

University of Massachusetts Medical School

eScholarship@UMMS

University of Massachusetts and New England
Area Librarian e-Science Symposium

2014 e-Science Symposium

Apr 9th, 12:00 AM

What to Do about Data: An Overview of Guidelines and Policies for Dataset Collection Development

Sarah Young
Cornell University

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



Part of the [Collection Development and Management Commons](#)



This work is licensed under a [Creative Commons Attribution-Noncommercial-Share Alike 4.0 License](#).

Repository Citation

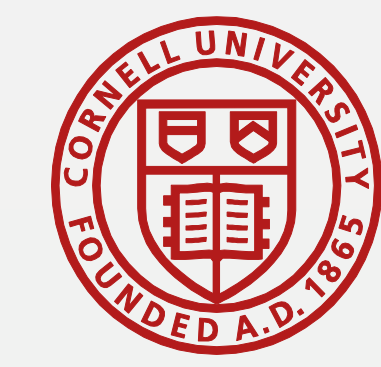
Young, S. (2014). What to Do about Data: An Overview of Guidelines and Policies for Dataset Collection Development. *University of Massachusetts and New England Area Librarian e-Science Symposium*. <https://doi.org/10.13028/9hk9-p212>. Retrieved from https://escholarship.umassmed.edu/escience_symposium/2014/posters/6

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.

What to do about data?



An overview of guidelines and policies for dataset collection development

Sarah Young, Health Science and Policy Librarian, Cornell University, sy493@cornell.edu

Considerations in dataset purchasing

Why purchase data?

Secondary datasets are increasingly important to researchers as they attempt to answer questions, make predictions and test hypotheses in new and powerful ways. For libraries that strive to provide information to support research needs, these datasets can be considered a **'new currency' in collection development**. There are many unique considerations in the collection and acquisition of datasets.

Purpose

The purpose of this overview was to get a sense of **current approaches to dataset collection development** at other research institutions, to determine key considerations in dataset purchasing, and to highlight particular challenges in implementing a dataset collection development program.

Methods

Currently existing dataset collection development policies, guidelines and programs were gathered from **web searches of academic library websites, calls to listservs and personal communications**. A total of 18 policies, guidelines, or programs were identified and considered in this work. A literature review was conducted with a focus on the collection of **commercially available datasets**.

Datasets should comply with the library's existing **storage capabilities**. Confidential data requires special storage and access considerations.

Datasets that come with **adequate documentation and relevant metadata** are preferred. Consider the language and ease of cataloguing.

Datasets purchased should be **institutionally accessible to all faculty, students and staff**. Terms should be in accordance with those for other electronic resource purchases made by the library. Consider fair use and the rights of scholars to data derivatives.

The amount the library is willing and able to contribute to a given dataset should be considered, with **joint purchases** between the library and the researchers when possible.

The commercial supplier of the data and the data itself should be vetted for **quality and reliability**, and long-term access ensured.

Data should be provided in a format that can be **supported by the library and used by the researcher**. Consider readability in commonly used statistical software.

Datasets with a broad subject appeal to the research community, **supporting the mission of the institution**, should be prioritized. Consider currency, the value of historical data, and geographic scope. Will the value of a dataset increase or decrease over time?

Storage needs

Cost

Quality

Documentation

Format

Terms of Use

Scope and Relevance

Getting a dataset collection development program off the ground:

Getting the word out

Liaison librarians and subject selectors can and should be involved in working with researchers and faculty across disciplines, particularly in the beginning stages of the dataset evaluation process. They can help determine if free datasets, or datasets already held in library collections, meet researcher needs and can get the word out to departments.

Handling requests

Requests can be handled on an **ad-hoc basis or via formal application** procedures. Two institutions examined in this study provided an online application process through which researchers could apply for library support for dataset purchasing (University of Cincinnati and the University of Illinois).

Negotiating licenses

License negotiation can be **lengthy and tedious**; commercial vendors selling datasets are often used to working with individual researchers, not libraries or institutional licensing arrangements.

Datasets in the Workflow

Decide whether datasets will be treated like other electronic acquisitions. Licenses may be negotiated by e-resource acquisitions departments with expertise in negotiating terms of use. datasets should be integrated into the normal cataloguing workflow, and should be considered a part of the digital preservation program.

Acknowledgements

Thanks to all of those who took the time to thoughtfully respond to listserv inquiries!

For references and links to dataset collection development policies, please see [handout](#).