What Goes Where When in Managing Research Data?

David B. Lowe

University of Connecticut - Storrs, david.lowe@lib.uconn.edu

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 4.0 License.

http://escholarship.umassmed.edu/escience_symposium/2014/program/3

This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in University of Massachusetts and New England Area Librarian e-Science Symposium by an authorized administrator of eScholarship@UMMS. For more information, please contact Lisa.Palmer@umassmed.edu.
What Goes Where When in Managing Research Data?

David B. Lowe
Data Management Services Librarian
UConn Libraries
4/9/2014
University of Massachusetts and New England Area e-Science Symposium
## The Current State:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Products of research</td>
<td>Products of research</td>
<td>Data Collected, Formats and Standards</td>
<td>Data Collected, Formats and Standards</td>
<td>Data Collected, Formats and Standards</td>
<td>Data Collected, Formats and Standards</td>
<td>Data Collected, Formats and Standards</td>
<td>Data Collected, Formats and Standards</td>
<td>Data Collected, Formats and Standards</td>
<td>Data Collected, Formats and Standards</td>
<td>Data Collected, Formats and Standards</td>
<td>Data Collected, Formats and Standards</td>
</tr>
<tr>
<td>Access to data, and data sharing, distribution, and provision</td>
<td>Access to data, and data sharing, distribution, and provision</td>
<td>Policies and provision for re-use, redistribution, and products of derivatives</td>
<td>Policies and provision for re-use, redistribution, and products of derivatives</td>
<td>Policies and provision for re-use, redistribution, and products of derivatives</td>
<td>Policies and provision for re-use, redistribution, and products of derivatives</td>
<td>Policies and provision for re-use, redistribution, and products of derivatives</td>
<td>Policies and provision for re-use, redistribution, and products of derivatives</td>
<td>Policies and provision for re-use, redistribution, and products of derivatives</td>
<td>Policies and provision for re-use, redistribution, and products of derivatives</td>
<td>Policies and provision for re-use, redistribution, and products of derivatives</td>
<td>Policies and provision for re-use, redistribution, and products of derivatives</td>
</tr>
<tr>
<td>Archiving of data</td>
<td>Archiving of data</td>
<td>Roles and responsibilities</td>
<td>Roles and responsibilities</td>
<td>Roles and responsibilities</td>
<td>Roles and responsibilities</td>
<td>Roles and responsibilities</td>
<td>Roles and responsibilities</td>
<td>Roles and responsibilities</td>
<td>Roles and responsibilities</td>
<td>Roles and responsibilities</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
</tr>
</tbody>
</table>

**Means of sharing:**

- Data format and dissemination
- Data format and preservation
- Data format and accessibility

**Plans for archiving and preservation:**

- Plans for archiving and preservation
- Plans for archiving and preservation
- Plans for archiving and preservation
- Plans for archiving and preservation
- Plans for archiving and preservation
- Plans for archiving and preservation

**Data types and metadata standards:**

- Data types and metadata standards
- Data types and metadata standards
- Data types and metadata standards
- Data types and metadata standards
- Data types and metadata standards
- Data types and metadata standards

**Roles and responsibilities:**

- Roles and responsibilities
- Roles and responsibilities
- Roles and responsibilities
- Roles and responsibilities
- Roles and responsibilities
- Roles and responsibilities

**Period of data retention:**

- Period of data retention
- Period of data retention
- Period of data retention
- Period of data retention
- Period of data retention
- Period of data retention

**Expected data:**

- Expected data
- Expected data
- Expected data
- Expected data
- Expected data
- Expected data

**Data format and dissemination:**

- Data format and dissemination
- Data format and dissemination
- Data format and dissemination
- Data format and dissemination
- Data format and dissemination
- Data format and dissemination

**Data storage and preservation of access:**

- Data storage and preservation of access
- Data storage and preservation of access
- Data storage and preservation of access
- Data storage and preservation of access
- Data storage and preservation of access
- Data storage and preservation of access

**Additional possible data management requirements:**

- Additional possible data management requirements
- Additional possible data management requirements
- Additional possible data management requirements
- Additional possible data management requirements
- Additional possible data management requirements
- Additional possible data management requirements
So do you feel like this?
Utility of Suggested Solutions in the Context of Innovative Environments

“It depends.”  “One size fits all.”
Data Life Cycle Model
(Digital Curation Centre, UK)
From Cycle to Linear Timeline

