Managing Your Research Data at IUB

Stacy Konkiel
E-Science Librarian
skonkiel@indiana.edu

9/21/2012
Overview

- Introduction
- Funding Agency Mandates
- IU Resources
  - Preparing your data
  - Storage (Filesystem & Archival)
  - Access
  - Preservation
- Q&A
Introduction

Stacy Konkiel, E-Science Librarian
skonkiel@indiana.edu

• Background in Scientific Publishing & Digital Research Preservation
• Scholarly Communication Department, Libraries
• Coordinate IU Data Management Service
Introduction

IUB Data Management Service

• Preparing your data
• Storage (HIPAA-compliant)
• Preservation
• Access
• Data Management Plan consultations
  • Lab-based
  • Funder required
Funding Agency Mandates:
Overview

Data Management Plans (DMPs)

Data Access ("Open Data")
Funding Agency Mandates: Overview

Funding Agency Mandates: Overview

• Organization = Easier Work
• Replicated Data = Safe(r) Data
  • Digital data is more fragile than analog data
• Open Data = More Citations (Piwowar et al, 2010)
• Accelerate discovery

(Houston, 2011)
Funding Agency Mandates:
Overview

ROI for funding agencies

- Fund more research
- More eyes on data
- Research dollars go further
- More results found
Funding Agency Mandates:
IU Solutions

IUB Data Management Service
• DMP Consultations
• Cyberinfrastructure for data management

Proposal Development Services (OVPR)
• Finding funding
• Grant writing assistance

Office of Research Administration (ORA)
• Grant compliance

Responsible Conduct of Research (RCR) & DM
• Poynter Center for Research Ethics
• RCR classes via ORA
Funding Agency Mandates:
What to address

Types of Data


http://bit.ly/OGt87o
Funding Agency Mandates: What to address

Formats

- Community standards
- Commonly used
- Interoperable
Funding Agency Mandates:
What to address

Metadata

Ex. Dublin Core

Title: Christina’s World
Creator: Wyeth, Andrew
Date: 1948
Subject: Painting, American Artists
Format: Painting
    Tempera on panel
    32 1/4 x 47 3/4"
Funding Agency Mandates: What to address

Metadata

Ex. Darwin Core

ScientificNameID: Bolborhynchus aymara
DecimalLatitude: -23.8169444
DecimalLongitude: -65.4847222
Year: 2005
IdentificationID: 52356
Preparations: tissue, round skin, other

Image via http://aymary.wz.cz/members.htm
Funding Agency Mandates: What to address

Privacy & Confidentiality
• Interrelated with issue of access
• Subject to IRB regulations?

Reuse & Distribution Policies
• Also subject to IRB regulations
• How do you want others to…
  Use your data? (Non-commercial only?)
  Credit your work?
  Share your work with others?
Funding Agency Mandates: IU Solutions

Determining Types of Data, Formats, Metadata
• E-Science Librarian
• Metadata Analyst & other library metadata specialists

Privacy, Confidentiality, Reuse & Distribution Policies
• IU Research Policies
• Legal Counsel
Funding Agency Mandates: What to address

Data Storage

• 3 copies in separate locations
• Yes
  Stable, short and long term storage for life of project+
  Attention to sensitive data issues
  Departmental/University tech support
• No
  Local storage (on lab computers, personal laptops, thumb drives)
Funding Agency Mandates: IU Solutions

Data Storage ("filesystem" storage)

- **Research File System** (UI TS)
  - 100 GB
  - Individual and team accounts
  - HIPAA-aligned, not encrypted
  - Regularly backed up
  - Small files that are updated frequently
Funding Agency Mandates: IU Solutions

Data Storage ("filesystem" storage)

- Data Capacitor (UITS)
  - Big Red, Quarry, Mason
  - 10 TB
  - Not backed up, regularly cleared
  - Large files
Funding Agency Mandates:
IU Solutions

Data Storage ("filesystem" storage)
• **XSEDE** (multi-institution)
  Visualizations
  Software
  "Science Gateways"
  Big Data
  For life of project (some persistent storage in SGs)
Funding Agency Mandates: IU Solutions

Data Storage ("filesystem" storage)

- **IU CTSI “Service Cores”** (IU, PU, ND)
  Subject-specific (social and physical sciences)
  Storage space & capabilities vary
  Computing and project planning

- **Advanced Information Technology Core** (UITS)
  Electronic personal health information (ePHI)
  HIPAA-protected information
  Biomedical research
  Paid service
Funding Agency Mandates: IU Solutions

Data Storage ("archival" storage)

- **IU Box**
  - 50 GB
  - Syncs with local storage
  - Permanent

- **Scholarly Data Archive (SDA)**
  - Where you back up RFS and Data Capacitor files
  - 50 PB
  - Permanent tape storage
Funding Agency Mandates: What to address

