Best Practice Poster:
Normalizing Decentralized Metadata Practices Using Business Process Improvement Methodology: A Data-Informed Approach to Identifying Institutional Core Metadata

Emily Porter
Emory University, USA
eporter@emory.edu

Keywords: metadata standards; element sets; benchmarking; best practices; business process improvement; thematic analysis; quantitative analysis; qualitative analysis.

Environment, Context, and Techniques

The Emory University Libraries and Emory Center for Digital Scholarship have developed numerous digital collections over the past decade. Accompanying metadata originates via multiple business units, authoring tools and schemas, and is delivered to varied destination platforms. Seeking a more uniform metadata strategy, the Libraries' Metadata Working Group initiated a project in 2014 to define a set of core, schema-agnostic metadata elements relevant to local content types.

Quantitative and qualitative techniques commonly used in the field of Business Process Improvement were utilized to mitigate complex organizational factors. A key research deliverable emerged from benchmarking: a structured comparison of over 30 element sets, recording for each standard its descriptive element names, their required-ness, and general semantic concepts.
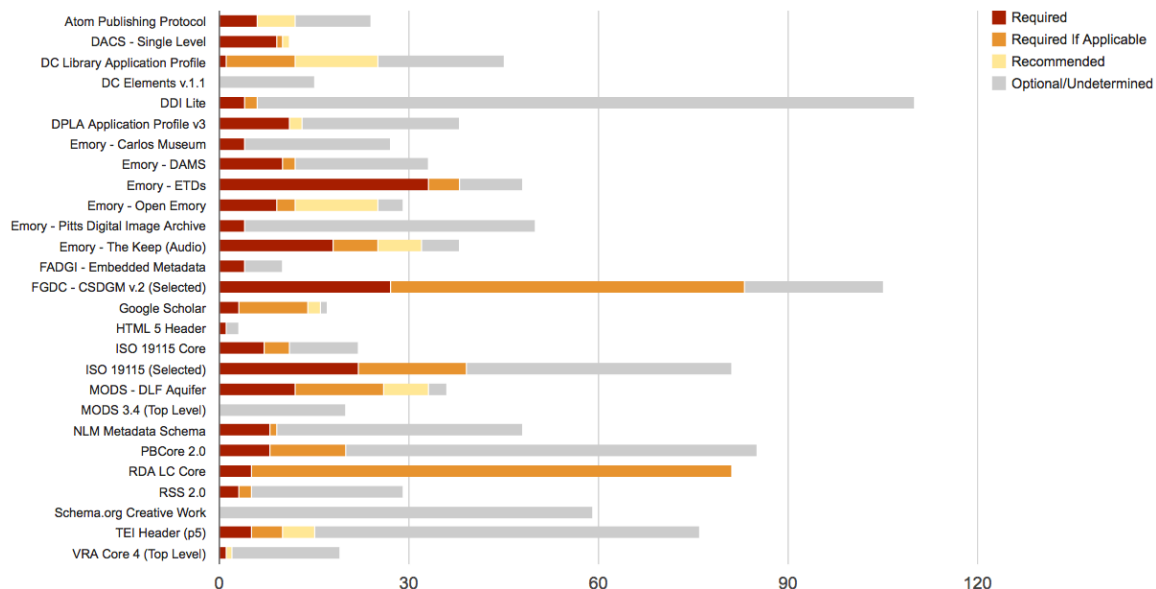

FIG. 1. Descriptive Elements by Schema/Standard: Quantity and Requirements (Selected Sources).

Additional structured data collection methodologies included a diagnostic task activity, in which participants with varying expertise created (simple) Dublin Core records for selected digital content. A survey of stakeholders provided greater context for local practices. Multiple public-facing discovery system interfaces were inventoried to log search, browse, filter, and sort options, and available web analytics were reviewed for user activity patterns correlating to these options.

