



Evidence Summary

Digital Libraries that Demonstrate High Levels of Mutual Complementarity in Collection-level Metadata Give a Richer Representation of their Content and Improve Subject Access for Users

A Review of:

Zavalina, O. L. (2013). Complementarity in subject metadata in large-scale digital libraries: A comparative analysis. *Cataloging & Classification Quarterly*, 52(1), 77-89.
<http://dx.doi.org/10.1080/01639374.2013.848316>

Reviewed by:

Aoife Lawton
Systems Librarian
Health Service Executive
Regional Library & Information Service, Dr. Steevens' Hospital
Dublin, Ireland
Email: aoife.lawton@hse.ie

Received: 29 Jul. 2014

Accepted: 14 Oct. 2014

© 2014 Lawton. This is an Open Access article distributed under the terms of the Creative Commons-Attribution-Noncommercial-Share Alike License 4.0 International (<http://creativecommons.org/licenses/by-nc-sa/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly attributed, not used for commercial purposes, and, if transformed, the resulting work is redistributed under the same or similar license to this one.

Abstract

Objective – To determine how well digital library content is represented through free-text and subject headings. Specifically to examine whether a combination of free-text description data and controlled vocabulary is more comprehensive than free-text description data alone in describing digital collections.

Design – Qualitative content analysis and complementarity comparison.

Setting – Three large scale cultural heritage digital libraries: one in Europe and two in the United States of America.

Methods – The researcher retrieved XML files of complete metadata records for two of the digital libraries, while the third library openly exposed its full metadata. The systematic samples obtained for all three libraries enabled qualitative content analysis to uncover how metadata values relate to each other at the collection level. The researcher retrieved 99 collection-level metadata records in total for analysis. The breakdown was 39, 33, and 27 records per digital library. When comparing metadata in the free-text Description metadata element with data in four controlled vocabulary elements, Subject, Geographic Coverage, Temporal Coverage and Object Type, the researcher observed three types of

complementarity: one-way, two-way and multiple-complementarity. The author refers to complementarity as “describing a collection’s subject matter with mutually complementary data values in controlled vocabulary and free-text subject metadata elements” (Zavalina, 2013, p. 77). For example, within a Temporal Coverage metadata element the term “19th century” would complement a Description metadata element “1850–1899” in the same record.

Main Results – The researcher found a high level of one-way complementarity in the metadata of all three digital libraries. This was mostly demonstrated by free-text data in the Description element complemented by data in the controlled vocabulary elements of Subject, Geographic Coverage, Temporal Coverage, and Object Type. Only one library demonstrated a significant proportion (19%) of redundancy between free-text and controlled vocabulary metadata. An example of redundancy found included a repetition of geographic information in both a Description and Geographic Coverage metadata elements.

Conclusion – The author reports high levels of mutual complementarity in the three cultural heritage digital libraries studied. The findings demonstrate that collection-level metadata which includes both free-text and controlled vocabulary is more representative of the intellectual content of the collections and improves subject access for users. The author maintains that there is no standard for collection-level metadata descriptions, and that this research may contribute to best practice guidelines in this area. It is unclear whether the digital libraries studied had written policies in place on how to describe collections and if those policies were adhered to in practice. The author expresses a need for further research to be conducted on collection-level metadata in other domains, such as science and interdisciplinary digital libraries, and on other scales (e.g., regional or state collections) and geographic regions beyond Europe and the United States.

Commentary

Although there is an abundance of general research and literature on metadata, digital libraries, interoperability, and standards, there is little specific research on complementarity of collection-level metadata. This research highlights this gap and adds value to the current body of research on this topic.

The methods of data collection and presentation of results score high on Glynn’s critical appraisal checklist (Glynn, 2006). This work builds on a previous study by the same author, which compared the free-text Description metadata field in multiple digital libraries (Zavalina, 2011). The current work extends the analysis to include three digital libraries of similar content and size and the comparative analysis of free-text Description element with four controlled vocabulary subject metadata fields. The sample size of metadata records from each library is similar, making the analysis sufficiently precise. Consent was obtained from developers of two of the digital libraries who agreed to provide the records for the content analysis. There is a detailed description of the level of comparative analysis carried out. However, there is no description of any tools that may have been used to compare the metadata records and retrieved digital content samples. A more detailed description of the methodology used would have been useful.

The results are well presented with five figures that visually demonstrate the level of complementarity between metadata elements. The extensive descriptive detail provided by the author enables further replication of this study on other digital library collections. The author thanks individuals at two universities for feedback on her study; however it is unclear whether the method of the content analysis used has been externally validated beyond this feedback. The author refers to the importance of complementarity of metadata in digital collections as an enabler of subject based access to collections. This study would

be strengthened with the inclusion of evidence about how users of cultural heritage collections browse or search collections by subject.

Based on the evidence presented in this paper, digital library managers, repository managers and cataloguers should consider the collection-level metadata that they are currently implementing and whether this is something that needs attention. Specifically, the potential for levels of redundancy observed in this study is an important finding. Cataloguers could save time by avoiding entering duplicate data, such as identical geographic information in both Description and Geographic Coverage elements. As the growth of digital library collections is set to continue, adherence to standards of metadata descriptors and best use of descriptive content including free-text and controlled vocabulary elements will become ever more important to discoverability. The author signposts two guidelines of relevance to cataloguers: the *Framework of Guidance for Building Good Digital Collections* established by NISO, and the *Guidelines for Digital Libraries* being prepared by the International Federation of Library Associations and Institutions in conjunction with the World Digital Library Project.

References

- Glynn, L. (2006). A critical appraisal tool for library and information research. *Library Hi Tech*, 24(3), 387-399. <http://dx.doi.org/10.1108/07378830610692154>
- Zavalina, O. L. (2011). Free-text collection-level subject metadata in large-scale digital libraries: A comparative content analysis. In T. Baker, D. I. Hillman, & A. Issac, (Eds.), *Metadata harmonization: Bridging languages of description. Proceedings of the DCMI International Conference on Dublin Core and Metadata Applications* (pp. 147.157). DCMI. Retrieved from <http://dcpapers.dublincore.org/pubs/index>