

# Implementing Preservation Repositories For Digital Materials: Current Practice And Emerging Trends In The Cultural Heritage Community

---

A Report by the PREMIS Working Group  
September 2004

---

PREMIS—Preservation Metadata: Implementation Strategies  
A Working Group Jointly Sponsored By OCLC and RLG

## **Abstract**

In Winter 2003-2004, the PREMIS working group conducted a survey aimed at gathering information on key aspects of planned and existing preservation repositories for digital materials. Survey questions touched on a variety of areas, such as mission, funding, preservation strategy, and access policies, but with an overarching focus on current practice for managing preservation metadata in digital archiving systems. Survey responses were received from nearly fifty institutions located in thirteen countries, and included libraries, archives, museums, and other institutions. Analysis of the responses suggests that the digital preservation community is beginning to coalesce around several emerging trends in the use and management of preservation metadata, which are enumerated and discussed at the conclusion of the report.

© 2004 OCLC Online Computer Library, Inc. and The Research Libraries Group, Inc. OCLC: 6565 Frantz Road, Dublin, OH 43017-3395 USA, <http://www.oclc.org/>; RLG: 2029 Stierlin Court, Suite 100, Mountain View, CA 94043-4684 USA, <http://www.rlg.org/>.

Reproduction of substantial portions of this publication must contain this copyright notice.

## **Suggested citation:**

OCLC/RLG PREMIS Working Group. 2004. "Implementing Preservation Repositories for Digital Materials: Current Practice and Emerging Trends in the Cultural Heritage Community." Report by the joint OCLC/RLG Working Group Preservation Metadata: Implementation Strategies (PREMIS). Dublin, O.: OCLC Online Computer Library Center, Inc. Available online at: <http://www.oclc.org/research/projects/pmwg/surveyreport.pdf> (PDF:668K/66pp.)

The following members of the PREMIS Implementation Strategies Subgroup contributed to this report:

George Barnum  
Charles Blair  
Olaf Brandt  
Priscilla Caplan, Chair  
Robin Dale  
Keith Glavash  
Rebecca Guenther  
Cathy Hartman  
Nancy Hoebelheinrich  
Mela Kircher  
John Kunze  
Brian Lavoie  
Vicky McCargar  
Evan Owens  
Angela Spinazze  
Stefan Strathmann  
Robin Wendler  
Hilde van Wijngaarden  
Deborah Woodyard

With staff assistance from Jennifer Childree, FCLA.

# TABLE OF CONTENTS

EXECUTIVE SUMMARY .....	5
Introduction .....	9
1 Characterization of Respondents .....	11
2 Mission and Services.....	13
2.1 What is the mission of the preservation repository? Is the repository strictly dedicated to long-term preservation only or does it serve other goals too? .....	13
2.2 Who can deposit materials into your preservation repository?.....	14
2.3 What services does your preservation repository provide? (Check all that apply.).....	16
2.4 Does your repository manage and track non-digital versions of digital materials as well? .....	17
2.5 How is the preservation repository funded? (Check all that apply.) .....	18
3 Models and Policies.....	18
3.1 What types of materials are accepted by this preservation repository? .....	18
3.2 How are materials obtained by the preservation repository? (Check all that apply.).....	19
3.3 What kinds of agreements does the repository have for obtaining materials? (Check all that apply.).....	19
3.4 Are there formal signed contracts or agreements with customers/depositors? 19	
3.5 What are the policies or practices of the preservation repository regarding access to the materials? (Check all that apply.).....	25
3.6 How is your preservation repository informed by the Open Archival Information Systems model (OAIS)? What features of the repository would you say conform to OAIS? What features, if any, do not?.....	26
3.7 In which ways do you find the OAIS reference model useful? Are there ways in which the model is unhelpful? .....	27
4 Architecture, Storage, and Preservation Processes .....	28
4.1 What is the relationship between access and preservation copies of materials stored in your preservation repository? (Check all that apply.) .....	28
4.2 (Answer only if both access and preservation copies are stored) Do access copies get preservation treatment (e.g. migration)? Is the relationship between access and preservation copies maintained by the repository? If so, how?....	29
4.3 How are metadata and materials stored within the preservation repository? For example, one repository might zip metadata together with content files and store the zip file as a single entity. Another repository might store metadata in relational database tables and content files as individual entities in a file system. A third repository might use multiple models for different types of materials. Please describe all of the models that apply. ....	33
4.4 What preservation strategies are your preservation repository implementing now? (Check all that apply.) Why did you chose these particular strategies? .	35

