

Research Data Service Campus Annual Report

2015-2016

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.”

2015-2016 In Review

2015-2016 Goals

- Closely monitor evolving data policies from agencies and publishers and inform the Illinois research community of the policy impacts on funding and publication.
- Expand discipline coverage and increase campus participation in basic data management training and consultation for improved research efficiency.
- Cultivate a knowledgeable community to enable expert service on management, curation, and stewardship of data.
- Define campus-level and service-level policies that articulate the roles, responsibilities, and commitments to stewardship of research data.
- Prototype, beta-test, and launch infrastructure that supports current and emerging data publication and preservation needs.
- Increase awareness of the RDS through delivery of consistent, clear, frequent, and multi-line communication.

Major Activities and Accomplishments

In just 2 years since the RDS was formally established, the RDS has succeeded in becoming a central, customer-facing service that addresses researchers’ data stewardship and management questions and needs. Primary services offered directly through the RDS include advising on data management planning and implementation, providing data management training, and developing and operating a repository for public access to Illinois research data.

While the RDS’s backbone comes from the strong interaction with the staff of many relevant Library units (including the subject libraries, Scholarly Commons, Preservation, and Archives), the program also partners with campus units including the iSchool (formerly GSLIS), OVCR, Technology Services, and NCSA to enable more ready adoption and use of current services (e.g. Box.com) or establish additional services (e.g. the Active Data Service). The RDS has also come to also serve as a “clearing house” for a wide variety of data-related questions, and we depend on our relationships with and

knowledge of other campus data resources, such as Technology Services Security, the Office of Protection of Human Subjects, and IT Professionals to connect researchers with the help they seek.

In 2015-2016, the RDS's major accomplishments include:

- Improved and increased our DMP review and data management implementation consultations
- Developed and launched the Illinois Data Bank, a repository to provide public access to Illinois' research data
- Established Illinois' position as leaders by authoring several works on research data services and data curation and by securing a Sloan Foundation planning grant in collaboration with 5 peer institutions to develop a data curation network
- Contributed to campus-wide data related organizational efforts including the Ad Hoc Data Management Task Force, the Illinois Data Science Institute white paper, and the Research IT Initiative
- Improved communications, connected with new areas of campus, and reached over 500 Illinois faculty, staff, and students
- Experimented with indirect data management training by initiating or participating in new training activities across campus to support students, including the Data Science Across Disciplines Focal Point, Software Carpentry, and Data & Python Office Hours

Outreach

Data management is fundamental to research, yet in practice it is rarely articulated as such. Researchers at the beginning of their career often pick up habits based on ad hoc practices and local research culture. As a result, data management becomes either an invisible skill or—increasingly—a *visible* incompetency. Because of this we consider it core to our mission to provide training and outreach to improve basic data management skills of researchers across campus.

In 2015-2016, we presented or co-presented over 40 workshops and seminars across campus that were either wholly or partially dedicated to data management. Through these presentations, we've reached over 500 students, faculty, administrators, IT professionals, and business office professionals across 12 colleges and 5 research institutes. See Table 1 below for the breakdown of outreach coverage (not including Savvy Researcher workshops). This year we've only missed coverage in the Colleges of Labor and Employment Relations and Media, and we expanded coverage overall. Our DMP consultations have more than doubled from 13 in 2014-2015 to 28 in 2015-2016. We've also reviewed FOAs that require much more complex DMPs, in one case 6 different types of data management/sharing plans, including 1) Data Sharing Plan, 2) Data Management and Quality Control Plan, 3) Data and Safety Monitoring Plan, 4) Genomic Data Sharing Plan, 5) Sharing Model Organisms, and 6) Software Sharing Plan. While this is a situation where NIH seems to be increasing their data management expectations, on the other end of the spectrum several agencies have not yet formalized policies that make DMPs a requirement, despite indicating they would do so in their implementation

plans. Back channel communications indicate that some agencies are in limbo regarding enforcement of their public access plan (e.g. see <http://bit.ly/2aMHZfZ>). The highly variable situation suggests that at least for the foreseeable future, DMPs will continue to be a necessary but irregular component of grant applications.

