Auditing an Institutional Data Repository With Respect to the CoreTrustSeal Certification Requirements

Jonathan Petters
Data Management Consultant/Curation Services Coordinator
Virginia Tech University Libraries

Naina Pisharoti
Graduate Assistant
Aerospace and Oceanic Engineering
Welcome to VTechData, the data repository for Virginia Tech!

Virginia Tech's Data Repository is a platform for highlighting, preserving, and providing access to the work generated by the Virginia Tech Community.
About VTechData

- VTechData highlights, preserves, and provide access to research products (e.g. datasets) of the Virginia Tech community
- Meeting journal and funder requirements for data sharing/availability
- Open access repository
- Virginia Tech researchers can self-deposit
- Curators suggest dataset/metadata enhancement by depositors, meeting minimum metadata requirements
- 74 published datasets since 2016
Internal Assessment – Why CoreTrustSeal?

- evaluation process looks at technical, curatorial, and organizational infrastructure
- lightweight repository certification process (relative to TRAC)
- gaining in notoriety within research data circles (e.g. Enabling FAIR Data Project)
- Facilitate conversation around Libraries’ digital infrastructure development/support
Assessment Process

- Interviewed Data Services and Information Technology Services personnel within Libraries towards 15 of the 16 CoreTrustSeal requirements
  - Requirement regarding funding and staffing out of scope
- For each CoreTrust Seal requirement authors gave
  - Ratings for VTechData
  - Recommendations for improving VTechData
Report Completed February 2019

- VTechData does not yet meet the CoreTrustSeal Repository Requirements
- Improvements and documentation required towards digital curation, digital preservation, and systems administration
- Report available online
Challenge for CoreTrustSeal Audit

- VTechData gives Virginia Tech researchers a way to provide access to their data
- Discipline-agnostic; serve political scientists, aerospace engineers, ecologists...
- As curators we are not in a position to evaluate the quality/reusability of all the data or metadata to which we provide access, now or in the future
  - Data Curation Network may lead to this capability for generalist repositories
Definition of Designated Community

- “the Designated Community of VTechData comprises researchers at Virginia Tech who can deposit and publish their data therein”
- With this definition, 3 of the 16 requirements (8, 11, 14) were deemed not applicable – would this have been alright towards CoreTrustSeal certification?
- Defining Designated Community an active point of discussion
  - CoreTrustSeal Review of Requirements
  - ‘Access is What we are Preserving’: But for Whom?
Thanks!

Questions?
<table>
<thead>
<tr>
<th>Rating</th>
<th>Description</th>
<th># Reqs at this Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Not Applicable</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>The Repository has not considered this yet</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>The Repository has a theoretical concept</td>
<td>4</td>
</tr>
<tr>
<td>3*</td>
<td>The Repository is in the Implementation Phase</td>
<td>6</td>
</tr>
<tr>
<td>4*</td>
<td>The guideline has been fully implemented in the Repository</td>
<td>2</td>
</tr>
</tbody>
</table>
OAIS Designated Community definition

- An identified group of potential Consumers who should be able to understand a particular set of information in ways exemplified by the Preservation Objectives. The Designated Community may be composed of multiple user communities. A Designated Community is defined by the Archive and this definition may change over time. (draft OAIS Reference Model v3)