Thematic analysis was performed on all benchmarking, system profiles, and web analytics data to map the results to a common set of conceptual themes, facilitating quantification and analysis. A weighted scoring model enabled the ranking of elements' themes: the highest scoring concepts then explicated as an initial set of core elements, mapped to relevant standards and schemas.

Acknowledgements

References

Atom Syndication Format – Introduction. (2007). Retrieved March 31, 2014, from http://atomenabled.org/developers/syndication/.

Berkman Center for Internet & Society at Harvard Law School. (2003). RSS 2.0 Specification (RSS 2.0 at Harvard Law). Retrieved March 31, 2014, from http://cyber.law.harvard.edu/rss/rss.html

Data Documentation Initiative. (2014). DDI Lite (Recommended Elements). Retrieved Month DD, YYYY, from http://www.ddialliance.org/sites/default/files/ddi-lite.html.

DCMI-Libraries Working Group. (2004). DC-Library Application Profile (DC-Lib). Retrieved Feb 13, 2014, from http://dublincore.org/documents/library-application-profile/.

Digital Library Federation. (2009). Digital Library Federation/Aquifer Guidelines for Shareable MODS Records, Version 1.1. Retrieved April 30, 2014, from https://wiki.dlib.indiana.edu/download/attachments/24288/DLFMODS_ImplementationGuidelines.pdf.

Digital Public Library of America. (2013). Metadata Application Profile, Version 3. Retrieved Month DD, YYYY, from http://dp.la/info/developers/map/.

Dublin Core Metadata Initiative. (2012). Dublin Core Metadata Element Set, version 1.1. Retrieved March 25, 2014, from http://www.dublincore.org/documents/dces/.

Embedded Metadata Working Group, Smithsonian Institution. (2010). Basic Guidelines for Minimal Descriptive Embedded Metadata in Digital Images. Retrieved April 7, 2014, from http://www.digitizationguidelines.gov/guidelines/GuidelinesEmbeddedMetadata.pdf.

Federal Geographic Data Committee. (1998). CSDGM Graphical Representation. Retrieved April 23, 2014, from http://www.fgdc.gov/csdgmgraphical/index.html.

Google Scholar. (n.d.). Inclusion Guidelines for Webmasters. Retrieved April 17, 2014, from http://scholar.google.com/intl/en-US/scholar/inclusion.html#indexing.

International Standardization Organization (ISO). (2003). *Geographic information — Metadata*. New York: American National Standards Institute.

Library of Congress. (2013). MODS User Guidelines (Version 3). Retrieved February 28, 2014, from http://www.loc.gov/standards/mods/userguide/index.html.

Library of Congress. (2014). LC RDA Core Elements. Retrieved April 8, 2014 from http://www.loc.gov/aba/rda/pdf/core_elements.pdf

Miller, S. (2011). *Metadata for Digital Collections*. New York: Neal-Schuman Publishers, Inc.

PBCore. (2011). Elements. Retrieved April 21, 2014, from http://www.pbcore.org/elements/.

Schema.org. (n.d.). CreativeWork. Retrieved April 9, 2014, from http://schema.org/CreativeWork.

Society of American Archivists. (2013). Describing Archives: A Content Standard. Retrieved May 2, 2014, from http://files.archivists.org/pubs/DACS2E-2013.pdf.

Text Encoding Initiative. (2014). 2 - The TEI Header. Retrieved April 7, 2014, from http://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html.

U.S. National Library of Medicine. (2004). NLM Metadata Schema. Retrieved April 17, 2014, from http://www.nlm.nih.gov/tsd/cataloging/metafilenew.html.

VRA Core Oversight Committee. (2007). VRA Core 4.0 Element Description. Retrieved March 31, 2014, from http://www.loc.gov/standards/vracore/VRA_Core4_Element_Description.pdf.

W3C. (2014). 4.2 Document metadata. Retrieved July 1, 2014, from http://www.w3.org/TR/html5/document-metadata.html