	DMP Creation Consultation	Data Management Consultation	Informational Meeting/Seminar
<i>ACES</i>	✓	✓	✓
<i>AHS</i>	✓	✓	
<i>Business</i>		✓	
<i>Education</i>		✓	
<i>Engineering</i>	✓	✓	✓
<i>Fine + Applied Arts</i>	✓	✓	
<i>Graduate College</i>			✓
<i>iSchool</i>	✓	✓	✓
<i>LAS</i>	✓	✓	✓
<i>Law</i>		✓	
<i>School of Social Work</i>	✓		
<i>Veterinary Medicine</i>	✓		
<i>University Library</i>		✓	✓
<i>Beckman Institute</i>			✓
<i>IGB</i>	✓	✓	✓
<i>iSEE</i>	✓		
<i>NCSA</i>	✓	✓	✓
<i>PRI</i>	✓		✓

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; IGB, Carl R. Woese Institute for Genomic Biology; iSEE, Institute for Sustainability, Energy, and Environment; NCSA, National Center for Supercomputing Applications; PRI, Prairie Research Institute.

As we build awareness for the Research Data Service on campus, we are acutely aware of the fact that many people with data management needs do not recognize, or are unable to articulate, the impact of poor data management on their research. In the same way that other preemptive or preventative activities, such as estate planning, nutrition, or physical fitness, are difficult to prioritize, data management rarely escalates to being an imperative and is most visible once something has already gone wrong. Because of this blind spot, we've worked to be creative in our outreach efforts and experiment with interjecting data management training in areas that are seemingly tangential, such as the Data Science Across Disciplines Focal Point, Software Carpentry, and Data & Python Office Hours. We found these new training activities to be successful because each of them reflect an area where we are seeing increasing interest from researchers across campus, i.e., Data Science and research

programming. This strategy has allowed us to meet researchers where their immediate interests lie while also providing data management guidance along the way. Additionally, these efforts have kept our own internal understandings of research challenges and perspectives current and honest. These initiatives serve as a form of “loss leaders,” and if done judiciously and strategically, are of benefit to our overall mission to enable and support data stewardship and management of research data. We will continue to experiment in the coming years, as we are able and as these efforts are effective.

Illinois Data Bank Launches

Planning and hiring were the focus of 2014-2015, and with the on-boarding of a developer and two data curators, a major effort this year was development and launch of the Illinois Data Bank (<https://databank.illinois.edu/>). The Illinois Data Bank is a free, self-serve publishing platform that centralizes, preserves, and provides persistent and reliable access to Illinois research data. In addition to development of the technical infrastructure and web application, substantial effort was undertaken to ensure that at launch all aspects of operating the Illinois Data Bank were considered and addressed so that both the system and our associated services would run professionally. As such, critical components of the Illinois Data Bank that also required development included policies and workflows, adoption and adaption of metadata schema, and iterative user testing.

The Illinois Data Bank is optimized specifically for data, which is material that essentially supports or contributes to research findings that someone else might want to use to re-analyze or extend those findings (e.g. text or csv files, chemical or biological spectra, data from text mining or topic modeling, etc.). IDEALS will still serve as the Library's repository for reporting of research and scholarship as it is optimized for publications, reports, presentations, posters, etc.

Illinois Data Bank Datasets:

- can be linked to related materials, such as articles, theses, code, and other datasets
- can include files of any format and up to 15 GB via self-deposit (larger via special arrangement)
- can be deposited for immediate release or temporarily embargoed
- receive a stable, unique identifier (DOI) for persistent access and ease of citation
- are registered in a central, world-wide catalog for better discovery
- are professionally managed and curated
- are preserved at the University Library for a minimum of 5 years

During the summer of 2016 we reached out to researchers on campus with known data sharing needs to seed the repository in a “soft launch” phase. Through these efforts 7 datasets were publically available as of June 30, 2016. Reception was positive and this period was helpful in identification of bugs, testing of communication and curation workflows, and to hone messaging. The Illinois Data Bank will be broadly announced to campus initially on August 28, 2016 via eWeek with subsequent marketing via the News Bureau and other avenues.

Contribution to Campus Data Initiatives

In addition to our core services, the RDS team contributed to several data-related activities on campus in the 2015-2016 year. The focus on data as an asset cuts across nearly all sociality sectors and similarly the University's interest in data is not restricted to research. The Provost, CIO, and VCR established an Ad Hoc Data Management Task Force to assess the current campus resources for data management across all sectors of the University. The Task Force is structured as a group of six individuals, of which the RDS director is member, from diverse areas of campus. This group's primary responsibility is to lead the Task Force activities, and a much larger group of data area representatives (e.g. admissions data, financial data, library data, etc.) has been named and will be engaged as needed. The Task Force will be in commission through the 2016-2017 and may go longer.

