ROI: Why a Framework is needed?

Ruth Duerr

This work is licensed under a Creative Commons Attribution v4.0 License.
Outline

• Definitions
• Background
• Repositories and some societal questions
## Definitions

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
<th>Implications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Curation</td>
<td><em>Processes that add value to foster discovery and reuse</em></td>
<td><em>Allows others to reuse data; supports cross-disciplinary research</em></td>
</tr>
<tr>
<td>Preservation</td>
<td><em>Providing enough representation information, context, metadata, etc. to support use and interpretation by agents other than the original data producer</em></td>
<td><em>Ability to use own data in the future (e.g., 5 yrs out)</em></td>
</tr>
<tr>
<td>Archiving</td>
<td><em>Activities that enable long-term retention of digital materials (e.g., replication, fixity, identifiers, etc.)</em></td>
<td><em>Provides identifiers for sharing, citation, etc.</em></td>
</tr>
<tr>
<td>Storage</td>
<td><em>Basic physical storage with backup and restore services</em></td>
<td><em>Allows basic sharing</em></td>
</tr>
</tbody>
</table>

### Curation Stack Model

- Curation
- Preservation
- Archiving
- Storage

Tiers of Maturity Assessment

- **Organizations (Capability)**
  - **Repository Processes Maturity** (e.g., CMMI Data Management Maturity)
  - **Repository Procedures Maturity** (e.g., ISO 16363:2012 - trustworthiness)

- **Portfolios (Asset Management)**
  - **Asset Management Maturity** (e.g., National Geospatial Dataset Asset)

- **Individual Datasets (Practices)**
  - **Stewardship Practices Maturity** (e.g., NCEI/CICS-NC DSMM)
Background

• What is the value of a repository?
  • What data does it hold?
  • What services can/does it provide?
  • How mature are those services?
  • What audiences can it serve?
What is the value of a data set?

- Analytic potential
  - Preservation readiness
    - Organization
    - Metadata
    - Formats
  - Number of potential user communities
  - Fitness for purpose
Metadata's Role as a Repository Differentiator

- Do they accept whatever they are given?
- Do they mandate provider's provide quality metadata and documentation?
- Do they help researchers create good metadata and documentation?

Who pays for metadata?
(Data organizational issues are similar)
User communities

• Who might it be relevant to?
  • Ruling out communities/applications is easier than identifying new ones
  • “Field data is small-scale, hard to find, uniquely described (PI-defined and organized) often of wide interest when compiled” - Ted Scambos
• Intellectual property and privacy issues?
  • Redaction vs restriction?
  • Data sharing rules
Repository Differentiators

- Can you actually get the data?
- Can you only get it in the form ingested?
  - Will it be migrated to a newer compatible format if that format dies?
- Can new audiences get representations usable by them?
- What more advanced services are available?

Who pays for preservation?
Who pays for curation?
Any ROI framework needs to make these differentiations clear
Questions?