Access

• What will you share?
• How will you share it?
• What provisions will be made to share data securely?
Funding Agency Mandates:
IU Solutions

Access

- **IU Box**
  - Role-based access
  - Share with anyone
  - No HIPAA, FERPA, FISMA, GLBA, ePHI data

- **IU Alfresco Share**
  - Role-based access
  - Multi-institution
  - No ePHI data allowed
Funding Agency Mandates: IU Solutions

Access

• **IUScholarWorks Repository**
  - Open Access repository
  - Indexed by Google, et al
  - No sensitive data unless de-identified, etc

• **WinSCP**
  - SFTP
  - Access via username and password
  - Desktop client
Funding Agency Mandates: What to address

Preservation

• As formats become obsolete, who will migrate data? Think .BMP to .JPG to .PNG

• As media degrades, who will transfer data?

• In the future, who will know enough about your research (or even your field) to work with your data?
  Preservation metadata = Rosetta Stone
Funding Agency Mandates:
IU Solutions

Preservation

• **IUScholarWorks Repository**
  Storage in perpetuity
  Format migration
  Interface for OA data stored on SDA
  File integrity validation, embargoes
  Managed by IU Libraries

• **Scholarly Data Archive** (SDA)
  Permanent, perpetual tape storage
  Mirroring & redundancy
### LEADII Vortex2 Archive Dataset

<table>
<thead>
<tr>
<th>Title</th>
<th>LEADII Vortex2 Archive Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author</td>
<td>Pale, Beth</td>
</tr>
<tr>
<td>Date</td>
<td>2011-02-25</td>
</tr>
<tr>
<td>Date(s)</td>
<td>May 01, 2010 - June 15, 2010</td>
</tr>
<tr>
<td>Geographic / Spatial</td>
<td>Oklahoma</td>
</tr>
<tr>
<td>Information</td>
<td></td>
</tr>
<tr>
<td>Methodology</td>
<td>Executing 214 workflows, using 109,568 CPU hours and generating 215 GB of data and over 9,100 2D products, LEAD II produced short-term, highly accurate weather forecasts each morning and made the results instantly available to researchers in the field using mobile phones and a field viewer.</td>
</tr>
<tr>
<td>File Information</td>
<td>Each 28GB zipped tarball contains at the top level a directory &quot;out&quot; which consists of a number of directories, each containing the results of a single forecast. The directory, out/forecast_20100501090000EDT_run001, for instance is for 9:00 May 1st, 2010 (Eastern Daylight Time). Within a forecast directory are approximately 9 data products (sometimes less than 9 data products because visualization generation for some variables failed). Each data product is in its own file, and described by a metadata file of the same name but ending in .metadata.xml.</td>
</tr>
<tr>
<td>Location</td>
<td><a href="http://hdl.handle.net/2022/12987">http://hdl.handle.net/2022/12987</a></td>
</tr>
<tr>
<td>Type</td>
<td>Dataset</td>
</tr>
<tr>
<td>Embargo release date</td>
<td>2212-04-11</td>
</tr>
</tbody>
</table>

### External Files

- [http://purl.dlib.indiana.edu/iusw/data/2022/12987/LEADII-Vortex2-dataset.tar.qzaa](http://purl.dlib.indiana.edu/iusw/data/2022/12987/LEADII-Vortex2-dataset.tar.qzaa)
- [http://purl.dlib.indiana.edu/iusw/data/2022/12987/LEADII-Vortex2-dataset.tar.qzab](http://purl.dlib.indiana.edu/iusw/data/2022/12987/LEADII-Vortex2-dataset.tar.qzab)
- [http://purl.dlib.indiana.edu/iusw/data/2022/12987/LEADII-Vortex2-dataset.tar.qzad](http://purl.dlib.indiana.edu/iusw/data/2022/12987/LEADII-Vortex2-dataset.tar.qzad)
- [http://purl.dlib.indiana.edu/iusw/data/2022/12987/LEADII-Vortex2-dataset.tar.qzae](http://purl.dlib.indiana.edu/iusw/data/2022/12987/LEADII-Vortex2-dataset.tar.qzae)
Funding Agency Mandates: 
IU Solutions 

Preservation 
• Librarian Consultations 
  Best home for research (IU and non-IU) 
  Preservation and metadata consultations
Q&A

Download this presentation at:

> http://hdl.handle.net/2022/14682 <

Contact

Stacy Konkiel

skonkiel@indiana.edu

IU Data Management Service: http://scholarworks.iu.edu/data
IU Libraries Data Management Guide: http://www.libraries.iub.edu/data