

Research Data Service (RDS) Annual Report

July 1st 2015 – June 30th 2016

Submitted by: Heidi Imker (imker@illinois.edu)

I Unit Narrative

1. 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.

	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.

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.

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. We believe 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 for the 2015-2016 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 from the moment of 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 adaptation of a 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 for us to identify bugs, test communication and curation workflows, and 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).

2. Major challenges

As a fledging program, the RDS has several challenges related to expectations. This was somewhat heightened during the year as the Research IT efforts began and there was some interpretation of the RDS being under the preview of Research IT. This situation has been skirted.

We were perhaps too successful in bringing in highly competent and skilled staff for the RDS. We struggle with boundaries. The RDS supports researchers on campus in their data management needs, and our work towards that must take priority. Working with the AUL for Research, RDS policies will be set to outline how, when, and to what extent out-of-scope activities will be considered.

Another challenge is the Illinois Data Bank retaining a disproportionate amount of our focus. It's a tangible and comfortable area for us as it aligns better with LIS skills and the Library's existing IR work. Deposits are also measurable therefore metrics and physiological rewards are naturally built into data publishing. Data management itself is more nebulous and can feel like "self help." However, the need for excellent data management practices spans across all phases of research

and all types of research. Not only does data management training have more immediate and far-reaching benefits to researchers, sharing of high quality data will continue to suffer from poor management unless practices are improved. We are actively working to creatively address data management by thinking about it through the lens of personal information management and behavioral psychology.

The highly variable issues brought to our attention have reinforced the need for such a service such as the RDS. While we field questions the best we can, a challenge is that we fail at times in providing researchers with a consistent set of services, policies, and practices, because our campus' services, policies, and practices *are not consistent*. Take a subset of questions posed:

- *"Subject: cost of archiving petabyte-scale data?"*
- *"RDS may also provide some more general database components such as authentication and access control services."*
- *"I was also wondering if there is any university-wide practice regarding data sharing?"*
- *"Chris just came by and said you were the one to talk to about getting larger group folders in Box."*
- *"What steps should I take to de-identify the individual raw data to prepare them for publication?"*
- *"Right now we are using a shared web page to upload and download information from different groups. I'd like to see what ideas you folks have about how that can be done."*
- *"He indicated that it amounted to about 3 file cabinet drawers of material."*
- *"I am not sure who is the appropriate person or unit to contact, and thought I would start with you."*
- *"This is almost perfect timing as we are having these very discussions right now within our unit, as we are trying to figure out ways to get people more access to our data."*

Lastly, as in last year, I'll reiterate that our ever-loving challenge will always be awareness on campus.

3. Significant changes

The RDS developed and launched the Illinois Data Bank, in consultation with other library and campus units.

4. Contribution to Library-wide programs

Information Services:

We are the university's resource for research data management, preservation, and archiving best practices. We also are the university's resource for funding agency and publisher research data sharing policies. In addition to email contact, we have a

Wordpress website and a promotional flier that we distribute at events.

Instructional Services:

The RDS provides instructional services on data management, federal agency policy developments, DMP requirements and creation, research data repositories, data publication, and post-publication review of data. These are either done as part of the Scholarly Commons's Savvy Researcher Workshop series or in campus venues external to the Library.

Scholarly Communication:

We regularly promote the Scholarly Commons during RDS presentations. In FY 2015-2016, this has resulted in impressions on > 500 people. The RDS has also participated in Savvy Researcher workshops for Data Management, the content of which was both created and presented collaboratively with members of the Scholarly Commons and other members of the Library. Data Curator Elizabeth Wickes also consulted on data-oriented books to add to the SC library to assist in computational data processing and management consultations. The Director also participates in the Data Services Committee and the RDS curator Elizabeth Wickes provides Data and Python Open Hours through the Scholarly Commons. We value our partnership with the Scholarly Commons and will continue to work closely with the unit.

Assessment:

No assessments were undertaken this year specifically by the RDS.

