

Research Data Service (RDS) Annual Report

July 1st 2014 – June 30th 2015

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

I Unit Narrative

1. 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.¹ On-going, committed funding was provided by the Offices of the Provost, the Chancellor, 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.² 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)

2. Major challenges

As a new program, the RDS has several challenges related to expectations emanating from both external and internal directions. As a campus-facing program that exists through partnerships with several other campus units (above), not surprisingly, the RDS's scope and responsibilities have been subject to interpretation. We strive to manage those expectations and clearly articulate our mission, but we also strive to meet, or at least facilitate the meeting of, the campus' data-related needs which are frequently wide, uncharted, and highly idiosyncratic. Occasionally, questions aimed at the RDS are entirely misdirected which simply require us to put the researcher in touch with the correct contact. However, more often than not, needs require collective engagement from multiple people within the Library or units on campus. For example, inquiry about the level, and to whom, can human subject data be shared. Such inquiries require our knowledge of data sharing policies (or nascent policies) from funding agencies, emerging discipline practices, and erratic publisher expectations. However, such a question also requires the knowledge of the Institutional Review Board, as the consent given at the time of the study is essential to a decision. Other challenges have included inquiries about data on physical media, obsolete media, data retention, and requests to preserve so-called Big Data (terabyte to petabyte scale) from NCSA, Beckman, PRI, and others. The RDS will continue to work with partners to identify, and likewise develop, solutions collaboratively whether they are service, policy, or infrastructure orientated.

Internal to the Library, we also struggle with a broader understanding of the RDS's role on campus and within the Library. The RDS was meant to be a collaborative program rooted at the Library. As such as the RDS is fortunate to have unusual financial support, administrative buy-in, and campus visibility. This is an extraordinarily unique situation and one that is widely envied by other academic institutions. Likewise, there is exceptional interest in research data from Librarians and Library staff at Illinois. However, it is also somewhat awkward in that necessarily our attention must remain fixed on external researcher needs which, paradoxically, has led some who are more internally focused to wonder how the RDS contributes to the Library. We can counter this partially by being vocal about our mission and priorities and by also providing professional development opportunities for those interested in research data. While this has been accomplished through travel support to data-related conferences from the AUL for Research and engagement with the eResearch Implementation Committee (currently the RDS Committee) and the RDS Interest Group, more specific and immersive opportunities are required. In FY 2015-2016, we will aim to increase these opportunities.

As attested by the lack of robust platforms, it's clear that the repository technology required to ingest, preserve, and provide access to research data will be challenging. Research data is diverse and there are untold numbers of file types and associated proprietary and/or homegrown code and software required to render re-use of the data a possibility. Metadata schemes inevitably confound both Librarians and researchers. Likewise, research data is frequently

complex meaning a dataset may contain hundreds, thousands, or even tens of thousands of files, it may be dynamically produced, or representation may require an interactive interface. Research data may be also be oversized. The RDS and Repository Management Team have worked closely in FY 2014-2015 to identify a strategy to develop a research data-specific repository and, critically, *integrate* content into the Medusa preservation system. This will allow us flexibility in the system build but also provide broader utilization and support for Medusa.

Finally, our ever-loving challenge will be awareness on campus. Researchers on campus do not and will not come to think of the Library for stewardship of research data on their own. This an intellectual leap that must be actively triggered. Furthermore, as Alan Cabbage, VP of University Relations at Northwestern, said recently at the CIC CLI Annual conference, universities are communicating with a “parade” as students, staff, and faculty come on to and off of campus. Therefore, for the foreseeable future, we must continue to be highly active and visible on campus. It bears repeating, we cannot wait for researchers to come to us. However, a major concern is that we scale back on outreach due to lack of bandwidth. This is an unfortunate coping mechanism often employed, but the RDS will work to resist this as it inevitably means that only smaller, frequently random, sub-sets of users are served. Ultimately, it also hampers the ability both articulate needs and actualize a strategic vision for the unit.

Although these issues are wildly challenging, they in fact heavily underscore the need for the Research Data Service on campus. It also provides a view into the potential future of the RDS, one in which the program becomes a widely recognized and relied-upon resource for research data on campus. The diversity and breadth of the issues likewise attests to the appropriateness of the RDS’s home at the Library. Not only is it a neutral, well-connected, collaborative environment, the Library has the best expertise to address new community and agency expectations around data access and preservation.

