# 3. Technical and administrative metadata standards

Metadata Standards and Applications

# Goals of session

- To understand the different types of administrative metadata standards
- To learn what types of metadata are needed for digital preservation
- To learn the importance of technical, structural and rights metadata in digital libraries

# Types of administrative metadata

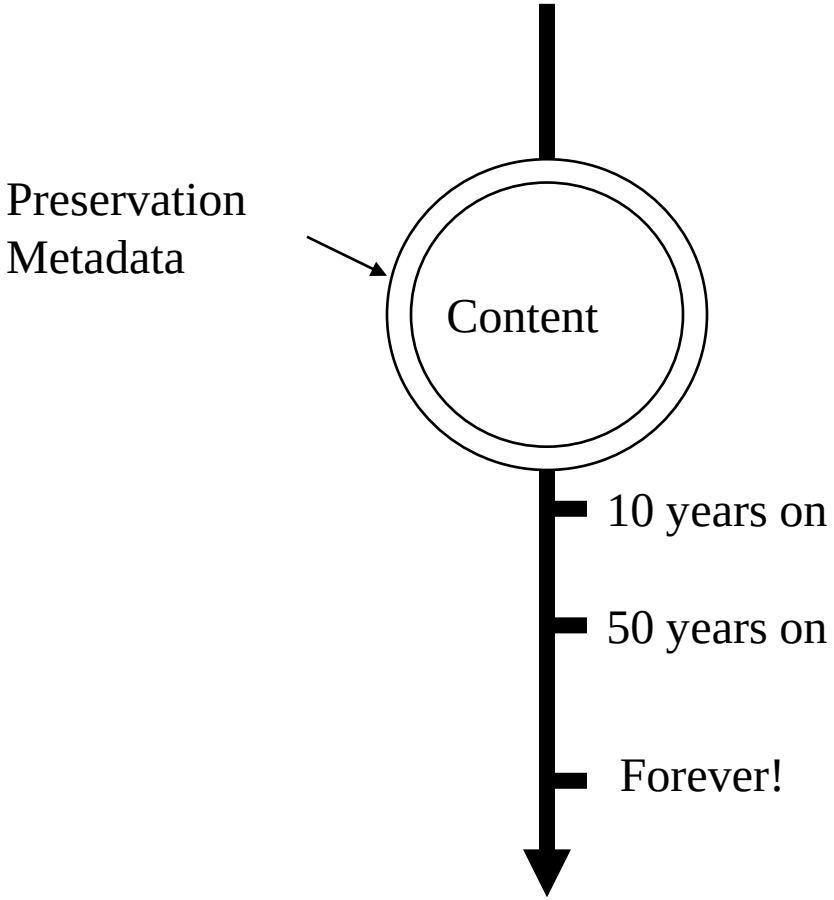
- Provides information to help manage a resource or the metadata about a resource
  - Preservation metadata
    - Technical characteristics
    - Information about actions on an object
  - Structural metadata may be considered administrative; indicates how compound objects are put together
  - Rights metadata
    - Access rights and restrictions
    - Preservation rights and restrictions

# Preservation Metadata: PREMIS

- PREMIS is:
  - Common data model for organizing/thinking about preservation metadata
  - Guidance for local implementations
  - Standard for exchanging information packages between repositories
- http://www.loc.gov/standards/premis/

# Preservation Metadata Includes:

- Provenance
  - Who has had custody/ownership of the digital object?
- Authenticity:
  - Is the digital object what it purports to be?
- Preservation Activity:
  - What has been done to preserve it?
- Technical Environment:
  - What is needed to render and use it?
- Rights Management:
  - What IPR must be observed?




