



Long-term Repositories: Taking the Shock out of the Future

An APSR event, hosted by the National Library of Australia
August 31 and September 1, 2006

Day 1 - Implementing Preservation Metadata with PREMIS

Presenters and Abstracts

Rebecca Guenther

Networking & Standards Specialist, Library of Congress

The Promise of PREMIS: Background, Scope & Purpose of the Data Dictionary for Preservation Metadata

Released in May 2005, the PREMIS Data Dictionary for Preservation Metadata defines and describes implementable, core preservation metadata with broad applicability to digital repositories. This presentation will provide:

- a review of previous efforts that led to the formation of the PREMIS working group;
- an overview of the Data Dictionary, including its assumptions, data model, and technical neutrality;
- a review of some of the projects that have implemented the data dictionary since its release;
- a discussion of some of the implementation choices to be made with PREMIS especially using XML;
- an overview of the maintenance activity and future plans.

Chris Blackall

*Development Portfolio Co-ordinator,
Australian Partnership for Sustainable Repositories (APSR)*

The Architecture and Standards Environments in which PREMIS Needs to Operate

This presentation provides an overview of the information architectures and standards in which PREMIS will be implemented. The key standards related to PREMIS will be identified and discussed, as well as the crucial relationships between these standards and the underlying information architectures and systems adopted by institutions that are, taken together, the essential foundations for the long-term preservation and accessibility of digital information. To illustrate the relationship between standards and information architectures and systems, a range of examples will be discussed, drawn from APSR-supported projects that are using institutional repositories to preserve research information and data. Finally, based on these examples, there will be discussion in broad terms of the opportunities for improving digital preservation that lay ahead, but also some of the barriers.

Deborah Woodyard-Robinson

Digital Preservation Consultant, Woodyard-Robinson Holdings Ltd, Christchurch, New Zealand

Common Use Cases for Preservation Metadata

Metadata should exist to support function, but functions such as the preservation actions for digital material are unlikely to be designed before a repository is built. Examining the anticipated use cases and comparing these with experience from current projects will help to construct the appropriate context for the use of preservation metadata and PREMIS.

Jane Hunter

*Professorial Research Fellow,
School of Information Technology & Electrical Engineering, University of Queensland*

PREMIS and the Role of Preservation Metadata in Making Preservation Systems Workable

This presentation will discuss the role of preservation metadata in enabling preservation systems, including automating interaction with Web Services and external registries, such as format registries, to support preservation. The AONS (Automatic Obsolescence Notification System) project, which draws on these principles, will also be described.

Bronwyn Lee

Business Analyst, National Library of Australia

Preservation Metadata - Adapting or Adopting PREMIS for APSR

APSR aims to establish a centre of excellence in sustainable digital resource management, and partner universities are developing demonstrator repositories built on sustainability principles. This presentation outlines the results of a project commissioned by APSR to specify requirements for the collection of metadata needed for long term continuity of access to digital collections. The project was called PRESTA (PREMIS Requirements Statement) but it took a broader view than PREMIS alone.

The presentation will highlight the following areas of the project report:

- recommended preservation metadata elements including 'mandatory' elements;
- recommended supported file formats;
- recommended tools for automatic metadata extraction and their capabilities;
- gaps in preservation metadata collected in selected partner repositories and recommendations for enhancements;
- functional specifications and use cases for preservation events and event ('history') logging, a significant piece of the digital preservation framework not yet included in partner repositories;
- a draft METS profile for exchanging preservation metadata.

Peter Raftos

*Program Leader,
Digital Resource Services,
Australian National University*

Scott Yeadon

*Project Officer,
Australian National University*

Christiaan Kortekaas

*Senior Library Systems Programmer,
University of Queensland*

Sustainability through Interoperability: PREMIS, Fedora and DSpace

The PREMIS data dictionary identifies the core preservation metadata elements required to support the long-term sustainability of digital objects. In this presentation, an overview of the current level of PREMIS support within the Fedora and DSpace repository software is provided, as well as information on how PREMIS support will be improved in later versions of the software. In addition, the importance of repository interoperability is discussed and demonstrated by using the National Library of Australia's draft METS SIP/DIP to exchange repository objects in a meaningful way between Fedora and DSpace.

Mathew Black

Applications & Integration Architect, National Library of New Zealand

Tools and Techniques for Preservation Metadata Extraction and Collection

Presentation topics include: the tools and techniques that can be used to automatically extract preservation metadata embedded or associated with digital objects; what to do with the extracted metadata; what can and can't be done with current tools (including the National Library of New Zealand's Preservation Metadata Extract Tool); future technologies; workflows; validation/extraction stacks; and the importance of good metadata structure for effective preservation reporting and risk analysis.

