "Should we keep everything forever? Determining the long-term value of research data." Poster presented at: iPRES 2016, 13th International Conference on Digital Preservation, Bern, Switzerland. October 3-6, 2016 http://hdl.handle.net/2142/91659

- Whitmire, A., Imker, H.J., Jones, S. "Analyzing, interpreting, and implementing data management plans" DataONE Webinar Panel. November 8, 2016

- Wickes, E. "Reworking the workshop: designing data management workshops to align with behavioral change models" IASSIST Annual Conference 2017, Lawrence, Kansas. May 24, 2017

- Wickes, E. "Crafting High Impact Tabling Experiences" IASSIST Annual Conference, Lawrence, Kansas. May 26, 2017

- Fallaw, C. Metadata in the middle: Socio-technical infrastructure for Illinois Data Bank DOIs and related metadata" Poster presented at: IASSIST Annual Conference, Lawrence, Kansas. May 26, 2017 http://hdl.handle.net/2142/96212

Written Works

- Fallaw, C. Dunham, E.M., Wickes, E., Strong, D., Stein, A., Zhang, Q., Rimkus, K., Ingram, B., Imker, H.J. 2016 "Overly Honest Data Repository Development" *Code4Lib Journal* 34 http://journal.code4lib.org/articles/11980

- * Burnette, M.H., Williams, S.C., Imker, H.J. 2016. "From Plan to Action: Successful Data Management Plan Implementation in a Multidisciplinary Project" *Journal of eScience Librarianship* 5(1): e1101. http://dx.doi.org/10.7191/jeslib.2016.1101

- Zhang, Q., Imker, H., Li, C., Ludäscher, B., Senseney, M. 2016. "Using a Computational Study of Hydrodynamics in the Wax Lake Delta to Examine Data Sharing Principles" *International Journal of Digital Curation*, 11 (2): 138-155. http://dx.doi.org/10.2218/ijdc.v11i2.433

- * Johnston, L.R., Carlson, J.R., Hswe, P., Hudson-Vitale, C., Imker, H., Kozlowski, W., Olendorf, R.K., Stewart, C. 2017. "Data Curation Network: How Do We Compare? A Snapshot of Six Academic Library Institutions' Data Repository and Curation Services." *Journal of eScience Librarianship* 6(1): e1102. https://doi.org/10.7191/jeslib.2017.1102

* Both have consistently remained on this journal's list of 10 Most Popular Papers (based on average number of full-text downloads per day since the paper was posted).

2017-2018 for the RDS

Core Staffing

- Heidi Imker, Director, May 2014 - Present
- Colleen Fallaw, Research Programmer, April 2015 – Present
- Research Data Specialists (2), TBD
- Qian Zhang, CLIR Postdoc Fellow (w/ iSchool), August 2015 - August 2017
- Hoa Luong, Academic Hourly, May 2017 - Present
- Dena Strong, GSLIS Practicum Student, September 2015 – May 2016; ongoing AP development contribution, May 2016 – Present

Goals for 2017-2018

- Maintain an acceptable base level of service for core areas during periods of staff shortage
- Recruit excellent replacement staff
- Undertake a comprehensive evaluation of all aspects of the RDS to identify strengths, weakness, and provide holistic basis for future directions