**Makes digital objects self-documenting across time**

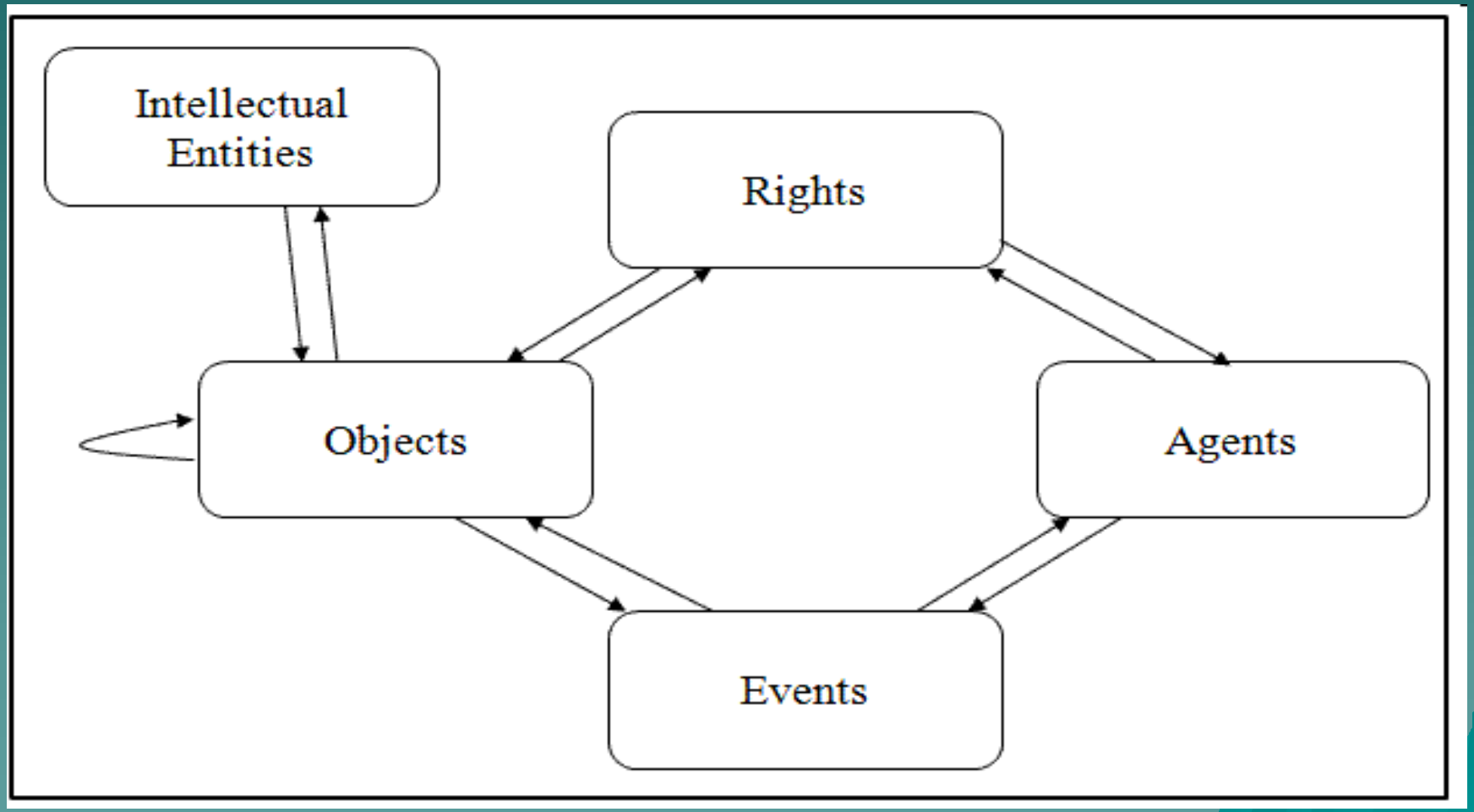
# PREMIS Data Dictionary

- PREMIS Data Dictionary for Preservation Metadata version 2.0
  - Detailed description of metadata elements
  - Introduction and supporting documentation
    - Guidelines to support implementation, use, management
  - Entity Hierarchical Listing
  - Based on institutional experiences in managing operational capacity for digital preservation
- Set of XML schema developed to support use of Data Dictionary

# What PREMIS is not …

- An "Out-of-the-box" solution: must be instantiated as metadata elements in repository system
- All encompassing: excludes business rules, format-specific technical metadata, descriptive metadata for access, non-core preservation metadata
- Lifecycle management of objects outside the repository
- A rights management solution: limited to permissions regarding actions taken within repository

