Apr 6th, 10:30 AM

Data Repositories: The Answer that Actually Came with a Question

Lisa Johnston
University of Minnesota - Twin Cities

Follow this and additional works at: http://escholarship.umassmed.edu/escience_symposium

Part of the Scholarly Communication Commons

This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 4.0 License.
Data Repositories

The answer that actually came with a question

Lisa Johnston
Research Data Management/Curation Lead,
University of Minnesota
What are data repositories?

“A repository (also referred to as a data repository or digital data repository) is a searchable and queryable interfacing entity that is able to store, manage, maintain and curate Data/Digital Objects. A repository is a managed location (destination, directory or ‘bucket’) where digital data objects are registered, permanently stored, made accessible and retrievable, and curated. Repositories preserve, manage, and provide access to many types of digital material in a variety of formats. Materials in online repositories are curated to enable search, discovery, and reuse. There must be sufficient control for the digital material to be authentic, reliable, accessible and usable on a continuing basis. (p3, footnote 2).”

Citation: The RDA Data Foundations and Terminology working group has a growing dictionary of data related terms that is searchable at http://smw-rda.esc.rzg.mpg.de/index.php/Main_Page
Why are data repositories useful?

- Governments have open data initiatives
  - Democratize research results
  - Release the potential of valuable/$$ data

- Federal/private funders increasingly require data sharing
  - Public access
  - Return on $$ investment ⇒ do new research

- Journals have data sharing policies
  - Increase transparency
  - Facilitate reproducibility

- Researcher/disciplinary culture shift in digital age
  - Ease of sharing ⇒ greater potential for reuse
  - Citation impact, reputation building
What kinds of data repositories exist?

Disciplinary Data Repositories

And 1516 more listings at http://www.re3data.org/
What kinds of data repositories exist?

General Data Repositories

- Figshare
- Zenodo
- Mendeley Data Beta
- Dataverse Project
- Open Science Framework
- http://openICPSR.org
What kinds of data repositories exist?

Institutional Data Repositories

- Dataverse
- Digital Commons
- DSpace
- EPrints
- Fedora
- HubZero
- Hydra
- Islandora
What kinds of data repositories exist?

Institutional Data Repositories

- Dataverse
- Digital Commons
- DSpace
- EPrints
- Fedora
- HubZero
- Hydra
- Islandora
What kinds of data repositories exist?

Institutional Data Repositories

Running...
- Dataverse
- Digital Commons
- DSpace
- EPrints
- Fedora
- HubZero
- Hydra
- Islandora
What kinds of data repositories exist?

Institutional Data Repositories

Running...
- Dataverse
- Digital Commons
- DSpace
- **EPrints**
- Fedora
- HubZero
- Hydra
- Islandora
What kinds of data repositories exist?

Institutional Data Repositories

Running...
- Dataverse
- Digital Commons
- DSpace
- EPrints
- Fedora
- HubZero
- Hydra
- Islandora
What kinds of data repositories exist?

Institutional Data Repositories

Running...

- Dataverse
- Digital Commons
- DSpace
- EPrints
- Fedora
- HubZero
- Hydra
- Islandora
What kinds of data repositories exist?

Institutional Data Repositories

Running...
- Dataverse
- Digital Commons
- DSpace
- EPrints
- Fedora
- HubZero
- Hydra
- Islandora
What kinds of data repositories exist?

Institutional Data Repositories

Running...
- Dataverse
- Digital Commons
- DSpace
- EPrints
- Fedora
- HubZero
- Hydra
- Islandora
How do I build an institutional data repository?

Blueprint for Institutional Data Repositories

Nancy McGovern’s Three-Legged Stool of building an organization's digital preservation program…

- technological infrastructure
- organizational infrastructure
- resources framework

Citation: http://www.iassistdata.org/downloads/2006/f2_mcgovern.pdf
How do I build an institutional data repository?

Blueprint for Institutional Data Repositories

Lisa’s three-legged stool for building a data repository (basically the same…)

- technological infrastructure
- organizational infrastructure
- resources framework

Image: https://placeshakers.files.wordpress.com/2010/08/stool-parody.jpg
DRUM
http://z.umn.edu/drum

Launched Nov 2014
Available to U of M researchers and provides:
- Open access
- Curation services
- Permanent identifiers (DOI)
- Flexible Licenses
- File download analytics
- Preservation
How do I build my institutional data repositories?

Organizational Infrastructure

- Libraries DM+C Initiative 2014-2015 with dedicated resources and a 19-person team
  - Business model
  - Policies
  - Services

- Existing libraries digital preservation framework

- Campus: New research data management policy

- Marketing: Official launch in March 2015

Citation: “The Supporting Documentation for Implementing the Data Repository for the University of Minnesota (DRUM): A Business Model, Functional Requirements, and Metadata Schema” at http://hdl.handle.net/11299/171761.
How do I build my institutional data repositories?

Organizational Infrastructure

- Libraries DM+C Initiative 2014-2015 with dedicated resources and a 19-person team
  - Business model
  - Policies
  - Services

- Existing libraries digital preservation framework

- Campus: New research data management policy

- Marketing: Official launch in March 2015

Citation: “The Supporting Documentation for Implementing the Data Repository for the University of Minnesota (DRUM): A Business Model, Functional Requirements, and Metadata Schema” at http://hdl.handle.net/11299/171761.
How do I build my institutional data repositories?

Organizational Infrastructure

- Libraries DM+C Initiative 2014-2015 with dedicated resources and a 19-person team
  - Business model
  - Policies
  - Services
- Existing libraries digital preservation framework
- Campus: New research data management policy
- Marketing: Official launch in March 2015

Citation: “The Supporting Documentation for Implementing the Data Repository for the University of Minnesota (DRUM): A Business Model, Functional Requirements, and Metadata Schema” at http://hdl.handle.net/11299/171761.
How do I build my institutional data repositories?