- 10+ years
- 3 years

- Light bulb
- Hand with dollar sign
- Pie chart
- Laboratory equipment
- Wooden loom
- Tombstone with "RIP" and dollar sign
<table>
<thead>
<tr>
<th>I. UConn Affiliate Resource:</th>
<th>maxGB free</th>
<th>GroupShareable (within UConn)</th>
<th>GlobalShareable (beyond UConn)</th>
<th></th>
<th></th>
<th>Link:</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECM/FileNet</td>
<td>TBD</td>
<td>Yes</td>
<td>No</td>
<td>secure</td>
<td>State-hosted; not single-sign-on; more for dept. documents</td>
<td><a href="http://ecm.uconn.edu">http://ecm.uconn.edu</a></td>
</tr>
<tr>
<td>EFS or &quot;P:\ drive,&quot; individual</td>
<td>10</td>
<td>No</td>
<td>No</td>
<td>secure</td>
<td>questionable long term solution</td>
<td><a href="http://fileserver.uconn.edu">http://fileserver.uconn.edu</a></td>
</tr>
<tr>
<td>EFS or &quot;P:\ drive,&quot; departmental</td>
<td>10/member</td>
<td>Yes</td>
<td>No</td>
<td>secure</td>
<td>questionable long term solution</td>
<td></td>
</tr>
<tr>
<td>filelocker.uconn.edu</td>
<td>2</td>
<td>Yes</td>
<td>Yes</td>
<td>secure</td>
<td>short-term storage, sharing</td>
<td><a href="http://filelocker.uconn.edu">http://filelocker.uconn.edu</a></td>
</tr>
<tr>
<td>github.uconn.edu</td>
<td>1</td>
<td>Yes</td>
<td>Yes</td>
<td>secure</td>
<td>not for large or A/V files, but code</td>
<td><a href="http://github.uconn.edu">http://github.uconn.edu</a></td>
</tr>
<tr>
<td>google.uconn.edu Docs/Drive</td>
<td>30</td>
<td>Yes</td>
<td>Yes</td>
<td>secure</td>
<td>not long-term solution</td>
<td><a href="http://google.uconn.edu">http://google.uconn.edu</a></td>
</tr>
<tr>
<td>RNAS storage device</td>
<td>15</td>
<td>Yes</td>
<td>Yes</td>
<td>secure</td>
<td>now B2B only via dept.; B2C coming; not free above allotment</td>
<td><a href="http://rnas.uconn.edu/">http://rnas.uconn.edu/</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>II. Departmental or Personal Resource:</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Departmental server</td>
<td>varies</td>
<td>varies</td>
<td>varies</td>
<td>convenient</td>
<td>questionable long term solution</td>
<td>(your department)</td>
</tr>
<tr>
<td>media/devices</td>
<td>varies</td>
<td>No</td>
<td>No</td>
<td>convenient</td>
<td>highly fragile</td>
<td>(your favorite electronics retailer)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>III. Cloud solutions:</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Box.com</td>
<td>10</td>
<td>Yes</td>
<td>Yes</td>
<td>convenient</td>
<td>terms/security questionable</td>
<td><a href="http://www.box.com/">http://www.box.com/</a></td>
</tr>
<tr>
<td>Evernote</td>
<td>60MB upload max/month</td>
<td>Yes</td>
<td>Yes</td>
<td>convenient</td>
<td>terms/security questionable</td>
<td><a href="http://www.evernote.com/">http://www.evernote.com/</a></td>
</tr>
<tr>
<td>Google Docs/Drive (not via UConn)</td>
<td>15</td>
<td>Yes</td>
<td>Yes</td>
<td>convenient</td>
<td>terms/security questionable</td>
<td><a href="http://drive.google.com">http://drive.google.com</a></td>
</tr>
<tr>
<td>(Apple) iCloud</td>
<td>5</td>
<td>photos +iOS photos +iOS</td>
<td>convenient</td>
<td>terms/security questionable</td>
<td><a href="http://www.icloud.com">http://www.icloud.com</a></td>
<td></td>
</tr>
<tr>
<td>(MS) onedrive.com (was &quot;skydrive&quot;)</td>
<td>7</td>
<td>Yes</td>
<td>Yes</td>
<td>convenient</td>
<td>terms/security questionable</td>
<td><a href="http://www.skydrive.com">http://www.skydrive.com</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IV. Coming to UConn:</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Library-managed Repository*</td>
<td>TBD</td>
<td>Yes</td>
<td>Yes</td>
<td>secure, managed, long-term</td>
<td>not free above allotment; for archiving static data</td>
<td><a href="http://classguides.lib.uconn.edu/datamanagement">http://classguides.lib.uconn.edu/datamanagement</a></td>
</tr>
</tbody>
</table>

*available 2014

*Also consider bibliographic software (with article storage/access) and for fee services from above clouds as well as Amazon’s S3 & Glacier.*
Active Data Working Space Decisions

Acronyms:
2. B2C (from options matrix): Business-to-Customer, viz.: UITS to you
3. DoD: Department of Defense (restricted for national security)
4. ECM (from options matrix): Enterprise Content Management
5. EFS (from options matrix): Enterprise File Servers
6. NDA: Non-Disclosure Agreement (restricted by contract with a business entity)
7. PII: Personally Identifiable Information (restricted by laws for individual privacy)
8. RNAS: Research Networked Attached Storage
Static Data Decisions

Start w/ data set to which open access is required and ownership is clear.

- Are the data accompanied by sufficient metadata and documentation such that if you, as a knowledge/researcher, were seeing these files for the first time, you could make use of them?  
  - no
  - Add sufficient metadata and documentation
  - yes

- Do the data have long-term value (>5 years)?
  - no
  - Is all data in the set static now? (As in read-only & ready to be archived)
    - yes
    - Use that solution.
    - no
    - Would you like your data to be managed, networked, preserved for you, with analysis tools?
      - yes
      - Use the Library’s managed repository. Note: In addition to open, will also support levels of access, including restricted/dark.
      - no
      - Investigate options with campus IT or with cloud solutions.
  - yes
    - First, consider an established subject-specific data set repository (e.g. ICPSR for social sciences) and second, the Libraries’ managed repository.

Solution in mind?
Tightening Compliance Evidence from NSF Review Panel

• Fresh warning/cautionary tale from March 2014!

• PI had written in DMP (without our input):
  – “The data will be kept on the PI’s data cluster for ten years beyond the life of the project. The long-term preservation of digital objects ... is the PI’s responsibility.”

• Reviewer responded:
  – “NSF specifies that it is the University, not the PI, that has ultimate responsibility for long-term preservation of the data.”
  – Also questions access for research community
Near Future Work

• More decision trees!
  – Internal workflows
  – Embargoing and licensing levels
  – Cost modeling

• Direct costing institutional infrastructure
  – Within OMB Circular A-21 stipulations
  – Also per NSF DMP FAQ #5:
    • does not have to be “unfunded mandate”