

Metadata Quality Assessment: A Phased Approach to Ensuring Long-term Access to Digital Resources

Daniel Gelaw Alemneh

University of North Texas; Digital Projects Unit,
Post Office Box 305190, Denton, Texas 76203, USA
E-mail: daniel.alemneh@unt.edu

Abstract

Maintaining usable digital collections necessitates maintaining high quality metadata about those digital objects. An effective metadata management approach can help institutions improve consistency, clarity of data lineage and relationships so that institutions can better use, reuse, and integrate resources. The metadata quality characteristics depend on various factors, including: type of resources, user perspectives, needs and priorities, which vary across groups of users. The metadata quality issue is particularly acute if there are a multiple institutions participating in collaborative digital projects, where a high level of interoperability is an important element. This poster demonstrates metadata quality assurance mechanisms by examining different quality assessment criteria, including metadata record completeness, consistency, accuracy, provenance, conformance to expectations, and other known substantive factors.

Introduction

The synergies of numerous emerging trends are shaping creation, access, use, and management of digital information resources. Digital collections are rapidly becoming an integral facet of any type of institutions around the globe. The volume and heterogeneity of digital resources grows daily, presenting challenges across cultural heritage institutions and other repositories of digital resources. The digital life cycle management starts from the point an item is selected for scanning (if not born-digital) and continues through image cleanup, metadata capture, derivative creation, and extends to ensuring long-term access.

Retrieval of information involves the user expressing requests by using terms from the common vocabulary and searching the file and matching requests with stored records. In order for end users to benefit fully from the development of various digital libraries in the increasingly self-structured Web 2.0 environment, service providers and collaborators need to maintain a high level of consistency across multiple data providers. Web 2.0 is indeed more than a buzzword and is the foundation for delivering a rich user experience to end users on the Web while leveraging the benefits of composite applications.

Metadata Management

Metadata is data that describes the characteristics of an item. Recognizing the critical role of metadata in any successful digital life-cycle management strategy, institutions that need to take responsibility for digital objects are increasingly implementing a metadata-based approach to ensuring long term access. The role of metadata in ensuring long-term access and management is analyzed, described, and commented upon by many researchers (Alemneh et al., 2002; Besser, 2002; Day, 2006; and Lavoie, 2008), among others.

Metadata is a systematic representation of an information-bearing object (text, images, audio, video, etc.) which points users to specific items on topics of interest. The creation of accurate metadata is fundamental to the discovery, use and reuse of digital contents. Significant progress has been made in raising awareness about the role of metadata in digital resource management and preservations. Many digital projects and initiatives believe that the backbone of their digital resources management function is the creation and maintenance of the detailed metadata associated with the digital object's significant properties.

If the digital library community is to provide optimal access to the diverse information resources and services available across digital repositories, all stakeholders must give high priority to the task of creating and maintaining the highest possible level of metadata quality.

Metadata Quality

Quality is a multidimensional concept and the two aspects of digital library data quality are the quality of the data in the objects themselves, and the quality of the metadata associated with the objects. Metadata errors occur in a variety of forms, but when errors exist, in whatever form, they block access to resources.

Metadata quality has a profound impact on the quality of services that can be provided to users. A good metadata enhances the value of a resource, provided that it is based on good analysis of the resource. A high quality metadata helps users find what they need, even when they are not sure themselves what they need. To fully understand what a good metadata record is, it is necessary to be both micro- and macro-minded. On the micro level, we concern ourselves with the specific mechanics of creating a quality metadata. On the macro level we put a metadata into the larger context of an information retrieval system.

The metadata quality characteristics depend on various factors. A number of metadata

researchers (Bruce and Hillman, 2004; Day, 2006; Guy et al., 2004; Jane et al., 2002; Thomas and Griffin, 1998; Moen, 1997; among others) have assessed metadata record quality by examining metadata record accuracy, provenance, consistency and logical coherence, timeliness, accessibility, subject term specificity and exhaustivity. Most agree that the appropriateness of any metadata elements needs to be measured by balancing the specificity of the knowledge that can be represented in it and queried from it and the expense of creating the descriptions. Although no consensus has been reached on conceptual and operational definitions of metadata quality; all emphasized the importance of metadata quality and noted that the quality of the metadata record that describes a digital object can affect the discovery, access and use of the resource.

