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

Creative Commons License
This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 4.0 License. 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.
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.1
• DRUM staffed by a distributed staffing model for data curation to handling the wide variety of data sets.

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

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

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

Curated Data Repository (Blue) (Approximately 100) vs. Uncurated Data Repository (Red) (Approximately 300)

Curated Data Repository (Blue) (Approximately 100) vs. Uncurated Data Repository (Red) (Approximately 300)

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

** 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 365 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.