In another area, as the Year of Cyberinfrastructure wrapped up in December of 2015, the steps were immediately taken to act on the tremendous amount of feedback gathered from campus researchers. The new Deputy CIO for Research John Towns and his office undertook an aggressive process to develop a comprehensive, balanced, and well-scoped proposal to develop and support Research IT on campus. The RDS participated in Research IT planning meetings, documentation of needs, and proposal writing and review. We expect that in the coming year campus will also take steps to address the policy, infrastructure, and support needs of researchers working with sensitive data, such as classified, PII, and FERPA data. As the Interdisciplinary Health Sciences Initiative (IHSI) scoped implementation of REDCap, a secure web platform for managing surveys and databases which is commonly used by medical schools (<https://projectredcap.org/>), the RDS has worked with them utilizing our library resources through the Scholarly Commons and the Savvy Researcher Workshop series to lead, host, and advertise REDCap training for researchers.

Additionally, as a group of campus researchers and administrators developed a white paper to form a new campus research institute, the Illinois Data Science Institute (iDSI), the RDS director served on the Steering Committee and articulated how the RDS and the Illinois Data Bank could support such an initiative.

Finally, the RDS director participated in the Center for Advanced Study's Spring Symposium by serving as a panel member to discuss the data science needs of researchers at Illinois (with Brunner, Wade, Turk, and Underwood).

Feedback from 2015-2016 Services

Data Management Training and Consultation

From College of LAS:

"I met with Elizabeth [RDS Data Curation Specialist] for help with a (to me, at least!) rather complicated R project. I have no coding background and have frequently been frustrated at how much background knowledge people expect me to have when I ask questions, but Elizabeth was able to meet me where I was and help me understand from there. She worked

with me patiently for over 2 hours (!!) and we were able to take a tangled mess of data from hundreds of separate Excel workbooks and turn them into one R dataset. She was very knowledgeable and was able to quickly find out anything she didn't know off the top of her head. She also helped me install some helpful programs for my job [and] left me with lots of resources with which to learn more about R - including an offer for specialized workshops based on what student researchers need! I have spent a lot of time screaming and fighting with R and after visiting Elizabeth at the Research Data Group [sic], I actually feel like we accomplished something in this program!”

From Applied Health Sciences:

I first contacted RDS regarding best practices for subject identifiers and data sharing across multiple data collection sites. Heidi and Elise [RDS Director and Data Curation Specialist] were very helpful and during our conversation I broached another topic—being asked to share brain imaging data for an about-to-be published study. They were helpful in working out the ramifications of this and based on their prompting, I contacted the IRB office. Which in turn led to re-reading of the original protocol, and we had not explicitly asked our patient to consent to sharing of individual (although de-identified) brain imaging data. The journal agreed with this and dropped their request for sharing individual-level brain imaging data. Over the past year, we’ve continued to discuss the new practice of data sharing of individual-level patient data.”

From the College of ACES:

“Your talks are always a highpoint of our seminar series among students and faculty. Thanks once again for your presentation.”

From Colleges of Engineering and ACES:

“Graduate students are often the primary people handling data associated with university research. Unfortunately, many of us are not educated in how to properly handle this data – both on the data management and data analysis side. Implementing good data management practices from the beginning can help safeguard against mistakes in data analysis (i.e. being able to re-trace steps or having a clean version of the data to check against). Currently, RDS is the major force on campus pushing to educate graduate students about how to properly manage data. Having an RDS presence in the Focal Point enabled us not only to educate 15 graduate students about the importance of properly handling data and how to implement good practices in their work, but also brought the message that data management is a critical and foundational aspect of any data-oriented pursuit.”

DMP Consultation

From College of Veterinary Medicine:

“I’ve found the DMP review service helpful in identifying areas that were not as clear as I had thought, and providing suggestions for improvement. As a new faculty member, it is

incredibly useful to know that experienced eyes are looking through this section of my proposals before submission. I've even had compliments from reviewers on the quality of my DMPs."

From the College of Engineering:

"The Research Data Service looked over my NSF Data Management Plan and provided a prompt, point-by-point review, including suggesting draft text and best practices for data archiving. As data management and archiving become increasingly important part of research and grant applications, the Research Data Service is a unique and valuable campus resource."

Illinois Data Bank

From the iSchool:

"Thank you so much for the quick turnaround, Elise [Data Curation Specialist]! Wow, that was service!! "

From Computer Science:

"I've uploaded my dataset to the website. It was extremely easy to use, and I'm very grateful for the resource. Thank you!"

See also upcoming Press Release from the Illinois News Bureau.

