

Should We Keep Everything Forever?: Determining Long-Term Value of Research Data

Bethany Anderson
University Archives
University of Illinois at Urbana-Champaign
bgandrsn@illinois.edu

Susan Braxton
Funk Library
University of Illinois at Urbana-Champaign
braxton@illinois.edu

Elise Dunham
Research Data Service
University of Illinois at Urbana-Champaign
emdunham@illinois.edu

Heidi Imker
Research Data Service
University of Illinois at Urbana-Champaign
imker@illinois.edu

Kyle Rimkus
Preservation Unit
University of Illinois at Urbana-Champaign
rimkus@illinois.edu

ABSTRACT

The University of Illinois at Urbana-Champaign's library-based Research Data Service (RDS) launched an institutional data repository called the Illinois Data Bank (IDB) in May 2016. The RDS makes a commitment to preserving and facilitating access to published research datasets for a minimum of five years after the date of publication in the Illinois Data Bank. The RDS has developed guidelines and processes for reviewing published datasets after their five-year commitment ends to determine whether to retain, deaccession, or dedicate more stewardship resources to datasets. In this poster, we will describe how the University of Illinois at Urbana-Champaign preservation review planning team drew upon appraisal and reappraisal theory and practices from the archives community to develop preservation review processes and guidelines for datasets published in the Illinois Data Bank.

Keywords

Innovative practice; appraisal; digital preservation; archival theory

1. INTRODUCTION

The Illinois Data Bank's [4] purpose is to provide University of Illinois at Urbana-Champaign researchers with a library-based repository for research data that will facilitate data sharing and ensure reliable stewardship of published data. The initiating goal that the IDB fulfills is that it provides a mechanism for researchers to be compliant with funder and/or journal requirements to make results of research publicly available. More broadly, the IDB endeavors to promote the discoverability and use of open research data by offering a preservation and access solution that is trusted by researchers at the University of Illinois at Urbana-Champaign. The Research Data Service (RDS) currently commits to preserve data and make it available for at least five years from the date of publication in the IDB.

In order to ensure that we are able to fulfill our commitment to stewarding University of Illinois at Urbana-Champaign research data in an effective and scalable manner, the RDS has established a policy framework that enables us to assess the long-term viability of a dataset deposited into the IDB. The RDS has developed guidelines and processes for reviewing published datasets after their five-year commitment ends to determine whether to retain, deaccession, or dedicate more stewardship

resources to datasets. Enacting a systematic approach to appraising the long-term value of research data will enable the RDS to allot resources to datasets in a way that is proportional to the datasets' value to research communities and its preservation viability.

2. PRESERVATION REVIEW

In this poster we will present the preservation review guidelines and processes we have developed within the context of archival appraisal theory and practice [1][2][3][5]. We will describe the automated measures we will implement to prioritize datasets for preservation review, as well as outline the Preservation Review Guidelines that preservation "Assessment Teams" will use to determine whether to retain, deaccession, or dedicate more stewardship resources toward datasets that undergo preservation review. The poster will also demonstrate the intended personnel make-up of "Assessment Teams" and examples of how dataset disposition will be documented and presented to IDB users.

The Illinois Data Bank Preservation Review Guidelines, which will be featured and expanded upon in this poster, are given in Table 1.

3. REFERENCES

- [1] Haas, J.K., Samuels, H.W. and Simmons, B.T. 1985. Appraising the records of modern science and technology: a guide. Massachusetts Institute of Technology.
- [2] Society of American Archivists, Technical Subcommittee on Guidelines for Reappraisal and Deaccessioning (TS-GRD). 2012. Guidelines for Reappraisal and Deaccessioning <http://www2.archivists.org/sites/all/files/GuidelinesForReappraisalAndDeaccessioning-May2012.pdf>.
- [3] UK Data Service. 2014. Collections Development Selection and Appraisal Criteria version 01.00 <https://www.ukdataservice.ac.uk/media/455175/cd234-collections-appraisal.pdf>
- [4] University of Illinois, Research Data Service. 2016. Illinois Data Bank <https://datbank.illinois.edu>.
- [5] Whyte, A. and Wilson, A. 2010. How to Appraise and Select Research Data for Curation. DCC How-to Guides. Edinburgh: Digital Curation Centre <http://www.dcc.ac.uk/resources/how-guides>

Table 1. Preservation review guidelines for the Illinois Data Bank

Evaluated by Curators/Librarians/Archivists

<i>Criterion</i>	<i>Consideration</i>
Cost to Store	What is the estimated cost of continuing to store the dataset?
Cost to Preserve	What is the estimated cost of continuing or escalating preservation for the dataset? Preservation actions may include file format migration, software emulation, and/or enhancement of preservation metadata.
Access	What do download and page view metrics indicate about interest in this dataset over time?
Citations	Has the dataset been cited in any publications?
Associated Publication Citations	If the dataset supports the conclusions of a publication, has that publication been cited in any other publications?
Restrictions	Does the dataset have any access or re-use restrictions associated with it?

Evaluated by Domain Experts

<i>Criterion</i>	<i>Consideration</i>
Possibility of Re-creation	Is it possible to create the dataset again?
Cost of Re-creation	If it is possible to create the dataset again, what would be the cost of doing so?
Impact of Study	Did the study that generated this dataset significantly impact one or more research disciplines?
Uniqueness of Study	Was the study that generated this dataset novel?
Quality of Study	Is the study that generated this dataset regarded as being of quality by domain experts?
Quality of Dataset	Is the dataset of quality according to domain experts?
Current Relevance	Is the dataset useful for addressing contemporary research questions according to domain experts?

Evaluated by Curators/Librarians/Archivists and Domain Experts

<i>Criterion</i>	<i>Consideration</i>
Availability of Other Copies	Is the copy of the dataset in the Illinois Data Bank the only one?
Understandability	Has the creator supplied sufficient metadata and documentation related to the dataset's creation, interpretation, and use in order to facilitate future discovery, access, and reuse?
Dependencies	Are the software, computing environment, or other technical requirements for using the dataset known? If so, are they available?
Appropriateness of Repository	Is there another trusted repository that, based on their collecting scope and user community, would be a better home for the dataset?