Day 2 - Certifying Digital Repositories: Using the RLG-NARA Audit Checklist

Presenters and Abstracts

Robin Dale

Program Officer, Research Libraries Group (RLG)

Certifying Digital Repositories: Development of the Audit Checklist by the RLG-NARA Taskforce

Over the last three years, the RLG - National Archives and Records Administration (US) Taskforce on Digital Repository certification has worked to develop a checklist which would allow objective evaluation of digital repositories and archives. The checklist was based upon international standards and community best practices for digital preservation as well as influences from information technology and information security standards. The goal was to develop metrics to evaluate three aspects influencing repository trustworthiness: the characteristics of the archiving organization that affect performance, accountability, and business continuity; technologies and technical infrastructure employed by the repository; and finally the processes and procedures utilized by the repository.

A draft checklist was released for public consumption and comment in August 2005, leading to further development. In addition, the checklist has been leveraged by the Center for Research Libraries Auditing and Certification of Digital Archives project and used in test audits of three digital repositories and archives. Through public comment, practical application, and other CRL project activities, we have made further progress toward the development of a fully-fledged program of audit & certification for digital repositories. This presentation will discuss checklist development, envisioned uses of the checklist, the emerging audit & certification process, potential audit & certification process outputs, and compliance issues.

Ron Jantz

Digital Library Architect, Rutgers University Libraries

Fedora Preservation Services - A Working Group Report

A major objective of the Fedora Preservation Services Working Group (WG) is to facilitate the creation of trusted digital repositories. To accomplish this end, the Working Group is specifying the requirements and architecture for services and technologies that can be readily integrated into the Fedora Framework. In the specification process, the WG is focused on the underlying capabilities to support digital object persistence, life cycle management, multidisciplinary collections, and management of the repository environment (e.g. storage, memory, operating system, etc). This presentation will provide a WG progress report with special emphasis on the concept architecture, key features such as event management, and an early view of the development plan. The discussion will also cover how Fedora preservation services relate to PREMIS and the audit checklist for Trusted Digital Repositories.

Stuart Barr

*Project Leader, Remote Sensing Infrastructure and Engineering,
Geospatial and Earth Monitoring Division, Geoscience Australia*

Australia's remotely sensed data archive - the next 25 years and beyond

The Australian Centre for Remote Sensing holds an unbroken archive of remotely sensed data covering all of Australia spanning in excess of 25 years. Sound practices initiated in the early years and maintained since then ensure that the entire ACRES archive remains accessible. Measurement of long-term temporal change of the Australian landscape was made possible through this archive; inevitably there will be other uses in the future for this unique archive that we cannot predict. This presentation discusses how a review of the audit checklist is being used to review practices to ensure that the archive remains accessible indefinitely.

Stephen Ellis

*Assistant Director-General, Digital Government Branch,
National Archives of Australia*

The Trusted Digital Repositories Checklist: Government Records & Archives Aspect

The National Archives of Australia must serve a number of key stakeholders through its programs for preserving digital records. This presentation discusses these key stakeholders, their expectations and the National Archives aims in meeting them, including a number of special considerations influencing digital preservation in the archives context. The presentation also describes the current management of digital records held by the National Archives, discusses the certification process and notes a number of key issues for compliance with the audit checklist. Possible enhancements to the audit checklist are also discussed.

Sophie Holloway

*National Manager, Australian Social Science Data Archive,
Australian National University*

ASSDA: A Trusted Digital Repository or a trusted digital repository?

Since 1981, The Australian Social Science Data Archive (ASSDA) at the Australian National University has been trusted by its designated community to preserve and redistribute social science data according to the wishes of the data creators and the needs of the community. ASSDA exists within its designated community, staff are trained within that community and all developments are guided by community leaders. Trust in ASSDA is strong, with deposition often being automatic and stipulated as essential by granting bodies like ARC. However, when using the Audit Checklist for the Certification of Trusted Digital Repositories (TDR) ASSDA has many challenges ahead, perhaps the greatest of which involves the funding cycles intrinsic to the university sector. This paper seeks to provide a case study of a digital repository and its issues for compliance to a TDR audit.

Gerard Clifton

Manager, Digital Preservation, National Library of Australia

Trust and the Web: Can the Audit Checklist Apply to Web Archives?

An increasing number of National Libraries and similar institutions are building Web archives: collections of materials representing their national or cultural documentary heritage as it is expressed in the online environment. The practice of Web archiving usually involves a harvest model, rather than a deposit model, and poses a number of documentation and validation issues that impact on a Web archive's ability to comply with the RLG/NARA audit checklist for Trusted Digital Repositories. This presentation explores some of these issues and proposes some ways in which the checklist may be applied in the Web archiving context.

Markus Buchhorn

*Head, ANU Internet Futures,
Lead, APAC Information Infrastructure Program & APAC Grid Services Architect,
Grid Services Coordinator, GrangeNet*

What Does Partial Compliance Mean?

This presentation will tackle the concept of partial compliance from several angles. One angle is on the questions of measurement, metrics and methods. The other angles look at the implied value of (partial) compliance to consumers, providers, and funders.