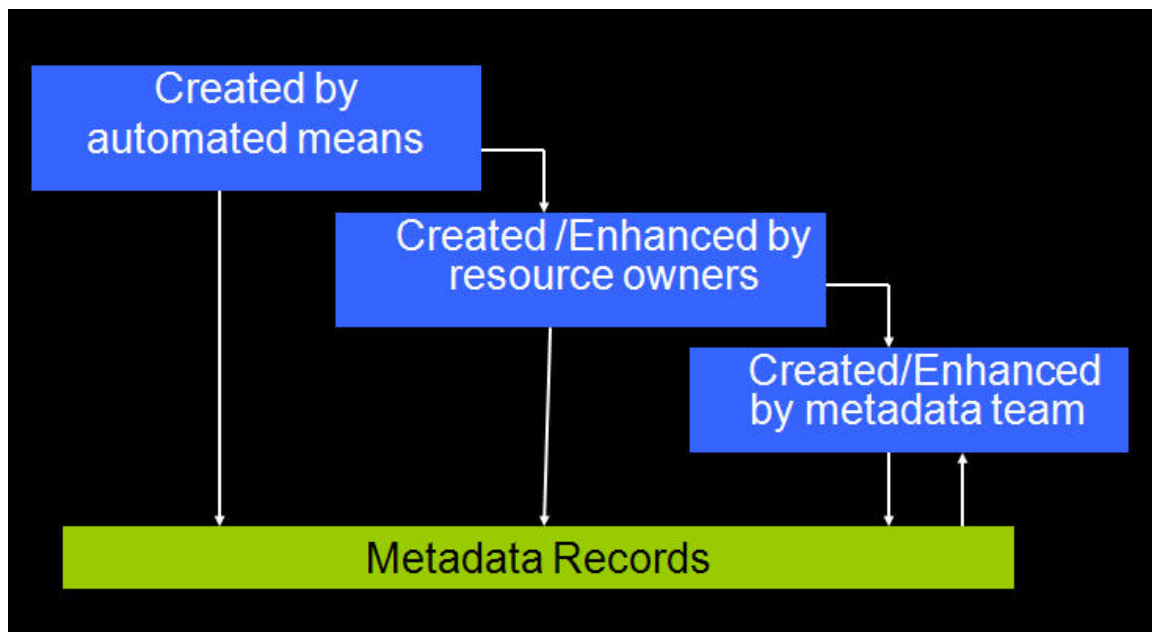


Figure 1. Metadata records creation and enhancement

Different institutions have developed a number of metadata quality assurance procedures, tools and associated quality assurance mechanisms. Figure 1 shows how the University of North Texas (UNT) digital projects approach metadata quality issues at various levels of digital resources life cycle. As depicted in Figure-2, different metadata analysis tools allow staff to assess the overall metadata quality in terms of how well are the existing (automatically and/or manually generated) metadata represent the existing digital resources in general.



Figure-2 UNT Libraries Metadata Analysis Tool

Figure-2 depicts some of UNT Libraries' metadata analysis tools that provide the ability to view the data in a multitude of engaging formats, including word clouds and geographic mapping functions, highlights usage and trends and points out errors and redundancies. The Metadata Analysis Tools build reports and aggregates statistics by data mining the digital objects metadata. The tools enable the digital projects team to provide timely feedback to contributors regarding metadata quality control issues.

Summary

Metadata errors, omissions and ambiguities result in problems with recall and precision and affect interoperability. Maintaining high quality metadata for every digital object requires a framework that provides the appropriate context needed to carry out quality assurance measures. As described in this document and summarized in Figure-8, like so many others, the UNT Libraries metadata team approaches metadata quality issues at various levels of the digital resources life cycle. The team continually reviews and refines the metadata creation processes and makes them up-to-date and useful in light of current requirements and developments in the field. Considering the complexities and multifaceted issues involved in determining the level of metadata quality required by all players, such a modular approach facilitates the flexibility and responsiveness required in such a diverse and collaborative digital projects environment.

