



Digital**Preservation**Coalition

# Preservation metadata: an overview

Richard Gartner  
23<sup>rd</sup> April 2013

Richard Gartner

Centre for e-Research  
Department of Digital Humanities

**KING'S**  
*College*  
**LONDON**

# What is metadata?

μετα = about or  
above

Meta-language

Meta-theory

Metadata = data about data

# Types of metadata

## Descriptive

Information on the intellectual content of an item

## Administrative

Information to enable the management of the item

## Structural

Information to enable the internal structure of an item and the relationships between its components to be rendered in a meaningful way

# Preservation Metadata

*“metadata that supports the process of long-term digital preservation...”*

*Specifically maintaining the:*

*Availability*

*Identity*

*Persistence*

*Renderability*

*Understandability*

*Authenticity of digital objects*

*over long periods of time”*

*Lavoie and Gartner, 2013*

Preservation metadata establishes a *context* around a preserved digital object that remains attached to the object over time.

### ***Provenance***

Information describing history of digital resource since its creation/capture

### ***Preservation Activity***

Information on the actions that have been taken to preserve the object over time

### ***Technical and interpretative environment***

Information on the technical requirements needed to access, render, and use the object

Information needed to understand or interpret the object and its contents

# PREMIS: the key preservation metadata standard

# Metadata standards

Caplan: a set of metadata elements and rules for their use which have been designed for a particular purpose\*

Collections: EAD (Encoded Archival Description)

Bibliographic data: MARC (MACHine Readable Cataloguing)

Anything (particularly electronic): DC (Dublin Core)

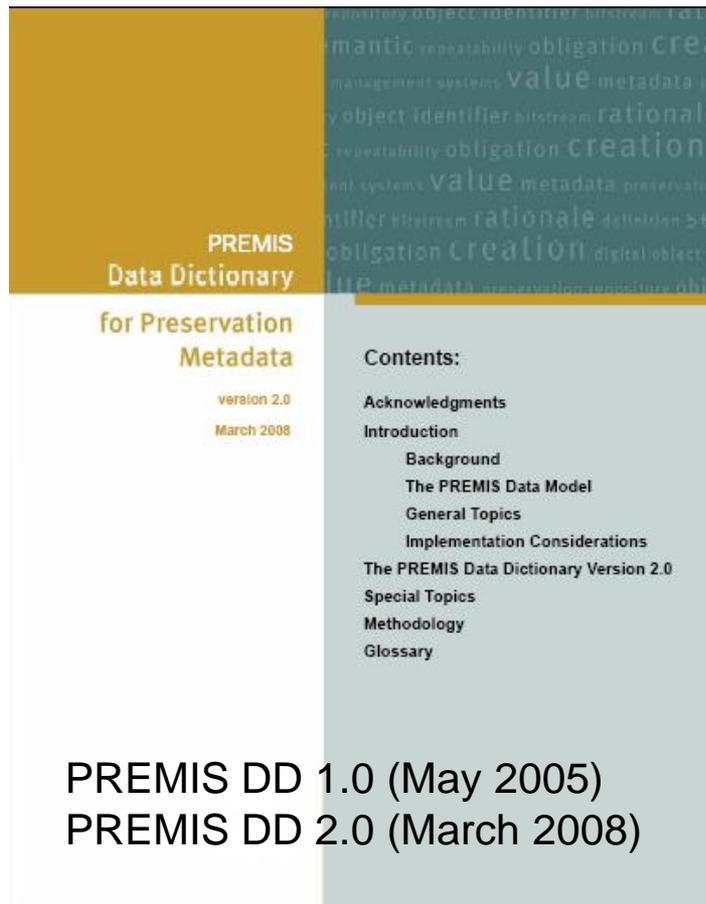
Technical information on an image: MIX (Metadata for Images in XML Schema)

Intellectual property rights: METSRights

Information for preservation of digital objects: PREMIS (PREservation Metadata: Implementation Strategies)

\*Caplan, Priscilla (2003) *Metadata fundamentals for all librarians*, Chicago, American Library Association.