Collections:

Although the RDS does not contribute to traditional collections, the concerted addition of research data into Library digital repository collections is a new direction for the Library.

Digital Content Creation:

Similarly, although the RDS does not contribute directly to traditional digital content creation, our explicit training of researchers in data management is an effort arm researchers with LIS best practices, albeit at the point of *external* digital content creation. This is beneficial to the Library itself in that it raises awareness of the Library's expertise in the digital age.

Staff Training and Development:

In July of 2015, we hosted University of Minnesota's Lisa Johnson at Illinois. She held a half-day workshop on Data Curation with garnered library attendance from both functional units and subject areas. She also met with several groups and shared the data repository and curation work they are doing at UMN.

Additionally, monthly seminars during the academic year through the RDS Interest Group are open to everyone. As mentioned above, travel funds to data-related conferences for FY 2015-2016 were provided by the AUL for Research. Finally, the RDS Committee also provides a mechanism to learn about research data and engage in data-related activities at the Library.

Public Engagement:

None

Advancement:

None

5. Progress made on Unit Annual Goals for FY15

- Closely monitor evolving data policies from agencies and publishers and inform the Illinois research community of the policy impacts on funding and publication.
 - Agencies have very notably lagged in implementation of policies to require DMPs, but we continue to actively monitor the situation.
- Expand discipline coverage and increase campus participation in basic data management training and consultation for improved research efficiency.
 - Coverage has improved with only 2 colleges without engagement as of the end of FY2015-2016. We continue to receive excellent feedback on seminars.
- Cultivate a knowledgeable community to enable expert service on management, curation, and stewardship of data.
 - Although all core RDS staff is new to data management and curation in this form and at this scale (e.g. entire university), our collective learning curve has been steep.
 - We've also engaged our fellow librarians in training with efforts such as Lisa Johnston's Data Curation workshop in July 2015.
 - See also professional accomplishments in the RDS Campus Annual Report.
- Define campus-level and service-level policies that articulate the roles, responsibilities, and commitments to stewardship of research data.
 - Understandably, campus administration is not eager to establish new campus policies. We will continue to engage with groups and prompt for clarity as situations arise. The Illinois Data Bank policies set a good foundation for the importance of data-centric policies.
- 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.
 - This has improved, but it is still a challenge. See goals for FY16.

6. Annual Goals for FY16

- 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

7. Number of GAs in FY15

- 0

8. GA funding

- n/a

9. GA major responsibilities and contributions

- n/a

II Statistical Profile

1. Facilities

- User seating counts (if applicable)
 - Not Applicable
- Number of hours open to the public per week (if applicable)
 - Not Applicable

2. Personnel

- Heidi Imker, Director (Associate Professor; 1 FTE)
- Colleen Fallaw, Research Programmer (Academic Professional; 1 FTE) April 2015-Present
- Elise Dunham, Data Curation Specialist (Academic Professional; 1 FTE) May 2015-Present
- Elizabeth Wickes, Data Curation Specialist (Academic Professional; 1 FTE) 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

3. User Services

- Gate Count (as reported during FY15 Sweeps Week).
 - Not Applicable
- Circulation (from Voyager circulation reports)
 - Not Applicable
- Reference interactions (from DeskTracker)
 - DeskTracker: 110
- Presentations (from the Instructional Statistics database)
 - Number of presentations to groups: 40
 - Number of participants in group presentations: >500

4. Other statistics

Web Statistics

From July 1, 2015 to June 30, 2016 the RDS website (researchdataservice@illinois.edu) had 4,300 users (new and returning) and total of 12,686 page views (numbers not corrected for bots or spam).

Librarian Training

In addition to the Staff Training and Development (4. *vide supra*), the RDS has also been supportive of GLSIS programs as they reach out to LIS students and international audiences, presenting at the following 6 classes/workshops:

- GLSIS Academic Libraries Seminar (w/ Maria Bonn)
- Mortensen Center International Librarians
- Chinese Scholar Librarian Visits (July 2015 and June 2016)