2015-2016 Leadership in the Profession

The Research Data Service takes seriously its role in extending Illinois' commitment to data stewardship by forming a community of knowledge around the management, curation, and stewardship of data, both on the Illinois campus and beyond. We have worked diligently to develop this knowledge and visibility throughout the 2015-2016 year. For example, we provided detailed comments to a Request for Feedback from DataCite on their Dublin Core Application Profile and RDF Mapping Documents on behalf of the RDS and a group of content discovery experts in the Library. To build our local knowledge, in July of 2015 the University of Minnesota's Lisa Johnston held a half-day workshop on Data Curation at the University Library. The 22 participants joined from across the library as well as the iSchool, the College of Engineering, and NCSA. Additional leadership accomplishments are as follows:

External Presentations

- "Research Data & the Art of Simultaneously Meeting Needs and Picking Battles" Imker, H.J. University of Iowa, November 2015 (invited)
- "Capitalizing on Research Data: Management, Dissemination, and Archiving" Imker, H.J. University of Iowa, November 2015 (invited)

Written Works

- “Overlooked and Overrated Data Sharing: Why so many scientists are confused and/or dismissive” Imker, H.J. in *Curating Research Data*, edited by Lisa R. Johnston. ACRL, 2016 *in press*
- “A Participant Agreement for Minting DOIs for Data Not in a Repository” Braxton, S.M., Anderson, B.G., Burnette, M.H., Habing, T.G., Mischo, W.H., Shreeves, S.L., Williams, S.C., Imker, H.J. in *Curating Research Data*, edited by Lisa R. Johnston. ACRL, *in press* 2016
- “Pre-Metadata Counseling: Putting the DataCite relationType Attribute into Action” Dunham, E.M., Wickes, E.D., Stein, A., Fallaw, M.C., Imker, H.J. in *Curating Research Data*, edited by Lisa R. Johnston. ACRL, *in press* 2016

Blog Posts

- “Meta-Service: How Illinois’ Research Data Service Benefits the Library” Imker, H.J. *eScience Community Blog*, December 2015
<http://esciencecommunity.umassmed.edu/2015/12/14/meta-service-how-illinois-research-data-service-benefits-the-library/>
- “Processing Digital Research Data” Dunham, E.M. *bloggERS!: The blog of SAA’s Electronic Records Section*, June 2016 (invited)
<https://web.archive.org/web/20160601152817/https://saaers.wordpress.com/2016/05/11/processing-digital-research-data/>
- “How to approach selecting a license for data release” Wickes, E.D. *Data Carpentry Blog*, June, 2016 (invited)
<http://web.archive.org/web/20160816013106/http://www.datacarpentry.org/blog/data-licensing/>

Awards

- Grad College Focal Point Grant, Rando, H.M., Byrne, D.M., Stein, A., Imker, H.J., August 2015 – August 2016
 - “Data Science Across Disciplines” graduate level class and seminar series
- Alfred P. Sloan Foundation Planning Grant, Johnston, L.R., Carlson, J. Hswe, P., Imker, H.J., Kozlowski, W.A., Stewart, C., Hudson-Vitale, C. April 2016 – April 2017
 - Plan for development of a network of shared data curation expertise across multiple universities
- RDA/US Data Share Fellowship from the Research Data Alliance (RDA), Dunham, E.M., May 2016 – Nov 2018
 - Develop an organization-wide curation strategy for the records of the RDA

Feedback from 2015-2016 Leadership

From the Illinois Mortenson Center for International Library Programs:

“On behalf of everyone at the Mortenson Center, I wanted to thank you for the fantastic data management session that you gave to our associates this summer! They really enjoyed it and learned a lot from the hands-on activities (I did too). Thank you for sharing your expertise and time with us! I also wanted to let you know that your session and the following visit to the iSchool inspired [a student] to begin seriously considering coming here for a PhD to study the most effective ways to manage and share data [in his region] - he’s already started his application.”

From the Interdisciplinary Health Sciences Initiative:

“RDS provided expert knowledge to the hiring decision for an open position in database development. This advisement was crucial to the success of the hire, and thereby data flow and the success of data projects within IHSI. RDS is a terrific resource on campus and we look forward to partnering with them again soon.”

From the University of Iowa Library:

“Learning about the Illinois RDS experience and their service model helped the UI Libraries and the campus in general to situate ourselves in the RDS landscape and make realistic and effective plans for the future to provide the best possible service for our researchers.”

2016-2017 for the RDS

Core Staffing

- Heidi Imker, Director, May 2014 - Present
- Colleen Fallaw, Research Programmer, April 2015 - Present
- Elise Dunham, Data Curation Specialist, May 2015 - Present
- Elizabeth Wickes, Data Curation Specialist, May 2015 - Present
- Qian Zhang, CLIR Postdoc Fellow (w/ GSLIS), August 2015-Present
- Dena Strong, GSLIS Practicum Student, September 2015 – May 2016; ongoing AP development contribution, May 2016 – present

Goals for 2016-2017

- 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