Universities: Diverse By Design

Recent Illinois Campus Highlights of Published Research

- Proceedings of the National Academy of Science
- Psychological Science
- Nature (2)
- Journal of Interpersonal Violence

http://illinois.edu/lb/iList/1171
Universities: Diverse By Design

RNAseq, gut microbiota, xylanolytic bacteria, hemicellulose, human nutrition, stars, divided attention, dual-task performance, distracted driving, aquarium, avian influenza, captive wildlife, emergency management, outbreak response, zoo, zoonotic disease, bullying, sexual harassment, adolescents, biosynthesis, enzyme mechanisms, X-ray crystallography
Universities: Diverse By Design

RNAseq, gut microbiota, xylanolytic bacteria, hemicellulose, human nutrition, stars, divided attention, dual-task performance, distracted driving, aquarium, avian influenza, captive wildlife, emergency management, outbreak response, zoo, zoonotic disease, bullying, sexual harassment, adolescents, biosynthesis, enzyme mechanisms, X-ray crystallography
Illinois Research Data Service (RDS)

Provide the Illinois research community with the expertise, tools, and infrastructure necessary to manage and steward research data.
Illinois RDS Timeline

- ARL eScience Institute
- Yr of Data Stewardship
- RDS Proposal
- RDS in Campus Strategic Plan
- RDS Hiring Initiated

2011  2012  2013  2014
Illinois RDS Timeline

2011
- Data Services Program
- ARL eScience Institute

2012
- Yr of Data Stewardship
- eResearch Task Force

2013
- RDS Proposal
- Data Management Outreach
- eResearch Implementation Committee

2014
- RDS in Campus Strategic Plan
- RDS Hiring Initiated
Top-Down Drivers

OSTP: INCREASING ACCESS TO THE RESULTS OF FEDERALLY FUNDED SCIENTIFIC RESEARCH

Key points:

- free public access to research data
- data management planning
- citation of datasets
- private sector collaborations
- education and workforce development
- assessment of data preservation needs
- development of sustainable solutions

“requiring researchers to better account for and manage the digital data resulting from federally funded scientific research”

http://www.whitehouse.gov/blog/2013/02/22/expanding-public-access-results-federally-funded-research
Bottom-Up Drivers
Bottom-Up Drivers
Current RDS Emphases

**Outreach**

• Improving local data stewardship has the most universal impact

**Infrastructure**

• Rentable Active Data Storage
• Data Deposit for Public Access (Repository)
RDS Expertise: Data Librarians

Current

- Scholarly Commons
  - Data Services
  - Digital Humanities
  - Statistical Software

Future

- Hiring – Data Curation Specialists
RDS Expertise: Workshops

• Intro to Data Management and (Data) Publication
• Improving Your Data Organization and Security Strategies
• Using Data Ethically
• How to Find and Learn About Appropriate Statistical Methodologies in Your Field
• Metadata for Research Data: How to Understand Your Data and Find It Later
• Basics of Research Data Publication and Sharing
RDS Tools: Data Mgmt Planning

- Data Management Plans (DMPs)
  - What data produced, how will you make it accessible, how will you preserve it, what restrictions are there on access

- **NSF** - 2011
- **NIH** - 2003 (>500K)
- **DOE** - 2014

Gates Foundation
“The foundation will require that data underlying the published research results be immediately accessible and open. This too is subject to the transition period and a 12-month embargo may be applied.” Nov 2014
RDS Tools: Data Mgmt Planning

- online wizard for creating a ready-to-use DMPs
- help text, links to resources, and Illinois-specific suggestions
- login with Illinois NetID and password to automatically create an account
- create, save, and edit plans
RDS Tools: Workflows

- Directory Structures
- Workflow Diagraming
- Box
- GitHub
- SharePoint
- etc.
RDS Tools: Minting DataCite DOIs

Digital Object Identifier (DOI)

A string that specifies a unique object within the DOI system.

- doi:10.13012/J83X84KP
- http://dx.doi.org/10.13012/J83X84KP

DataCite
Helping you to find, access, and reuse data

EZID
Long-term identifiers made easy
RDS Tools: Minting DataCite DOIs

- makes research data more discoverable and accessible
- enables reuse and verification
- allows the impact of data to be tracked (via citations)
- enables the ability to recognize and reward data producers
RDS Tools: Minting DataCite DOIs

**DOI resolver**
Resolve a DOI string (e.g. 10.5284/1000164) by entering it below:

Submit

**Metadata Search**

Search
RDS Infrastructure: Active Storage

- Partnership with National Center for Supercomputing Applications (NCSA)
- Competitively priced/TB-year
- Minimum 1 TB-month increments
- Accessible via:
  - network file share
  - file transfer services (ftp, etc.)
- Anticipated access to campus computing cluster
- Anticipated disaster recovery option
RDS Infrastructure: Data Deposit

IDEALS Institutional repository
- Can accommodate files <2GB, static, flat
- Intended long-term preservation

Illinois Data Bank (in planning)
- Broader deposit capability
- Medium term commitment to near-line access
- Build out access and preservation capabilities based on collection appraisal
Data Collection Management

- High volume, unstructured or variably structured, fragile, variable documentation, volatile

- Collection policies and criteria developed with Archives and Preservation Units
Data Preservation

- Bit-level preservation
- Format conversation
- Software bundling
- Loss of function
- Loss of value
- Loss of validity
Data Appraisal

Prioritize effort based on:

• Access frequency
• Citation frequency
• Associated publication citation frequency
• Impact/magnitude of study/scholar
• Uniqueness
• Level of description
• Preservability
• Current relevance
Takeaways

• No easy answers for the extreme and complex
• No quick solutions are sustainable or even desirable
• Except the unglamorous
  o Data Management
  o Storage
  o Access
  o Preservation