Organizational Infrastructure

- Libraries DM+C Initiative 2014-2015 with dedicated resources and a 19-person team
  - Business model
  - Policies
  - Services
- Existing libraries digital preservation framework
- Campus: New research data management policy
- Marketing: Official launch in March 2015

Citation: https://www.lib.umn.edu/dp/digital-preservation-framework
How do I build my institutional data repositories?

Organizational Infrastructure

- Libraries DM+C Initiative 2014-2015 with dedicated resources and a 19-person team
  - Business model
  - Policies
  - Services

- Existing libraries digital preservation framework

- Campus: New research data management policy

- Marketing: Official launch in March 2015

Citation: https://policy.umn.edu/research/researchdata
How do I build my institutional data repositories?

Organizational Infrastructure

- Libraries DM+C Initiative 2014-2015 with dedicated resources and a 19-person team
  - Business model
  - Policies
  - Services
- Existing libraries digital preservation framework
- Campus: New research data management policy
- Marketing: Official launch in March 2015

Image: http://continuum.umn.edu/email/2015/drum/
Case Study: Building the Data Repo for the U of M

Technical Infrastructure

- DRUM part of existing IR (DSpace):
  - Metadata schema
  - Collection home page
  - Record view

- Meet federal funding requirements

- Curation procedures
  - Sensitive Data
  - Readme Template
  - Transform File Formats

IR: https://conservancy.umn.edu/
Case Study: Building the Data Repo for the U of M

Technical Infrastructure

- DRUM part of existing IR (DSpace):
  - Metadata schema
  - Collection home page
  - Record view

- Meet federal funding requirements

- Curation procedures
  - Sensitive Data
  - Readme Template
  - Transform File Formats

Case Study: Building the Data Repo for the U of M

Technical Infrastructure

- DRUM part of existing IR (DSpace):
  - Metadata schema
  - Collection home page
  - Record view

- Meet federal funding requirements

- Curation procedures
  - Sensitive Data
  - Readme Template
  - Transform File Formats

Case Study: Building the Data Repo for the U of M

Technical Infrastructure

- DRUM part of existing IR (DSpace):
  - Metadata schema
  - Collection home page
  - Record view

- Meet federal funding requirements

- Curation procedures
  - Sensitive Data
  - Readme Template
  - Transform File Formats

Case Study: Building the Data Repo for the U of M

Technical Infrastructure

- DRUM part of existing IR (DSpace):
  - Metadata schema
  - Collection home page
  - Record view

- Meet federal funding requirements

- Curation procedures
  - Sensitive Data
  - Readme Template
  - Transform File Formats

IR: https://conservancy.umn.edu/
Case Study: Building the Data Repo for the U of M

Technical Infrastructure

- DRUM part of existing IR (DSpace):
  - Metadata schema
  - Collection home page
  - Record view

- Meet federal funding requirements

- Curation procedures
  - Sensitive Data
  - Readme Template
  - Transform File Formats

IR: https://conservancy.umn.edu/
Case Study: Building the Data Repo for the U of M

Technical Infrastructure

- DRUM part of existing IR (DSpace):
  - Metadata schema
  - Collection home page
  - Record view

- Meet federal funding requirements

- Curation procedures
  - Sensitive Data
  - Readme Template
  - Transform File Formats

Handout: https://www.lib.umn.edu/datamanagement/sensitive
Case Study: Building the Data Repo for the U of M

Technical Infrastructure

- DRUM part of existing IR (DSpace):
  - Metadata schema
  - Collection home page
  - Record view

- Meet federal funding requirements

- Curation procedures
  - Sensitive Data
  - Readme Template
  - Transform File Formats

This readme.txt file was generated on <YYYYMMDD> by <Name>

GENERAL INFORMATION

1. Title of Dataset:
2. File Information:
   A. Filename:
   B. Short description:
   C. Filename:
   D. Short description
   E. Filename:
   F. Short description:
   G. If data set includes multiple files related to one another, include relationship here:

3. Principal Investigator Contact Information
   A. Name:
   B. Institution:
   C. Address:
   D. Email:

4. Associate or Co-investigator Contact Information
   A. Name:
   B. Institution:
   C. Address:
   D. Email:

Download: z.umn.edu/readme
Case Study: Building the Data Repo for the U of M

Technical Infrastructure

- DRUM part of existing IR (DSpace):
  - Metadata schema
  - Collection home page
  - Record view

- Meet federal funding requirements

- Curation procedures
  - Sensitive Data
  - Readme Template
  - Transform File Formats

Excel Archival Tool

Download: Excel Archival Tool Software (Github download) http://z.umn.edu/exceltool
Case Study: Building the Data Repo for the U of M

Resources Infrastructure

- Funding for DataCite DOIs
- Staffing Model
- Training for new curation staff
Case Study: Building the Data Repo for the U of M

Resources Infrastructure

- Funding for DataCite DOIs
- Staffing Model
- Training for new curation staff

Case Study: Building the Data Repo for the U of M

Resources Infrastructure

- Funding for DataCite DOIs
- Staffing Model
- Training for new curation staff

DMC Scientific Data Curator Training Guide 2015

18 Training Activities
Conclusions

● Greater exposure of library services on campus

● Data repositories open up new conversations and opportunities (e.g., big data, campus networks, policies)

● Rewarding work to partner with researchers to publish their data sets.
Thanks and Questions

Keep building those stools...