open.michigan

Author(s): Paul Conway, Ph.D., 2010

License: Unless otherwise noted, this material is made available under the terms of the **Creative Commons Attribution–Noncommercial–Share Alike 3.0 License**: http://creativecommons.org/licenses/by-nc-sa/3.0/

We have reviewed this material in accordance with U.S. Copyright Law and have tried to maximize your ability to use, share, and adapt it. The citation key on the following slide provides information about how you may share and adapt this material.

Copyright holders of content included in this material should contact **open.michigan@umich.edu** with any questions, corrections, or clarification regarding the use of content.

For more information about **how to cite** these materials visit http://open.umich.edu/education/about/terms-of-use.

Any **medical information** in this material is intended to inform and educate and is **not a tool for self-diagnosis** or a replacement for medical evaluation, advice, diagnosis or treatment by a healthcare professional. Please speak to your physician if you have questions about your medical condition.

Viewer discretion is advised: Some medical content is graphic and may not be suitable for all viewers.





Citation Key

for more information see: http://open.umich.edu/wiki/CitationPolicy

Use + Share + Adapt

{ Content the copyright holder, author, or law permits you to use, share and adapt. }

PD-GOV Public Domain – Government: Works that are produced by the U.S. Government. (17 USC § 105)

Public Domain - Expired: Works that are no longer protected due to an expired copyright term.

PD-SELF Public Domain - Self Dedicated: Works that a copyright holder has dedicated to the public domain.

(cc) ZERO Creative Commons – Zero Waiver

(cc) BY Creative Commons – Attribution License

(cc) BY-SA Creative Commons – Attribution Share Alike License

(c) BY-NC Creative Commons – Attribution Noncommercial License

(cc) BY-NC-SA Creative Commons – Attribution Noncommercial Share Alike License

GNU-FDL GNU – Free Documentation License

Make Your Own Assessment

© FAIR USE

{ Content Open.Michigan believes can be used, shared, and adapted because it is ineligible for copyright. }

Public Domain – Ineligible: Works that are ineligible for copyright protection in the U.S. (17 USC § 102(b)) *laws in your jurisdiction may differ

{ Content Open.Michigan has used under a Fair Use determination. }

Fair Use: Use of works that is determined to be Fair consistent with the U.S. Copyright Act. (17 USC § 107) *laws in your jurisdiction may differ

Our determination **DOES NOT** mean that all uses of this 3rd-party content are Fair Uses and we **DO NOT** guarantee that your use of the content is Fair.

To use this content you should **do your own independent analysis** to determine whether or not your use will be Fair.

SI 640 DIGITAL LIBRARIES AND ARCHIVES

2010 Week 9: Metadata - OAIS and PREMIS

THEMES FOR THIS WEEK

- Administrative metadata
- Open Archival Information System
- PREMIS
- Integration of PREMIS and METS

ADMINISTRATIVE METADATA -- DRIGINS

- 1. Administrative
- 2. OAIS
- 3. PREMIS
- 4. Integration

- From 1998 on, metadata as the solution to nearly all digital preservation issues
- Administrative metadata supports content management from a variety of perspectives
 - Receiving content
 - Technical description
 - Quality assurance
 - Accountability
 - Changes made to content
- Models and standards preceded system development (just now catching up)

DAIS REFERENCE MODEL

- 1. Administrative
- 2. OAIS
- 3. PREMIS
- 4. Integration



Please see original image of at

Bruce Ambacher



The Consultative Committee for Space Data Systems

© FAIR USE CCSDS

- Origins in space science community
 - Why would space scientists need an archival standard?
- Very significant input from archivists
 - Bruce Ambacher of NARA
- Now an international standard
 - CCSDS 650.0-B-1 (blue book): Jan. 2002
 - ISO 14721:2003
 - Revisions being balloted (Sept. 2010)

OPEN ARCHIVAL INFORMATION

SYSTEM pen

- 1. Administrative
- 2. OAIS
- 3. PREMIS
- 4. Integration

- Reference Model standard(s) are developed using a public process and are freely available
- Information
 - Any type of knowledge that can be exchanged
 - Independent of the forms (i.e., physical or digital) used to represent the information
 - Data are the representation forms of information
- Archival Information System
 - Hardware, software, and people who are responsible for the acquisition, preservation and dissemination of the information
 - Additional OAIS responsibilities are identified later and are more fully defined in the Reference Model document