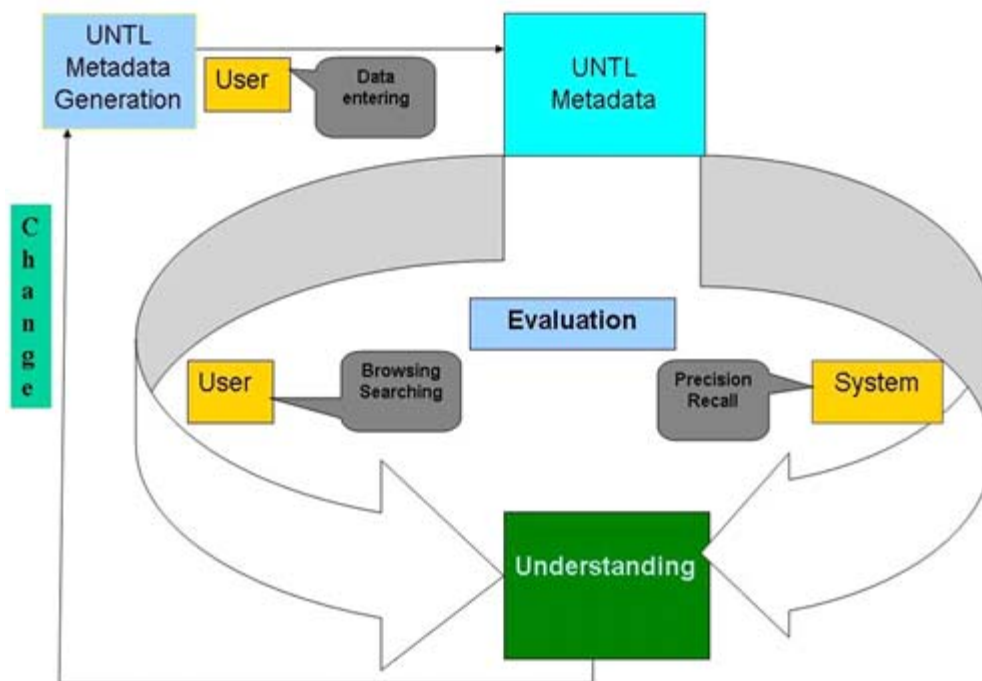


Figure-3 Quality assurance loop for UNT Libraries metadata workflow

Creation of quality metadata requires a community-wide approach. By federating and leveraging the power of open source software and commonly available tools, our community will be able to engage in scalable collaboration with the shared vision of building interoperable, usable, and durable digital libraries.

References

- Agnew, G. (2003). "Metadata Assessment: A Critical Niche within the NSDL Evaluation Strategy" Retrieved February 23, 2009 from http://eduimpact.comm.nsdl.org/evalworkshop/_agnew.php
- Alemneh, D. G., Hastings, S. K., and Hartman, C. N. (2002). "A Metadata Approach to Preservation of Digital Resources: The University of North Texas Libraries' Experience," *First Monday*, Vol. 7, No. 8 (August). Retrieved February 23, 2009 from http://firstmonday.org/issues/issue7_8/alemneh/index.html
- Barton J., and Robertson, J. (2005). "Designing workflows for quality assured metadata" *ETIS Metadata and Digital Repositories SIG meeting*, Retrieved February 23, 2009 from [http://mwi.cdlr.strath.ac.uk/Documents/2005%20CETIS%20\(MWI\).ppt](http://mwi.cdlr.strath.ac.uk/Documents/2005%20CETIS%20(MWI).ppt)
- Barton, J., Currier, S., and Hey, J. (2003). "Building quality assurance into metadata creation: an analysis based on the learning objects and e-prints communities of practice", *Proceedings of DC-2003*, Seattle, Washington Retrieved February 23, 2009 from http://www.siderean.com/dc2003/201_paper60.pdf
- Besser, H. (2002). The Next Stage: Moving from Isolated Digital Collections to Interoperable Digital Libraries. *First Monday*, 7(6). Retrieved February 23, 2009 from <http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/958/879>
- Bruce, T., & Hillman, D. (2004). The continuum of metadata quality: Defining, expressing, exploiting. In D. Hillman & E. Westbrooks (Eds.), *Metadata in practice* (pp. 238–256). Chicago, IL: ALA Editions.
- Day, M. (2006). The Long-term Preservation of Web Content. In Julien Masanes (Ed.), *Web Archiving* (pp. 177-199). Berlin Heidelberg: Springer-Verlag.
- Guy, M., Powell, A., and Day, M. (2004). "Improving the Quality of Metadata in Eprint Archives." *ARIADNE*, Issue 38, (January) Retrieved February 23, 2009 from <http://www.ariadne.ac.uk/issue38/guy/>
- Hillmann, D., Dushay, N., and Phipps, J. (2004). "Improving metadata quality: augmentation and recombination," *Proceedings of DC-2004*, Shanghai, China Retrieved February 23, 2009 from http://www.slais.ubc.ca/PEOPLE/faculty/tennis-p/dcpapers2004/Paper_21.pdf
- Lavoie, B. (2004). Preservation Metadata: Challenge, Collaboration, and Consensus. *Microform and Imaging Review* 33(3), 130-134.
- RLG (2008). Metadata Tools Forum Discussion Summary 2008-05-08. Retrieved February 23, 2009 from <http://www.oclc.org/programs/events/2008-05-08j.pdf>
- Zeng, M., and Mai, C. L. (2006). "Metadata Interoperability and Standardization – A Study of Methodology Part II." *D-Lib Magazine* Vol. 12, No. 6 Retrieved February 23, 2009 from <http://www.dlib.org/dlib/june06/zeng/06zeng.html>