

# Research Data Service Campus Annual Report

2014-2015

## 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.”*

## 2014-2015 In Review

### Goals

- Proactive outreach to increase awareness of agency and publisher data policies and to improve data management on campus.
- Scoping and development of infrastructure for private storage and public access to research data.

### Major Activities and Accomplishments

After several years of planning and coordination at the campus level, the Research Data Service (RDS) was formally adopted as a **campus strategic priority** through inclusion in the 2013-2016 Campus Strategic Plan.<sup>1</sup> On-going, committed funding was provided by the Offices of the Provost and Vice Chancellor for Research, and the program has been a point of accomplishment **highlighted by campus administrators**, for example, through inclusion in the Chancellor’s July address to the Board of Trustees.<sup>2</sup> The RDS’s home is in the University Library, and the program’s backbone comes from the strong interaction with the many relevant Library units (including the subject Libraries, Scholarly Commons, Preservation, and Archives) and a **cadre of talented and invested Librarians and Library staff**. Uniquely, the program also maintains **active partnerships with campus units** including GSLIS, OVCR, Tech Services, and NCSA. For kick-off in FY 2014-2015, the RDS’s major accomplishments include:

- **hired 3** full-time employees
- basic **communications** developed for and distributed through a Wordpress website and informational flier
- outreach to **over 600** Illinois faculty, staff, and students
- with the Library EZID Pilot sub-committee, began **minting DOIs for Illinois data**
- productive collaboration established with the Library’s Repository Management Team to begin developing a **dedicated research data repository**
- productive collaboration established with the Scholarly Commons and Librarians for Savvy Researcher **Data Management Workshops**
- productive collaboration established with GSLIS through hiring of a joint **CLIR post-doctoral fellow** (to start in August of 2016) and support of the **Beckman repository task**
- productive collaboration established with NCSA for development of a new **Active Data Storage** service
- **fostered champions** through working with researchers on data deposit and data repository scoping
- award of a **Focal Point grant** from the Graduate College for Data Science Across Disciplines program for FY 2015-2016 (with Metadata Librarian Ayla Stein and graduate students Halie Rando and Diana Bryne)

As we are a brand new service, our two primary goals for 2014-2015 were highly interdependent. Outreach is nothing unless there is something for people to use and infrastructure is nothing unless people use it.

During 2014-2015, the RDS reached >600 campus members through workshops, talks, and consultations. This works complemented current Library data services which itself fielded >300 data-related reference questions. Through exploratory partnerships and collaborations the RDS also played a crucial role in several currently funded research projects and participated in four funding proposals.

In addition to presenting or co-presenting **25 data management workshops**, we gave **15 presentations** across campus. Through these presentations, we've reached students, faculty, administrators, IT professionals, and business office professionals across **8 colleges and 4 research institutes**. See Table 1 below for the breakdown of outreach coverage (not including Savvy Researcher workshops). We are missing coverage in Business, General Studies, Fine and Applied Arts, Labor and Employment Relations, Law, Media, School of Social Work, and iSEE. The RDS has been consulted either directly or via Subject Librarians for **13 DMP reviews**.

	DMP Creation Consultation	Data Management Consultation	Informational Meeting/Seminar
<i>ACES</i>	✓	✓	✓
<i>AHS</i>		✓	✓
<i>Engineering</i>		✓	✓
<i>Education</i>	✓		✓
<i>GSLIS</i>			✓
<i>Graduate College</i>			✓
<i>LAS</i>	✓	✓	✓
<i>Veterinary Medicine</i>	✓		
<i>Beckman Institute</i>		✓	✓
<i>IGB</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; GSLIS, Graduate School for Library and Information Sciences; LAS, Liberal Arts and Sciences; IGB, Carl R. Woese Institute for Genomic Biology; NCSA, National Center for Supercomputing Applications; PRI, Prairie Research Institute.

Toward our goals of developing infrastructure, we hired a repository developer who is dedicated solely to the development of a research data repository. Two data curators were also hired who have duties to curate data and assist with repository design. We've created **30 DataCite DOIs** for Illinois data, which will aid in discovery and reuse. A large collection of datasets for a CS faculty member was ingested into IDEALS, assigned DOIs, and to-date the collection has had **548 downloads**. We also worked with NCSA throughout the year to scope and establish the Active Data Storage (ADS) service, which provides scalable, rentable storage for research data at an affordable cost. As of summer 2015, the ADS is testing the service with early adopters.

## Feedback from 2014-2015 Services

### *Data Management Training*

Examples from post-presentation survey of Postdoc Bootcamp through the **Graduate College**:

Q1: What aspects of the workshop were most useful?

"Short, relevant summary of good data management practices."

"Documentation - I did not do that for my PhD! Storage/Backup."

"All of it. I only wish there was more time."

### *DMP Consultation*

From the **College of Education**:

"Very impressive service and timely response. Very much appreciated. I have cut and pasted the statement you provided and exported it in a pdf file. So I am hoping this should be it. Thanks a lot for your help and introduction to this new system."

From the **Carl R Woese Institute for Genomic Biology**:

"...using the University Research Data Service for the distribution of derived datasets is a good idea. I wasn't enthusiastic about setting up a Web site for the project with links to results, this provides an easier and more sustainable alternative."

### *DMP Implementation*

From **College of ACES**:

"Dr. Imker and her colleagues at the Research Data Service have provided several useful resources over the past year for our use in carrying out two collaborative, interdisciplinary research projects. Services included several in-person consultations in which concrete suggestions were offered to trouble shooting data management and documentation challenges. In-person meetings were followed up with detailed summaries of topics discussed. Services and recommendations were individualized to specific needs of our projects and included referrals regarding computing resources and computer programming on campus, links to data documentation examples and protocols, specific suggestions regarding workflow management and versioning conventions, and recommendations for file format conversion for long-term usability (e.g., converting wmv video files to mp4 format)."