AIS INFORMATION DEFINITION =

A Dintormation is defined as any type of knowledge that can be exchanged, and this information is always expressed (i.e., represented) by some type of data

> In general, it can be said that "Data interpreted using its Representation Information yields Information"

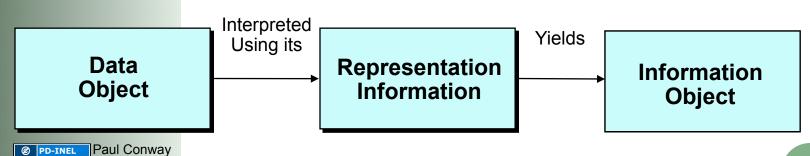
In order for this Information Object to be successfully preserved, it is critical for an archive to clearly identify and understand the Data Object and its associated Representation Information

1. Administrative

2. OAIS

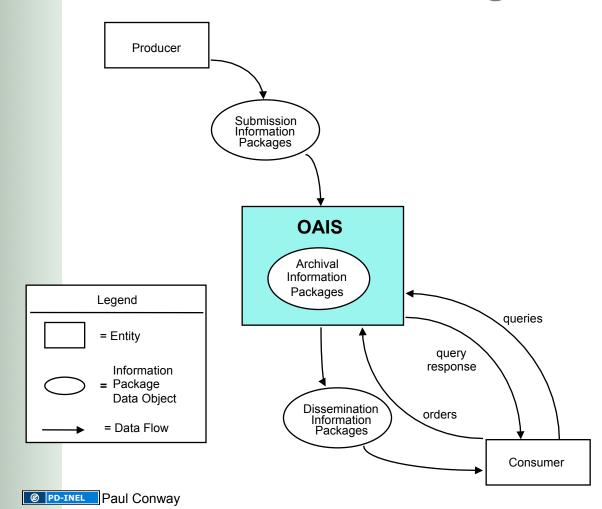
3. PREMIS

4. Integration



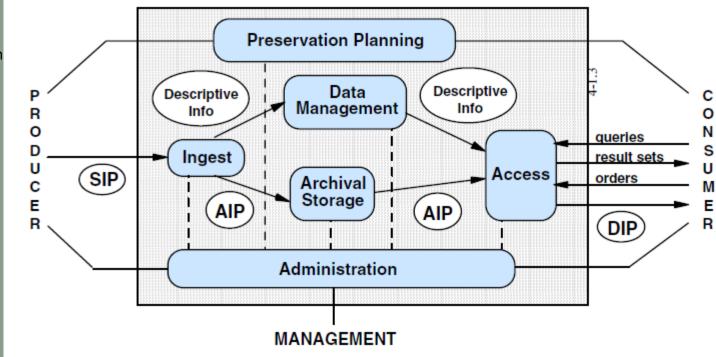
OAIS: External Data Flow Diagram

- 1. Administrative
- 2. OAIS
- 3. PREMIS
- 4. Integration



DAIS REFERENCE MODEL (SEC. 3-6)

- 1. Administrative
- 2. OAIS
- 3. PREMIS
- 4. Integration



Ø PD-INEL

Source Undetermined

OAIS-based E-Journal Archive Mapped to Harvard's Technical Infrastructure **Archive Producer** Administrative Administrative Administration registration and Admin DB Reporting interface profile Report Editorial Data Management Ingest Archival Storage Preservation process Planning Collect archive-Off-site Populate article-Transform Article catalog specific tape library SGML / XML level catalog / SIP to AIP collection mat claiming metadata optional PDF collection mat DB Automatic DB **Format** content replication registry Use SICI AIP Render page for name Issue description, Serial check-in if necessary **DRS** Preservation **Format** NRS/Admin issue claiming (repository) migration monitoring (naming) Archive (ILS) metadata Deposit Use DOI DRS deposit error for name Article page Access description (e.g., PDF) Consumer Deposit AIP deposit Transform Request for confirmation AIP to DIP archival data Transform to SIP acceptance archive DTD and normative formats Proof Proof read DIP **DIP** delivery articles read error Generate/ SIP Publisher Archive populate SIP article page Off-line article page entity set Content / metadata QC description description Syntactic and semantic Internal and external On-line SIP consistency Render SIP QC failure article page description (authentication/ authorization) Content / metadata Q¢ File-level QC: Article QC failure Syntactic and semantic Normalized Transform non-Rcvd all files, no (name resolution) website Internal and external SIP extras, checksum. normative to normative check issue manifest consistency against external A&I URN/URL Issue TOO SIP file error hash Aggregate and website, compress SIP METS file error Submission error Unpack, parse Client-side Server-side SIP Title/issue METS, extract submission submission **OPAC** manifest Ø PD-INEL Paul Conway