3. Significant changes

Staffing changes

- In FY 2014-2015, 3 FTE staff were hired.

Funding agency policy changes

- In FY 2014-2015, 15 federal agencies released implementation plans for public access. The plans outline the policies which are intended to be implemented in FY 2015-2016. In October of 2015, 8 agencies will be requiring Data Management Plans (DMPs) and another 2 will begin doing so in January of 2016. This will bring the total to 13 and will include all of the major agencies funding Illinois research.

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:

At the end of every talk given by the RDS, we present a slide that promotes the Scholarly Commons. In FY 2014-2015, this has resulted in impressions on > 600 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. The Director also participates in the Data Services Committee and the RDS graduate student, Bill Pooler, participated in GIS Day planning with the Scholarly Commons GIS specialist. 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. However, the RDS was included in the Year of Cyberinfrastructure survey that was administered to faculty and students in Spring of 2015 (data unpublished but forthcoming). In that survey, 15.9% respondents (n=327) were aware of the RDS. Considering the RDS had been in operation for less than a year at the time of the survey, this is fairly good. In comparison, 16.5% of respondents were aware of Public Affairs' WebTools and 15.9% were aware of the Maker Lab; both resources have been on campus longer than the RDS. However, this does indicate, as laid out in the challenges section, that we have significant work to do to reach the campus more broadly.

Collections:

Although the RDS does not contribute to traditional collections, the concerted addition of research data into Library digital repository collections is a new and expanding 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 to 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:

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 2014-2015 were provided by the AUL for Research. Finally, the RDS Committee (formerly the eResearch Implementation Committee prior to March 2015) 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

Our first year goals were minimal but highly inter-dependent:

Outreach

- Promote awareness of agency and publisher policies on campus
- Improve data management

Infrastructure

- Lay ground work for providing Illinois-hosted public access to research datasets
- Scope and establish rentable "active" (working) data storage with NCSA

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.

6. 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.³
- 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.

7. Number of GAs in FY15

- 1 @ 50%

8. GA funding source was from State Funds.

9. GA major responsibilities and contributions

William Pooler served as a 50% GA for the 2014-2015 academic year. Bill helped with various aspects of outreach (website, instruction materials) and collaborative work (webform for the Active Data Storage project with NCSA). He also developed a preliminary tool assessment protocol to evaluate data-related resources.

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
- William Pooler, RDS GA (50%) May 2014-May 2015

The positions that Fallaw, Dunham, and Wickes have filled were all posted and hired within FY 2014-2015. A CLIR-DLF postdoctoral fellow, a joint 50/50 position with GSLIS's CIRSS was also posted. Qian Zhang was interviewed in April and extended an offer. She will join us in August of 2015.

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: n/a for FY 2014-2015
 - Manual Statics: approx 58 one-on-one interactions with unique individuals via email
- Presentations (from the Instructional Statistics database)
 - Number of presentations to groups: 27
 - Number of participants in group presentations: 556

4. Other statistics

Web Statistics

We began tracking statistics for the Wordpress site (researchdataservice@illinois.edu) in late 2014. From December 18, 2014 to June 30, 2015, the RDS website had 1,136 users (new and returning) and total of 4,477 pageviews.

DMPTool

292 DMPs created | 350 unique users (from launch to date)

Librarian Training

The RDS has also been supportive of GSLIS programs as they reach out to LIS students and international audiences, presenting at the following 6 classes/workshops:

- GLSIS Science Librarians Leep Class (w/ Linda Smith, Peg Burnette, and Kelli Trei)
- GLSIS Scholarly Communications Seminar (w/ Maria Bonn)
- GLSIS Hong Kong Data Curation Workshop (virtual; with CIRSS)
- GLSIS Beijing Data Curation Workshop (virtual; with CIRSS)
- Mortensen Center International Librarians
- Mortensen Center OCLC Fellows

Footnotes

1. University of Illinois at Urbana-Champaign Campus Strategic Plan 2013-2016
http://strategicplan.illinois.edu/documents/2013-2016_Campus_Strategic_Plan.pdf
2. Wise, Phyllis. 'Dash Board Indictors'. 2015. Presentation to the Board of Trustees.
http://go.illinois.edu/Wise_BoT_July_2015
3. Most implementation plans indicate October 2015 as the effective date for new data management policies and requirements.