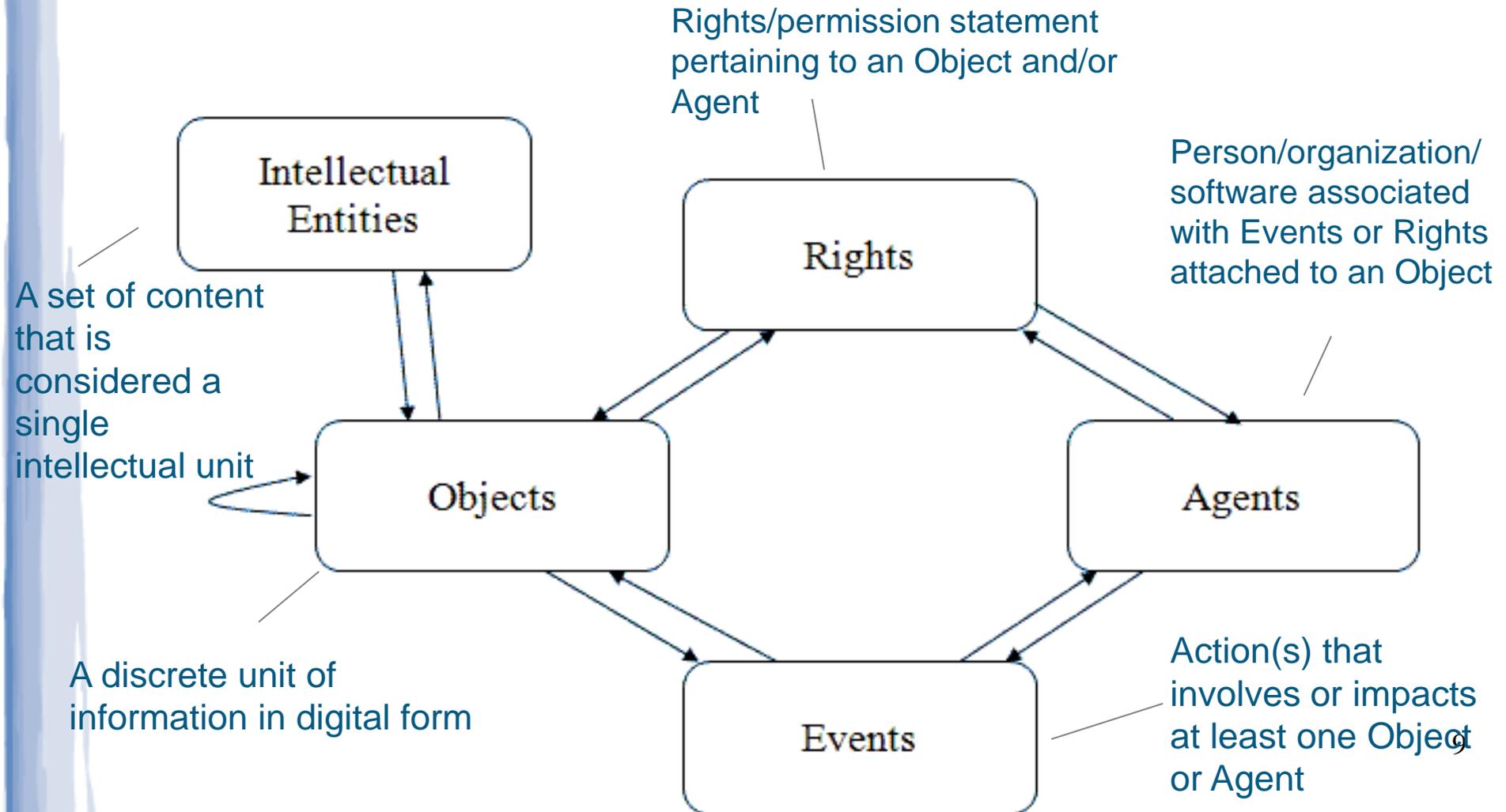
# Preservation Metadata Implementation Strategies (PREMIS)



