

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



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.

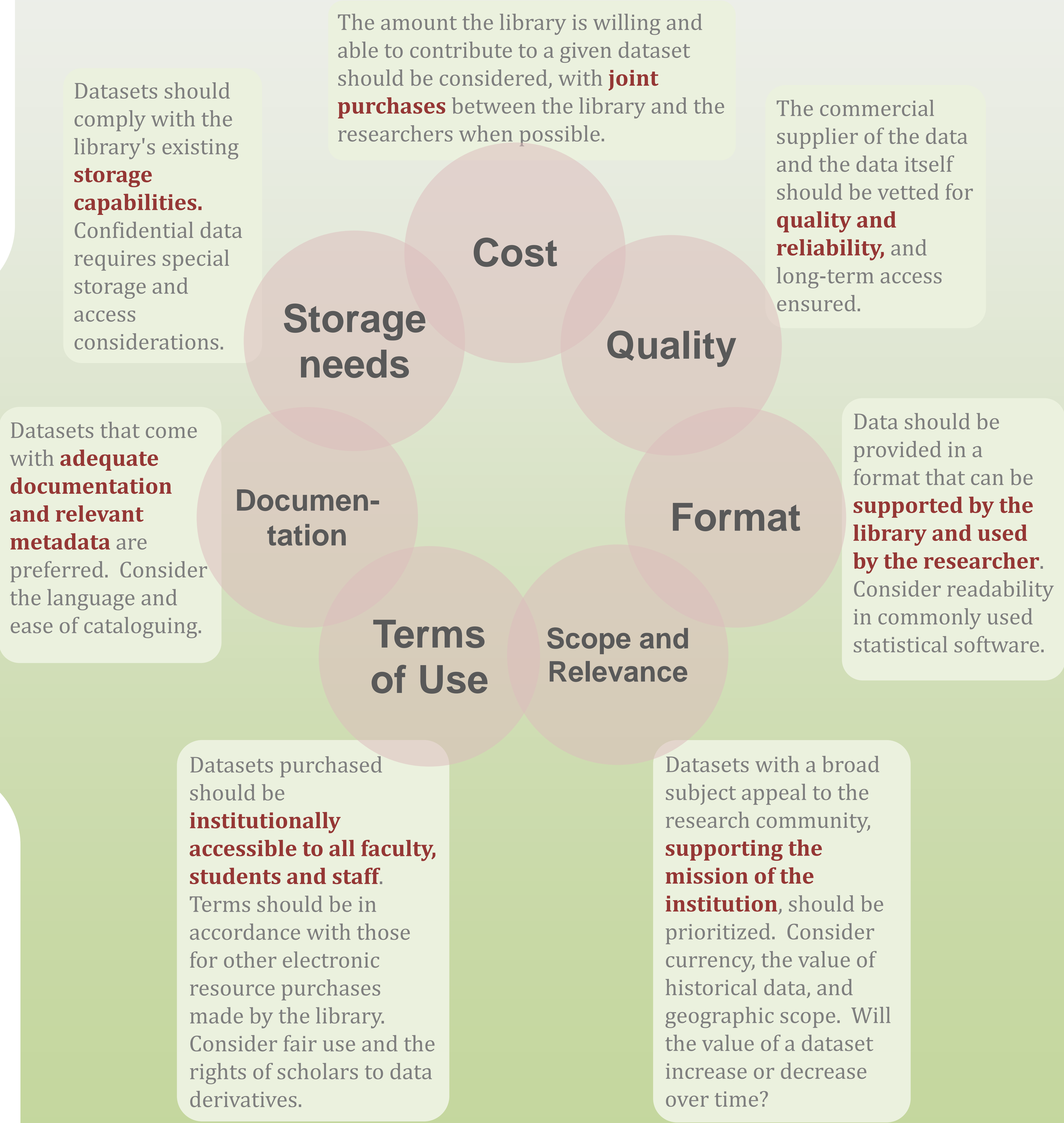
Purpose

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**.

Considerations in dataset purchasing



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](#).