# PREMIS Data Model



# Type of information covered in PREMIS (by entity type)

- Object
  - Object ID
  - Preservation level
  - Object characteristics (format, size, etc.)
  - Storage
  - Environment
  - Digital signatures
  - Relationships
  - Linking identifiers
- Event
  - Event ID
  - Event type
  - Event date/time
  - Event outcomes
  - Linking identifiers
- Agent
  - Agent ID
  - Agent name
- Rights
  - Rights statement
  - Granting Agent
  - Permission granted

# Why is PREMIS important to catalogers?

- As we take responsibility for more digital materials, we need to ensure that they can be used in the future
- Most preservation metadata will be generated from the object, but catalogers may need to verify its accuracy
- Catalogers will play a role in assessing and organizing digital materials, and will need to:
  - Understand the structure of complex digital objects
  - Determine significant properties that need to be preserved

# Technical metadata for images

- NISO Z39.87 andMIX
- Adobe and XMP
- EXif
- IPTC/XMP

- Some of these deal with embedded metadata in images

# Metadata for Images in XML (MIX)

- An XML Schema designed for expressing technical metadata for digital still images
- Based on the NISO Z39.87 Data Dictionary – Technical Metadata for Digital Still Images
- Can be used standalone or as an extension schema with METS/PREMIS

# Using MIX

- Includes characteristics that apply to all or most object types, e.g. size, format
- Format includes specific metadata for images, for example:
  - Image width
  - Color space, color profile
  - Scanner metadata
  - Digital camera settings
- http://www.loc.gov/standards/mix/

# Technical Metadata for Textual Objects (textMD)

- An XML Schema designed for expressing technical metadata for textual objects
- Developed at New York University; maintenance transferred to Library of Congress
- Includes format specific technical metadata for text
  - Byte order
  - Character set encoding
  - Font script … etc.
- http://www.loc.gov/standards/textMD/

# Technical Metadata for Multimedia (MPEG-7)

- A multimedia content description standard, associated with the content itself
  - Intended to allow fast and efficient searching
- Formally called Multimedia Content Description Interface
  - Does not deal with the actual encoding of moving pictures and audio (as MPEG-1, MPEG-2 and MPEG-4 do)
  - intended to provide complementary functionality to the previous MPEG standards

# MPEG-7: Description vs. Content

- Requirement that description must be separate from the audiovisual content
- Uses:
  - Descriptor (D): a representation of a feature defined syntactically and semantically.
  - Description Schemes (DS): Specifies the structure and semantics of the relationships between its components
  - Description Definition Language (DDL): an XML-based language to define the syntax rules

# Other Technical Metadata for Audio and Video

- LC developed XML technical metadata schemas for LC Audiovisual Prototype Project; these were widely implemented because of the lack of other schemas
- Audio and video technical metadata schemas under development by expert organizations
- Moving Image Collections (MIC) project is also experimenting with these:
  - http://mic.loc.gov/
- For more information on LC schemas:
  - http://www.loc.gov/rr/mopic/avprot/metsmenu2.html

# Structural Metadata

- Supports the intended presentation, use, and navigation of an object
- Binds the parts together; expresses relationships between parts of a multipart object
- Examples of structural metadata expressions:
  - EAD hierarchical structure
  - METS structMap
  - PREMIS relationship elements

# Rights Metadata

- Rights schemas with limited scope
- Rights Expression Language (REL) for managing intellectual property rights, particularly by rights owners
- Rights information is not well understood
  - Different laws in different jurisdictions
  - Machine actionable vs. human understandable
- Rights take different forms
  - Legal statutes, e.g. copyright
  - Contractual rights, e.g. licenses

# Rights Schema Examples

- Creative Commons
  - Allows creators to choose a license for their work
  - Simple rights statements that fit a lot of situations
  - http://creativecommons.org/
- METS Rights
  - Access rights for use with METS objects
  - Rights declarations
  - Rights holder
  - Context

# More Rights Schema Examples

- PLUS for images
- MPEG-21 REL for multimedia
- ONIX for licensing terms
- XRML/MPEG-21
- ODRL (Open Digital Rights Language)
- For a fuller discussion of rights languages, see the report written by Karen Coyle for the Library of Congress:
  - http://www.loc.gov/standards/relreport.pdf

# Exercise

- Provide administrative or technical metadata for the object used in the descriptive metadata exercise.