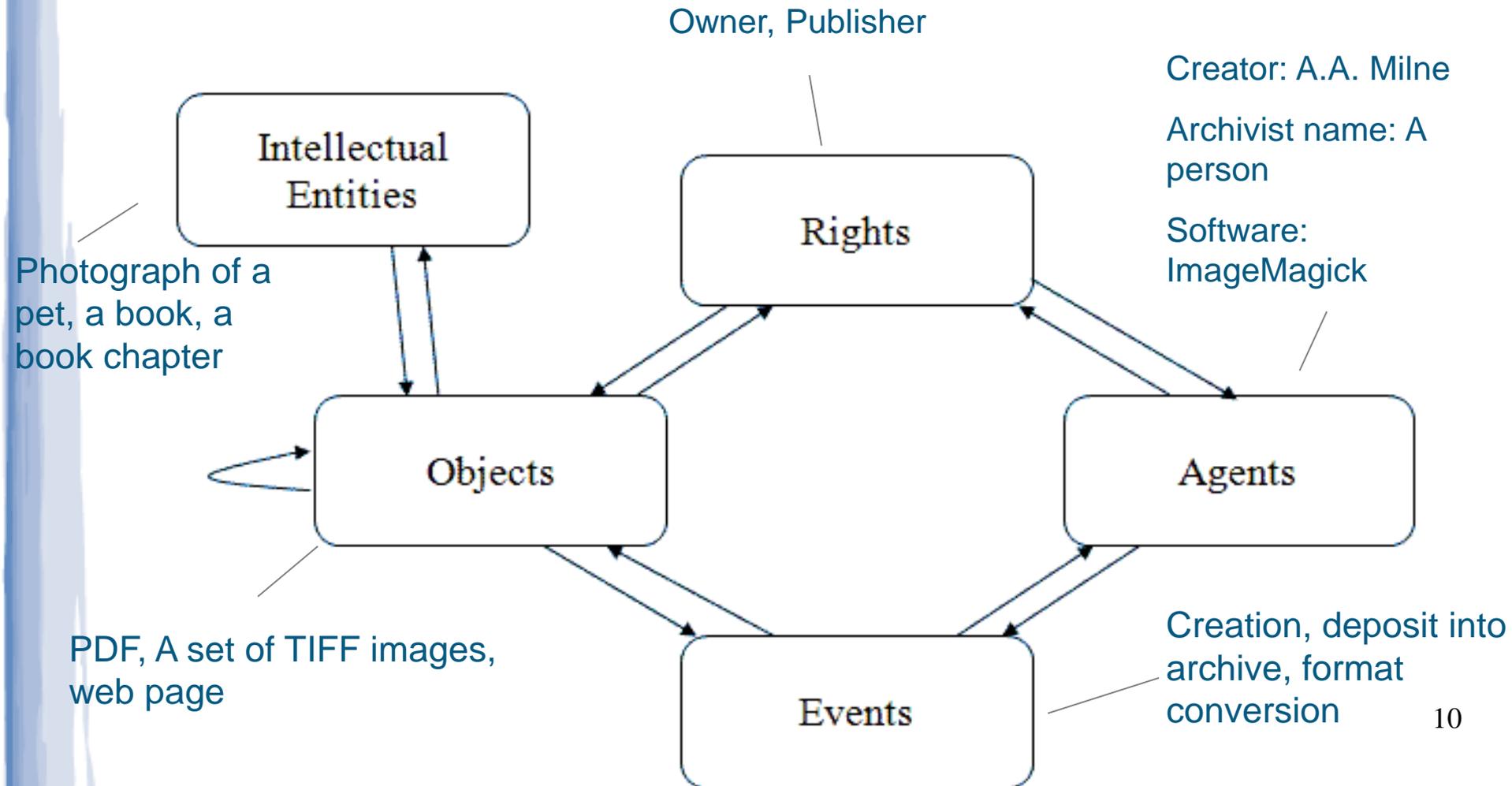
PREMIS DD 1.0 (May 2005)  
PREMIS DD 2.0 (March 2008)

- Data dictionary establishes core metadata - "things that most working repositories are likely to need to know in order to support digital preservation"
- Define an implementable set of "core" preservation MD elements
- Data Dictionary - Implementation independent, i.e. does not define how information should be stored
- XML schema has been developed

