

Research Data in eCommons @ Cornell: Present and Future

Wendy A. Kozlowski*, Dianne Dietrich, Gail Steinhart and Sarah Wright Cornell University Library, Ithaca, NY 14853 *wak57@cornell.edu

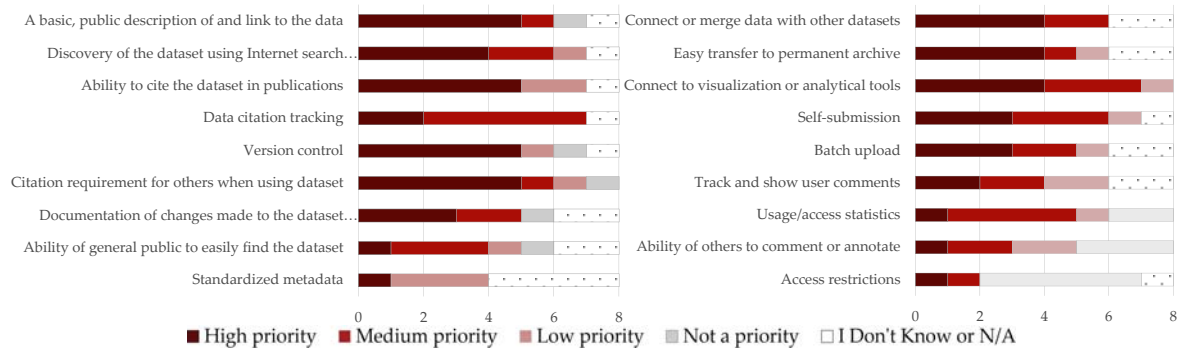
Introduction

As funding agencies increasingly prioritize sharing of research data, the role of institutional repositories (IRs) to house this material is likely to increase as well. By its very nature, data differs from the more traditional material housed in IRs such as publications, presentations, theses and dissertations. Given these distinctions, an effort to optimize functionality of eCommons to handle data could be helpful to accommodate future data deposits. To evaluate what potential eCommons users value in a repository for research data, we reviewed several sources of researcher feedback collected at Cornell University and elsewhere.

In the spring of 2012, 8 faculty and staff from Cornell University (CU) and Washington University in St Louis were interviewed using a modified Data Curation Profile (DCP) Toolkit¹. Researchers from a variety of disciplines were asked to prioritize features related to repository functionality (shown at right). Results are generally consistent with findings from a 2011 faculty survey on data management needs², DCPs completed at other institutions³ and other studies on data sharing⁴.

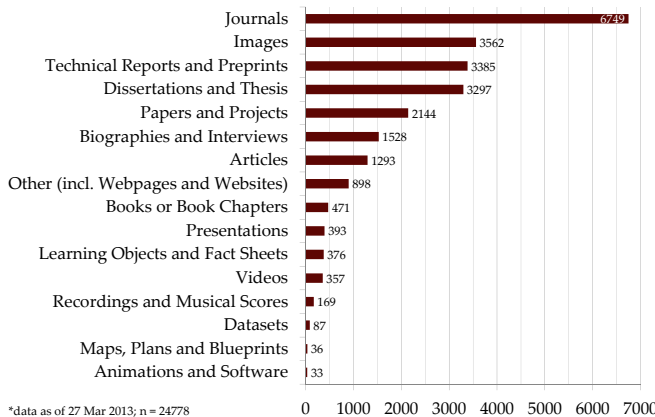
¹ <https://datacuratorprofiles.org>
² <http://dx.doi.org/10.7191/jeslib.2012.1008>
³ <http://hdl.handle.net/1853/28509>
⁴ doi:10.1371/journal.pone.0021101

What do researchers want?

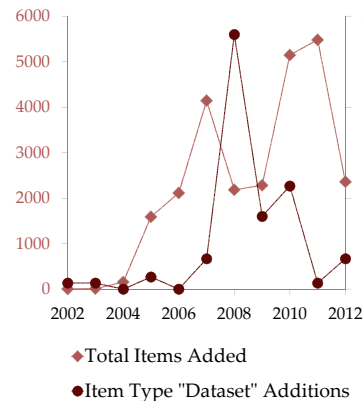


What does Cornell have?

Submitter-designated Item "Types" in eCommons*



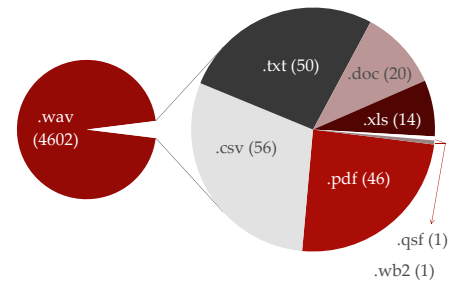
eCommons Submissions



Cornell University Library's IR, eCommons, is a DSpace powered repository available for materials in digital formats that may be useful for educational, scholarly, research or historical purposes. eCommons accepts research data with file sizes up to 1GB and individual collection sizes up to 10GB annually. By default, material is openly accessible via the web and under certain situations, access can be restricted to members of the Cornell community only and/or embargoed for a maximum of 5 years. Entries are assigned a persistent identifier (www.handle.net), and the CU Library is committed to preservation and to assuring long term access to contents. Upon deposit, users can assign an item type; presently, "dataset" items represent less than one half of one percent of total content (see figures, left). Datasets entries can be collections of multiple files; distribution of dataset file types is shown to the right.



Entry type "dataset" file extensions



How well are we meeting researcher needs and where can we go from here?

Key IR functions likely to be helpful to researchers	Assessment of current eCommons support	Considerations for the future of eCommons at Cornell
Discoverability via standard Internet search engines	Good, with some exceptions, such as incomplete indexing of large PDF's	In addition to Internet discoverability, DSpace 3.1 will offer enhanced search and browse features within the IR; upgrade planned for summer 2013.
Citation support (creation, export, tracking etc.)	Not currently supported	Explore creation of a suggested citation built in part from metadata; consider DOI assignment.
Version Control	Not currently supported	Item level versioning supported in DSpace 3.1.
Self-service submission	Available; current active registered users: 968 (564 have submitted)	Submission process may be additionally simplified using type-based metadata fields.
Access control by data owners	Access can be limited to a CU subgroup and limited embargos are allowed	Advanced embargo functionality supported in DSpace 3.1.
Infrastructure to allow for dataset updates (due to changes or addition of new data)	Datasets can be manually updated, but not without administrator support. Some datasets are updated by replacement, some by addition of new files.	Clearly articulated best-practices for dataset updates should be developed and added to eCommons usage policies.
Linking between data sets and related publications	Not currently supported	DSpace does not allow for this functionality, but linkages using VIVO and a CU metadata repository (sites.google.com/site/datastarsite) are currently in development.

