

PREservation Metadata: Implementation Strategies

*Rebecca Guenther
Library of Congress
Washington, D.C., USA*

Abstract

PREservation Metadata: Implementation Strategies, or PREMIS, is an activity jointly sponsored by OCLC and RLG, focusing on issues associated with implementing preservation metadata in digital archiving systems. It was established to develop a common, implementable core set of metadata elements for digital preservation, since most published specifications for preservation-related metadata are either implementation-specific or broadly theoretical. PREMIS is composed of nearly 30 international experts representing national and university libraries, museums, archives, government agencies, and the private sector. PREMIS was charged to define a set of semantic units that are implementation-independent, practically-oriented, and likely to be needed by most preservation repositories as well as to identify and evaluate alternative strategies for encoding, storing, managing, and exchanging the core elements within a digital archiving system.

Initiated in June 2003, PREMIS will conclude its activities in early 2005. This paper will report on the findings and conclusions of the PREMIS working group. The deliverables of the group consist of a survey of existing preservation repositories, a final report, and a data dictionary.

As part of the PREMIS work, existing preservation repositories were surveyed about their architectural models and metadata practices. This paper will summarize the results of this survey which addressed the mission, policy, economic, and technical aspects of digital repositories, as well as current practices for creating, managing, and maintaining preservation metadata within the repository environment. Analysis of the survey responses suggests that

the digital preservation community is beginning to coalesce around several emerging trends and best practices in the use and management of preservation metadata, which will be enumerated.

The second deliverable of the group is a data dictionary and final report. The semantic units that support long term preservation are represented in a data dictionary with implementation details and in METS-compatible XML schemas. In the course of this work, the group also developed a glossary of terms and concepts, a data model, and a typology of relationships. The paper will discuss the data model as well as key issues that arose in the consensus-building process during the development of the data dictionary.

The PREMIS working group report and data dictionary will be issued in March 2005, with the XML schemas to follow shortly thereafter. In addition to providing the data dictionary and accompanying documentation, the paper will also suggest follow-on activities to test the PREMIS data element set.

Biography

Rebecca Guenther is Senior Networking and Standards Specialist in the Network Development and MARC Standards Office of the Library of Congress. Her current responsibilities include work on national and international information standards, particularly metadata. Some of her current activities include: co-chair of PREMIS; rotating chair of ISO 639 Joint Advisory Committee on language codes; member of the National Information Standards Organization's Standards Development Committee; member of the Dublin Core Usage Board.