DAIS REFERENCE MODEL (SEC. 4-34)

- 1. Administrative
- 2. OAIS
- 3. PREMIS
- 4. Integration

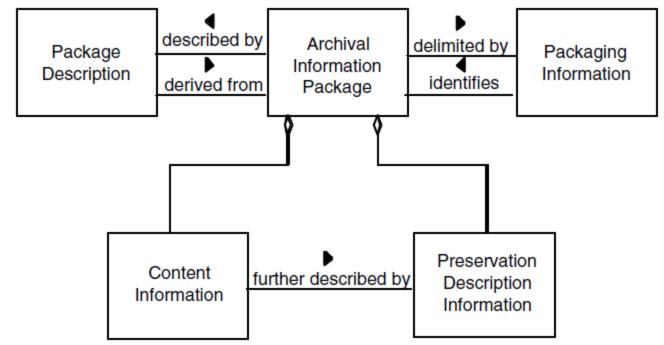


Figure 4-15: Archival Information Package (AIP)

RESERVATION DESCRIPTION NFORMATION

- Provenance Information
 - Describes the source of Content Information, who has had custody of it, what is its history
- Context Information
 - Describes how the Content Information relates to other information outside the Information Package
- Reference Information
 - Provides one or more identifiers, or systems of identifiers, by which the Content Information may be uniquely identified
- Fixity Information
 - Protects the Content Information from undocumented alteration

PRESERVATION METADATA

- 1. Administrative
- 2. OAIS
- 3. PREMIS
- 4. Integration

- Garrett & Waters (1996)
 - Components of information integrity
 - Content, fixity, reference, provenance, context
- OAIS Reference Model (2002)
 - P. 138: abandoned "content" and added "packaging" concept
- Multiple irreconcilable efforts to implement OAIS preservation metadata model, for example:
 - CEDARS (UK) Guide to Pres. Meta. (2002)
 - KB (Netherlands) model (2003)
 - British Library specification (2004)

REMIS WORKING GROUP

- 1. Administrative
- 2. OAIS
- 3. PREMIS
- 4. Integration

- June 2003: OCLC, RLG sponsored new international working group:
 - PREMIS: Preservation Metadata: Implementation Strategies
- Membership:
 - > 30 experts from 5 countries, representing libraries, museums, archives, government agencies, and the private sector
 - Co-Chairs: Priscilla Caplan (FCLA), Rebecca Guenther (LC)
- Objective 1: Identify and evaluate alternative strategies for encoding, storing, managing, and exchanging preservation metadata
 - PREMIS Survey Report (September 2004)
 - Snapshot of current practices/emerging trends related to managing and using preservation metadata in digital archiving systems
 - http://www.oclc.org/research/projects/pmwg/surveyreport.pdf
- Objective 2: Define implementable, core preservation metadata, with guidelines/recommendations for management and use

PREMIS DATA DICTIONARY

- 1. Administrative
- 2. OAIS
- 3. PREMIS
- 4. Integration

- May 2005: Data Dictionary for Preservation

 Metadata: Final Report of the PREMIS Working Group
 - 237-page report includes:
 - PREMIS Data Dictionary 1.0
 - o Context/assumptions, data model, usage examples
 - Set of XML schema to support implementation

Data Dictionary:

- Comprehensive view of information needed to support digital preservation
 - Guidelines/recommendations to support creation, use, management
- Based on deep pool of institutional experiences in setting up and managing operational capacity for digital preservation
- Received the 2005 Digital Preservation Award (UK) and 2006 Society of American Archivists Publication Award



http://www.loc.gov/standards/premis/



SCOPE

- 1. Administrative
- 2. OAIS
- 3. PREMIS
- 4. Integration

• What PREMIS DD is:

- Common data model for organizing/thinking about preservation metadata
- Guidance for local implementations

PRESERVATION METADATA

• Standard for exchanging information packages between repositories

• What PREMIS DD is not:

- Out-of-the-box solution:
 - need to instantiate as metadata elements in repository system
- All needed metadata:
 - excludes business rules, format-specific technical metadata, descriptive metadata for access, non-core preservation metadata
- Lifecycle management of objects outside repository
- Rights management:
 - limited to permissions regarding actions taken within repository



IAIS REFERENCE MODEL AND

- 1. Administrative
- 2. OAIS
- 3. PREMIS
- 4. Integration

- OAIS reference model specifies the Preservation Description Information (PDI)
- Description Information (PDI)

 PREMIS used the OAIS information model as a starting point

 PREMIS Data Dictionary developed the conceptual types of information objects into more than 100 semantic units.

 PREMIS Data Dictionary provided detailed descriptions and guidelines to implement these semantic units.

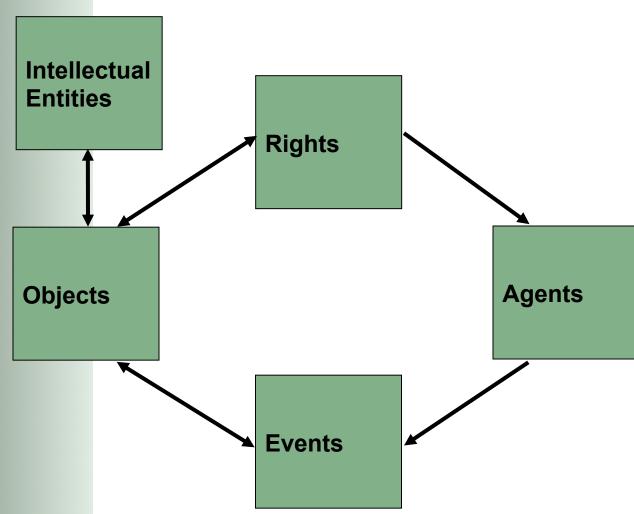
 All entities have reference (identification) information.

 PREMIS deals mostly with representation, context,

- provenance, and fixity information, in keeping with PREMIS definition of preservation metadata.

REMIS DATA MODEL

- 1. Administrative
- 2. OAIS
- 3. PREMIS
- 4. Integration



Fall 2010



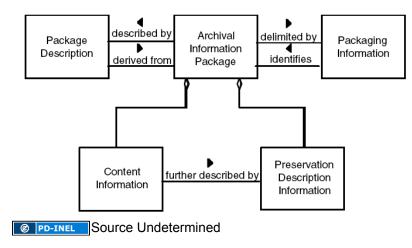
REMIS XML SCHEMAS

- 1. Administrative
- 2. OAIS
- 3. PREMIS
- 4. Integration

- One schema for each PREMIS entity in data model
 - Allows user to choose which parts of PREMIS to use
- PREMIS container schema
 - References schema for each entity type
 - Provides a container if it is desirable to keep some or all PREMIS metadata together
 - If using container requires at least an object which in turn requires objectIdentifier and objectCategory
 - Individual schemas may used alone or with container
- Semantic units in PREMIS schemas
 - XML is faithful to data dictionary
 - Only those units mandatory for all categories of objects are mandatory in object schema

