Johns Hopkins University Data Management Services: Reviewing Our First Year

David Fearon
Betsy Gunia

IASSIST June 7, 2012
Our service development: a social science analogy

Developing new services for data management support required understanding social factors, not just technical.

• Services emerging from data curation “social movement”
• Investigating
  – University culture for outreach
  – Data practices for consulting & curation support
  – Trends in public funding, data sharing & curation
Data curation “social movement”

<table>
<thead>
<tr>
<th>Emergence</th>
<th>Recognizing the need for curation &amp; access</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• eScience: digital data production outpaces sharing &amp; preservation</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coalescence</th>
<th>Policy &amp; requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Jan 2011: NSF Data Management Plan requirement</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Institutionalization</th>
<th>Building Services &amp; Infrastructure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• 2009: NSF DataNet funds Data Conservancy at JHU</td>
</tr>
</tbody>
</table>
Data Management Services at Academic Libraries

- Academic libraries in early stages of expanding services to support data curation & management

Data Management Planning Support

- Provide tools for planning
- Consult with researchers
- Assist in data archiving & sharing

JHU Data Management Services
## JHU Data Management Services

<table>
<thead>
<tr>
<th>Services</th>
<th>Preparing Data Management Plans</th>
<th>Archiving Research Data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Preparation</strong></td>
<td>Assistance with data management plans</td>
<td>Assist in implementing plan, preserve &amp; share research through JHU Data Archive</td>
</tr>
<tr>
<td><strong>Cost</strong></td>
<td>Funded by JHU schools using our service</td>
<td>2% of grant’s total direct cost</td>
</tr>
</tbody>
</table>

- **JHU Data Archive features (in development):**
  - Data Conservancy architecture: Accepts data from all disciplines
  - Feature-level discovery: Interoperable data access
  - Persistent identifiers: Supports long-term preservation

[DataConservancy.org](http://DataConservancy.org)
Investigating JHU Cultures to introduce Data Management Services

**Outreach Goals**
- Building awareness of services

**Approach**
- Finding appropriate communication channels

- Researchers
  - Research Projects Admin
  - Dept. Admins
  - Targeted workshops

- Deans
- Library Colleagues
Investigating JHU Cultures to introduce Data Management Services

Outreach Goals

Building awareness of services

Demonstrating Relevance of services

Approach

Finding appropriate communication channels

Consulting and customized support

Do DMPs matter to NSF?

Can’t I copy a template?

Why archive my data?

Investigating data practices

Investigating data practices
Understanding Data Practices

Data Management Plans

Data Archiving

Data Practices
Consulting on data management plans

- Final Plans
- Contact from Principal Investigator (PI)
- Background Research
- JHU DMS Comments on Plan
- PI Drafts Data Management Plan
- In-Person Consultation
- Info on PI
- Standards
- Data Repositories
- Questionnaire
### In-Person Consultation Tool: Questionnaire (sample questions)

<table>
<thead>
<tr>
<th>Section I: Data Products and Standards</th>
</tr>
</thead>
<tbody>
<tr>
<td>What naming conventions/schema will be used for your data, if any?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Section II: Data Storing and Long-Term Preservation</th>
</tr>
</thead>
<tbody>
<tr>
<td>During the project, how (i.e. media) and where (i.e. location(s)) will data be stored and who is responsible for it?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Section III: Data Sharing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Are there any data with privacy concerns to sharing (e.g., human subjects)? If so, what policies need to be adhered to and how will policies be enforced?</td>
</tr>
</tbody>
</table>
Social vs. Physical Science: Observations from data plans

Data Type/Qty
- Mixture of digital and physical material:
  - digital (documents, audio recordings, photos, database)
  - non digital (newspaper, interview notes, artifacts)
- Datasets on the scale of MBs and GBs, not TBs

Personal Identifiers
- Use of human subjects does not exempt one from sharing data
  - Graduate student received grant but was asked to modify data management plan and find something to share

Metadata
- More familiar with term “codebook” than “metadata”
- Metadata is more often created manually by the PI
### Data Flow for Survey on Opinion of the Global Extinction of Species

<table>
<thead>
<tr>
<th>Activity</th>
<th>Product</th>
<th>Associated Data Files</th>
</tr>
</thead>
</table>
| **Survey** | Conduct online Survey of Americans’ attitudes towards the extinction of species worldwide | **Raw Data**: Responses to Questions  
File Name: Raw_SpExt  
File Extension: xls  
Size: 10 MB |  
**Metadata**: Codebook  
File Name: Cbk_SpExt  
File Extension: xml  
Size: 10 MB |
| **Process** | Process Raw Data | **Processed Data**  
File Name: Proc_SpExt  
File Extension: xls  
Size: 10 MB  
File Name: Proc_SpExt  
File Extension: spv  
Size: 10 MB |
| **Analyze** | Statistically Analyze Processed Data | **Statistical Output**  
File Name: Stats_SpExt  
File Extension: spv  
Size: 10 MB  
File Name: GraphStats_SpExt  
File Extension: jnb  
Size: 15 MB |
Investigating trends in data curation

**Funders**  
new requirements & policy  
• Other funders adopting data management plan requirements

**Data curation**  
archiving, repositories, standards

• Expanding our expertise and knowledgebase

**Data sharing**  
Collaboration, data-intensive science

• E.g., academic credit for sharing data collections?
Thank you

David Fearon
Betsy Gunia

datamangement@jhu.edu
JHU Data Management Services
The Sheridan Libraries

http://dmp.data.jhu.edu/
(410) 516-0713