### *Data Publication and DOI Creation*

From **College of Engineering**:

"Research Data Services allowed us to host a large number of data files, with potential to be used in multiple publications, permanently on a reliable server. Moreover, the service was able to seamlessly handle large quantities of data. This service provided us with permanent digital object identifiers (DOI), and this capability helped us satisfy requirements of our publication venue. The system seemed flexible enough to accommodate our special needs. Hosting research data *persistently* is a challenge, and many of the other similar services are costly if the data size is as large as ours; the research data service at UI helped us overcome these challenges with minimal effort from our side."

## *Data Documentation and Repository Consultation*

From the **Beckman Institute**:

"RDS assisted the "INSIGHT" project in a study of data repositories. The "INSIGHT" project is a \$13.1M research contract funded by the Intelligence Advanced Research Projects Activity (IARPA), through their Strengthening Human Adaptive Reasoning and Problem Solving (SHARP) Program, to Professor Aron Barbey and team at the UIUC Beckman Institute. INSIGHT seeks to develop and assess a variety of cognitive training interventions to improve human adaptive reasoning and problem solving; neuroimaging with magnetic resonance imaging (MRI) technologies is an important part of the study. IARPA requested the INSIGHT team to lead a short-term study of data repositories suitable for SHARP, with a particular focus on the integration of behavioral and neuroimaging data. RDS worked closely with the INSIGHT project manager to identify, hire, train, and supervise students in the Graduate School of Library and Information Science; develop technical criteria for evaluating existing data repositories; educate the INSIGHT team on existing data and metadata standards; develop use cases, information workflows, and a data dictionary; and contribute to the SHARP data repository final report. RDS also participated in the March 16, 2015 site visit of the IARPA program manager and his staff. "RDS was a key player in the successful accomplishment of the SHARP data repository study," said Patricia Jones, INSIGHT project manager and Associate Director for Research at the Beckman Institute. "RDS made substantive intellectual and organizational contributions to the SHARP data repository task. IARPA is already impressed by the cognitive training, kinesiology, and neuroscience research at Illinois, and they are now even more impressed by the amazing technical support and leadership provided by RDS."

## 2014-2015 Leadership in the Profession

We gave presentations that described the Illinois Research Data Service program and Library data services in **10 different venues**. These included 2 on campus for GLSIS students, 2 on campus for the Mortensen Center for International Librarians, and 2 virtual presentations in Hong Kong and Beijing GSLIS Data Curation Workshops. Additionally, we were invited to present at the Association of Research Libraries (ARL) Fall Membership meeting,<sup>3</sup> the Nebraska Big Data Conference,<sup>4</sup> National Information Standards Organization (NISO) Data Management Webinar,<sup>5</sup> and the Standards for Reporting Enzymology Data (STRENDA) conference.<sup>6</sup>

## Feedback from 2014-2015 Leadership

### *Data Curation Workshop*

From the **Graduate School for Library and Information Science**:

"... special two-hour-long guest lecture on Research Data Services at Illinois played key role in making the [Hong Kong Data Curation Workshop] both popular and impactful. Heidi's talk resonated especially well with attendees who are on the cusp of creating their own RDS-type services. They found Heidi's insights very helpful and her presentation easy to follow and comprehend. Post-talk discussions were filled with interesting ideas and enthusiastic planning. The participants of the Institute left the meeting with a call for one or more follow-up meetings on data curation to help move things forward in Hong Kong, Macau, China, Taiwan and Singapore. I know that they would like the opportunity to have more contact and communications with Heidi at future events. Heidi, and Illinois Library, are justly being seen as a leaders in the DC and RDS domains."

## 2015-2016 for the RDS

At the start of the second year, the RDS is fully staffed with high qualified and energized professionals. The RDS staff will continue to work closely to complement the expertise and commitment already present at the University Library and to engage campus partners in strategic projects to achieve the second year goals as follows.

### 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 - December 2015

### Goals

- Closely monitor evolving data policies from agencies and publishers and inform the Illinois research community of the policy impacts on funding and publication.<sup>7</sup>
- 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.

### Footnotes

1. University of Illinois at Urbana-Champaign Campus Strategic Plan 2013-2016 [http://go.illinois.edu/Strategic\\_Plan](http://go.illinois.edu/Strategic_Plan)
2. Wise, Phyllis. "Dash Board Indicators." Meeting of the Board of Trustees. Chicago, Illinois: July 23, 2015. [http://go.illinois.edu/Wise\\_BoT\\_July\\_2015](http://go.illinois.edu/Wise_BoT_July_2015)
3. Imker, Heidi J. Panelist, "Data Management Programs: Getting to Concrete Results." 165<sup>th</sup> ARL Membership Meeting. Washington, DC: October, 8 2014
4. Imker, Heidi J. "Big Data in the Big Ten: Research Data at Universities." Nebraska Big Data Conference. Lincoln, Nebraska: November 7, 2014
5. Imker, Heidi J. "Building Capacity: Leveraging existing library networks to take on research data" Scientific Data Management: Caring for Your Institution and its Intellectual Wealth. NISO Virtual Conference. February 18, 2014
6. Imker, Heidi J. "(National) Data Services." 10th STRENDA Meeting. Glasgow Scotland: September 24, 2014.
7. Most federal agency implementation plans indicate October 2015 as the effective date for new data management policies and requirements.