4.5	What preservation strategies are your preservation repository planning to implement in the future? (Check all that apply.) Why did you chose these particular strategies? .....	39
4.6	What type of applications software is in use in your preservation repository? (Check all that apply.).....	41
5	Metadata.....	44
5.1	What categories of metadata are (or will be) stored by and used by your preservation repository? (Check all that apply.) .....	44
5.2	If you are using or planning to use metadata elements from one or more published scheme, which schemes are you using? (Check all that apply.).....	45
5.3	Does your repository record information about these types of entities? (Check all that apply.) .....	47
5.4	How is metadata obtained (or expected to be obtained) by the preservation repository? For example, is it submitted by depositors, extracted automatically by the repository's computer programs, other? If different methods are used for different sets of metadata, please note all of them. ....	49
5.5	Do you require contributors to your preservation repository to provide metadata with their contributions? If so, what are your requirements? Please provide sample documentation if possible. ....	50
5.6	How is metadata stored and updated in your preservation repository? If multiple methods are used, please explain. ....	51
5.7	Some preservation repositories create or plan to create normalized or migrated versions of digital materials. If this applies to your repository, how will metadata for the original and new versions be handled? What information (if any) will be recorded only for the version created by the repository? .....	51
5.8	What metadata do you feel is most important to record for preservation purposes?.....	52
6	DISCUSSION .....	53
7	REFERENCES .....	55
	APPENDIX A: Alphabetical List of Survey Respondents .....	56
	APPENDIX B: Survey Instrument - PREMIS Implementation Survey.....	57
	APPENDIX C: Follow-up Interview Questions .....	64

---

## EXECUTIVE SUMMARY

In June 2003, OCLC and RLG convened a Working Group, Preservation Metadata: Implementation Strategies (PREMIS), to focus on the practical aspects of implementing preservation metadata in digital preservation systems. The group has an international membership drawn from library, academic, museum, archive, government, and commercial sectors. One of the tasks in the PREMIS charge is to examine and evaluate alternative strategies for the encoding, storage, and management of preservation metadata within a digital preservation system.

To this end, in November 2003, the group distributed a survey about practices in digital preservation archiving. The survey included questions about business plans, policies, architecture and preservation strategies as well as metadata practices. Copies were sent directly to approximately seventy organizations known to be active or interested in digital preservation. The survey was also made available through several discussion lists. In February 2004, the group made a second distribution, sending thirteen more copies to museums, museum organizations and art institutes, and announcing the survey on lists targeting the museum community.

A total of forty-eight complete survey responses were received. Sixteen of the respondents were later selected for follow-up interviews to gather more information about some practices. Responses came from twenty-eight libraries, seven archives, three museums, and eleven other types of institutions. (A complete list of survey respondents is provided in [Appendix A](#).) Responses were received from thirteen different countries; 46% were from the U.S. Just under half of the organizations had at least some part of their preservation repository in production, while 70% reported being in some stage of planning or development. However, of these, only eleven institutions appeared to have realized an active preservation strategy (migration, emulation, normalization) in production.

Significant results include the following. (Numbers in parentheses refer to sections of this report containing these conclusions.)

The cultural heritage community has very little experience with digital preservation. We do not have enough experience to indicate whether the metadata these systems record, or plan to record, are adequate for the purpose. (1)

Most repositories serve the two goals of preservation and access. Less than a fifth could be called “dark archives.” (2)

National libraries, archives, university libraries, state libraries, and art museums all seem to have clearly defined constituencies in terms of who can deposit and make use of archived materials, as well as coherent sets of materials of concern. (2)

90% of respondents funded their repositories from their operational budget, while two-thirds used internal or external grant funds, in addition to, or instead of, operations funds. (2)

Differences between libraries and archives in terms of materials accepted are significant and reflect the difference in mission. All archives accepted electronic records and the majority accepted datasets and audio/video. Libraries showed less support for datasets and audio/video but more support for locally digitized materials and web resources. (2)

More than half of respondents had, or planned to have, formal signed agreements with depositors. Most of the agreements contain language that attempts to describe the uses to which repository content can be put, rather than describing in detail the mechanisms of preservation. (3)

Most respondents claim to have been informed by the Open Archival Information System (OAIS) framework, and most say that they at least partly conform to the model. Definitions of OAIS compliance vary, and there is strong demand for supplementary materials including reports and manuals supporting implementation. (3)

All respondents offered “secure storage” as a service; 92% offered, or planned to offer, preservation treatments, defined as normalization, migration, emulation, or other strategies designed to ensure long-term usability. (2)

The majority of institutions chose more than one strategy for preservation. Most (85%) are offering bit-level preservation. Beyond that, restrictions on submissions, normalization, migration and migration-on-demand are the four most popular strategies, in that order. According to the respondent’s future plans, the four most popular strategies, in order, will be migration, normalization, restrictions on submission, and migration-on-demand. Emulation is being used now by only 10% of respondents, but that doubles when future plans are considered. (3)

Most repositories are using some combination of commercially available, open source, and locally developed software. By far the majority are using a combination of software applications. Seventy different commercial and open source software products were specifically named. (4)

Most respondents are recording a wide range of types of metadata; more than half are recording elements of rights, provenance, technical, administrative, descriptive, and structural metadata. (5)

For non-descriptive metadata, METS (Metadata Encoding and Transmission Standard) was by far the most commonly used metadata scheme: 64% of libraries, 42% of archives, and 35% of other institutions used, or planned to use, METS. Z39.87 (Technical metadata for digital still images) was widely used by libraries but not others. (5)

Basing a local metadata scheme on other existing schemes is common, as is using more than one scheme in the repository. (5)







































































































































