Implementing a Cross-Institutional Staffing Model for Curating Research Data

The skills and expertise required to curate data cannot be fully automated nor reasonably provided by a few experts siloed at a single institution.

Method: In 2016-2017 six DCN partner institutions performed baseline assessments of local services, held focus groups with faculty researchers, ran controlled data curation pilots, and surveyed the library curator community to understand existing support and future plans for services in these areas. The results showed that multiple data curation experts are needed to effectively curate the diverse data types a repository typically receives and to keep up with changing trends and emerging tools that support research data best practice.

- Data sets observed across 6 DCN institutions in 2016-2017: 175
- Representation from scientific disciplines including agricultural, biological, engineering/applied, and physical sciences: 77%
- Different file types with ~30% tabular/spreadsheet files: 51%
- Average time spent curators spent processing the majority of datasets: 1 hour
- Percent of datasets lacking documentation: 47%
- Datasets that needed “Major” curation (e.g., documentation, transform file formats). Only 41 received this level: 77%
- Most datasets received in one day: 11

Levels of Curation Needed vs. Taken for Datasets Observed in 2016-2017 Planning Phase (n=175)

<table>
<thead>
<tr>
<th>Levels of Curation Needed</th>
<th>Levels of Curation Taken</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major</td>
<td>Local Curator</td>
</tr>
<tr>
<td>Major edits to the metadata</td>
<td>Local curator approves submissions and determines if the data should be submitted to the DCN for curation.</td>
</tr>
<tr>
<td>or major changes to the (new or missing) files</td>
<td>Local curator performs expert curation review and reports back to DCN coordinator.</td>
</tr>
<tr>
<td>Basic</td>
<td>Local Curator</td>
</tr>
<tr>
<td>Basic edits to the metadata</td>
<td>Local curator communicates needed actions back to the local curator.</td>
</tr>
<tr>
<td>or basic changes to the files</td>
<td>Local curator addresses curatorial issues and requests.</td>
</tr>
<tr>
<td>Minimal</td>
<td>Local Curator</td>
</tr>
<tr>
<td>No edits</td>
<td>Researcher</td>
</tr>
<tr>
<td>Curation Needed/Actions not documented</td>
<td>Local curator finalizes data submission and notifies DCN coordinator.</td>
</tr>
</tbody>
</table>

The Data Curation Network (DCN) Submission Workflow

The Data Curation Network (DCN) addresses this challenge by collaboratively sharing data curation staff across a network of partner institutions and data repositories.

The DCN brings the perspectives of researchers, librarians, administrators, and data curation subject experts in a cross-institutional network led by the University of Minnesota and including Cornell University, Penn State University, the University of Illinois, the University of Michigan, Washington University in St. Louis, Duke University, Johns Hopkins University, and the Dryad Data Repository. The DCN functions as:

- a well-coordinated and tiered staffing model that incorporates data curator expertise across a wide-variety of domains
- a technology-agnostic submission workflow that accommodates the various repository technologies in use (e.g. Samvera/Hydra, DataVerse, DSpace, BePress, etc.)
- standardized minimum levels of curation (CURATE Steps) that enable DCN Curators to prioritize their work and promote FAIR standards
- a sustainable, non-profit financial plan to support the DCN beyond the grant-supported implementation phase

and a professional development program that enables DCN curators to stay up-to-date with data best practices across domains, communities, nations, and beyond.

The Data Curation Network (DCN) Submission Workflow

- Submit: Local researcher self-deploys data set to local data curation service.
- Appraise: Local curator approves submissions and determines if the data should be submitted to the DCN for curation.
- Review & Assign: DCN Coordinator reviews submission and assigns to appropriate DCN curator based on file and discipline expertise.
- Curate: DCN Curator performs expert curation review and reports back to DCN coordinator.
- Take Action: DCN Coordinator communicates needed actions back to the local curator.
- Local curator assists local researcher with any actions (in person, via email, etc.).
- Re-submit: Local curator finalizes data submission and notifies DCN coordinator.
- Finalize: DCN curator reviews final data publication to ensure if DCN standards were met.
- Assess: DCN curator reviews final data publication to ensure if DCN standards were met.

Curator Needed: Curtatorexpected to be involved to curate the data. Curtatorexpected to be involved for tracking progress. "Uncurated" datasets are not included in this analysis.

Uncurated Data

Presenting scale and expertise challenges to individual institutions

Curated Data

At scale and with great efficiency through shared Data Curation Network

Local DCN Partner Institution Workflow

Ingest - Appraise and Select - DCN - Facilitate Access - Preserve - Long-Term

FAIRness (CURATE Steps) that enable DCN Curators to stay up-to-date with data best practices across domains, communities, nations, and beyond.

#datacuratonnetwork

Contributors: Lisa R. Johnston, (PI) University of Minnesota | Jake Carlson, University of Michigan | Cynthia Hudson-Vitale, Washington University in St. Louis | Heidi Imker, University of Illinois at Urbana-Champaign | Wendy Kooiowski, Cornell University | Robert Olendorf, Pennsylvania State University | Claire Stewart, University of Minnesota | Mara Blake, Johns Hopkins University | Joel Hermann, Duke University | Elizabeth Hull, Dryad Data Repository | Tim McGeary, Duke University

A web version of our DCN model is available with detailed curation workflows, staffing roles, draft FAQs, and technical functional requirements as “Data Curation Network: A Cross-Institutional Staffing Model for Curating Research Data” originally published July 27, 2017 online at https://sites.google.com/site/datacuratonnetwork/