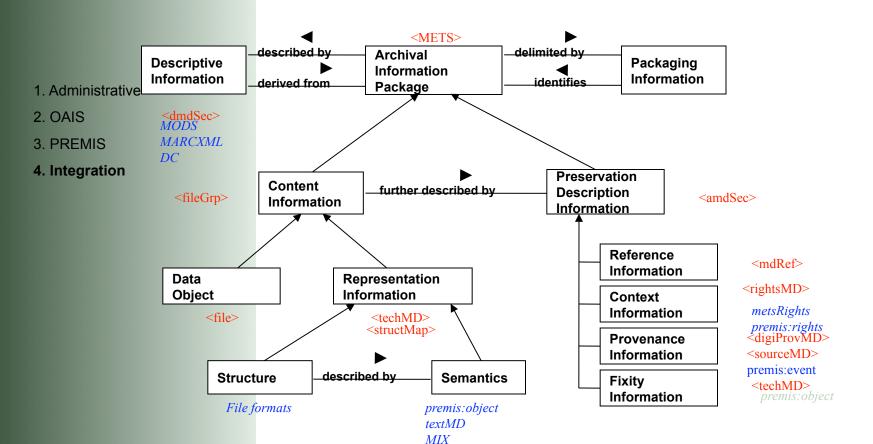
CONTAINER FOR XML APLEMENTATION

- 1. Administrative
- 2. OAIS
- 3. PREMIS
- 4. Integration



- Archival Information Package (AIP) may include much more metadata besides the preservation metadata
- A well defined container is usually necessary to group and appropriately associate these metadata with the data object
- For example: METS or MPEG-21 DID

OAIS and METS



Legend

Black Arial = OAIS

Red Times New Roman = METS Primary Schema

Blue Times New Roman Italics = Extension Schema

SSUES IN USING PREMIS WITH AFTS

- 1. Administrative
- 2. OAIS
- 3. PREMIS
- 4. Integration

- Which METS sections to use and how many
- Whether to record elements redundantly in PREMIS that are defined explicitly in the METS schema
- How to record elements that are also part of a format specific technical metadata schema (e.g. MIX)
- Recording structural relationships
- How to deal with locally controlled vocabularies
- Whether to use the PREMIS container

PREMIS AND METS SECTIONS

- 1. Administrative
- 2. OAIS
- 3. PREMIS
- 4. Integration

- Flexibility of METS requires implementation decisions
- You can't put all PREMIS metadata directly under amdSec
- What sections to use for PREMIS metadata?
 - Alternative 1
 - Object in techMD
 - Event in digiProvMD
 - Rights in rightsMD
 - Agent with event or rights
 - Alternative 2
 - Everything in digiProvMD
 - Alternative 3
 - Everything in techMD
- How many administrative MD sections to use?
- Experimentation will result in best practices

PREMIS IN METS ASSIGNMENT

- 1. Administrative
- 2. OAIS
- 3. PREMIS
- 4. Integration

- See assignment guidelines in Ctools
 - PowerPoint presentation with examples
- Use PREMIS Data Dictionary
 - http://www.loc.gov/standards/premis/
 - Page 130 ff
- Some information must be invented
 - Absence of local controlled vocabularies
- Consider using the PREMIS tools
 - http://www.loc.gov/standards/premis/tools for premis.php
- Goal of exercise is reading and interpreting the standards, not creating perfect XML

Thank you!

Paul Conway

Associate Professor
School of Information
University of Michigan
www.si.umich.edu

Additional Source Information

for more information see: http://open.umich.edu/wiki/CitationPolicy

Slide 6: ISO, http://www.iso.org/iso/home.html; Please see original image of at Bruce Ambacher at

http://ischool.umd.edu/content/bruce-i-ambacher; CCSDS, http://public.ccsds.org/default.aspx

Slide 8: Paul Conway

Slide 9: Paul Conway

Slide 10: Source Undetermined

Slide 11: Paul Conway

Slide 12: Source Undetermined

Slide 15: US Government, http://www.loc.gov/standards/premis/

Slide 16: US Government, http://www.loc.gov/standards/premis/; US Government, http://www.loc.gov/standards/premis/

Slide 17: US Government, http://www.loc.gov/standards/premis/

Slide 18: US Government, http://www.loc.gov/standards/premis/

Slide 19: US Government, http://www.loc.gov/standards/premis/

Slide 20: US Government, http://www.loc.gov/standards/premis/

Slide 19: Paul Conway

Slide 20: US Government, http://www.loc.gov/standards/premis/

Slide 21: US Government, http://www.loc.gov/standards/premis/; Source Undetermined

Slide 22: Paul Conway