Curating Research Data in DRUM: A workflow and distributed staffing model for institutional data repositories

Lisa Johnston
University of Minnesota - Twin Cities

Follow this and additional works at: https://escholarship.umassmed.edu/escience_symposium

Part of the Scholarly Communication Commons

This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 4.0 License.

Repository Citation

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.
Curating Research Data in DRUM
A workflow and distributed staffing model for institutional data repositories
Lisa R. Johnston, University of Minnesota – Twin Cities

Drum Submission Workflow

U of M researcher submits their data to DRUM using a custom submission workflow in Dspace.

DRUM coordinator receives notification and performs pre-acceptance appraisal.

Coordinator determines the type/discipline of data and assigns the submission to the appropriate DRUM curator.

Data curators open and review files. Detect missing information, validate software code, and create custom documentation.

Curators work closely with the author to augment the files and bring in the library liaison as needed.

Curators finalize the submission in DRUM and publish the dataset with a DataCite DOI.

Data Repository for the University of Minnesota (DRUM)
http://z.umn.edu/drumposter

- Open access IR-based data repository
- Beta launch November 2014
- Official campus launch March 2015
- Accepts datasets from our large and diverse academic community (no fees)
- All datasets are curated based on established procedures.
- DRUM staffed by a distributed staffing model for data curation to handling the wide variety of data sets.

Feedback from DRUM Submitters Post-Curation (n=17)

<table>
<thead>
<tr>
<th>Feature</th>
<th>Very Important</th>
<th>Somewhat Important</th>
<th>Not Important</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allow others free access</td>
<td>14 (82%)</td>
<td>5 (29%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Obtain persistent link</td>
<td>16 (94%)</td>
<td>1 (6%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Long-term preservation</td>
<td>16 (94%)</td>
<td>1 (6%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Curation assistance</td>
<td>16 (94%)</td>
<td>1 (6%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Full-text discoverability</td>
<td>15 (88%)</td>
<td>3 (18%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Track download statistics</td>
<td>16 (94%)</td>
<td>1 (6%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Monitor access to data</td>
<td>16 (94%)</td>
<td>1 (6%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Fulfill grant requirements</td>
<td>16 (94%)</td>
<td>1 (6%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Add CC license</td>
<td>16 (94%)</td>
<td>1 (6%)</td>
<td>0 (0%)</td>
</tr>
</tbody>
</table>

Would you recommend DRUM to a colleague at the U of M?

<table>
<thead>
<tr>
<th>Recommendation</th>
<th>Very easy</th>
<th>Somewhat easy</th>
<th>Somewhat difficult</th>
<th>Very difficult</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>14 (82%)</td>
<td>5 (29%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
</tbody>
</table>

Would you recommend working with DRUM curators?

<table>
<thead>
<tr>
<th>Recommendation</th>
<th>Very easy</th>
<th>Somewhat easy</th>
<th>Somewhat difficult</th>
<th>Very difficult</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>16 (94%)</td>
<td>1 (6%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
</tbody>
</table>

Curating Research Data in DRUM

Curators are assigned incoming submission based on subject area and data type; they review the data for usability and quality issues and work directly with the data authors to enrich the submission. Curation skills include generating custom metadata, arrangement and description of the objects, software-specific expertise, and file transformation for preservation needs.

Curated DRUM Submissions (Blue) vs. Uncurated Dataset Holding in Institutional Repository (Red) by Publication Date

Curating Research Data in DRUM

DRUM Submissions by College as of March 2016 (n=51)

- Business & Ed. 2%
- Administration 2%
- Veterinary Medicine 4%
- Biological Sci 6%
- Medical + Health Sci 8%
- Liberal Arts 10%
- Forestry, Ag, Natural Resources 17%
- Science & Eng 33%
- Education 6%
- Vet Medicine 4%
- Social Sciences 7%
- Engineering 16%
- Environmental 17%
- Humanites 33%
- Professional 17%
- Other 8%

** The University Digital Conservancy (the IR of the U. of Minnesota) collects unmediated dataset deposits from researchers. As of March 2016 the UDC housed a total of 301 datasets.

Distributed Data Curation

Curators are assigned incoming submission based on subject area and data type; they review the data for usability and quality issues and work directly with the data authors to enrich the submission. Curation skills include generating custom metadata, arrangement and description of the objects, software-specific expertise, and file transformation for preservation needs.

Curated DRUM Submissions (Blue) vs. Uncurated Dataset Holding in Institutional Repository (Red) by Publication Date

Curating Research Data in DRUM

DRUM Submissions by College as of March 2016 (n=51)

- Business & Ed. 2%
- Administration 2%
- Veterinary Medicine 4%
- Biological Sci 6%
- Medical + Health Sci 8%
- Liberal Arts 10%
- Forestry, Ag, Natural Resources 17%
- Science & Eng 33%
- Education 6%
- Vet Medicine 4%
- Social Sciences 7%
- Engineering 16%
- Environmental 17%
- Humanites 33%
- Professional 17%
- Other 8%

** The University Digital Conservancy (the IR of the U. of Minnesota) collects unmediated dataset deposits from researchers. As of March 2016 the UDC housed a total of 301 datasets.

Reference


Conclusion

An institutional data repository has the burden of collecting a diverse array of digital data, but, with appropriate staffing and careful procedures, each dataset (regardless of its’ distinctiveness) can be enriched for dissemination and reuse.