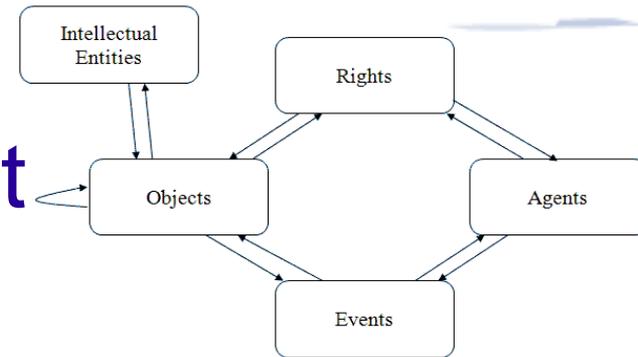
# PREMIS Data Model: Entities



# PREMIS Data Model: Example



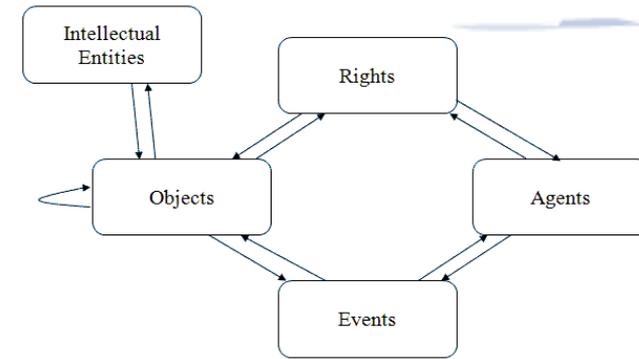
# PREMIS: Object



Technical properties that are applicable to all/most formats:

1. *Format*: Format name, version & format registry
2. *Size*: Size in bytes
3. *Fixity*: Info to verify if an object has been altered in an undocumented or unauthorized way.
4. *Creating Application*: Name, version, date data was created
5. *Inhibitors*: Features intended to inhibit access, use, or migration.

# PREMIS Events



- Aggregates information about actions that involves one or more Objects
- Metadata about an Event would normally be recorded and stored separately from the digital object.

# Audit trail / Provenance

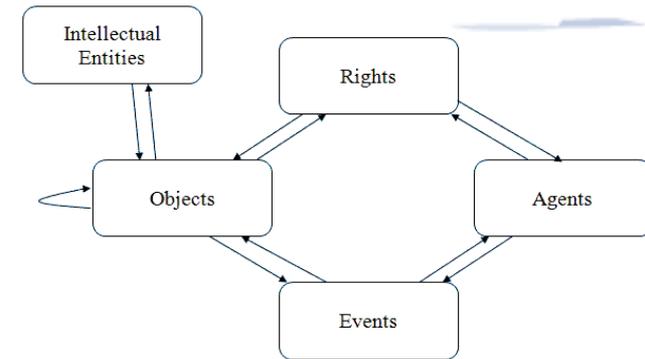
“Digital provenance documents the origin and chain of custody of a digital object, and any important events in the object's history.”

(Priscilla Caplan. 2006)

## Questions

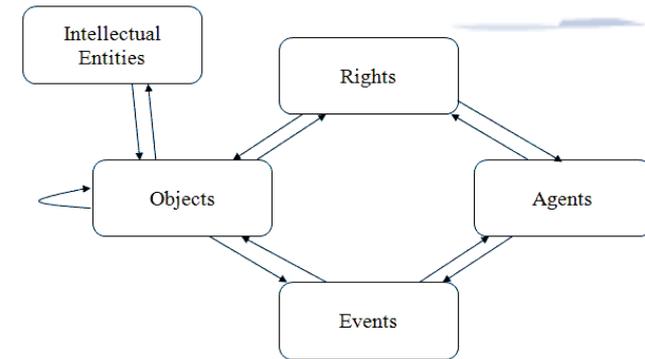
- “What actions have been performed on the resource?”
- “Who performed them and why?”
- “When could an error have been introduced into the management process?”

# PREMIS Events



Capture	obtain data from source
Fixity check	verify that an object has not changed in a given period
Migration	transformation of an object to a format suitable for distribution
Normalization	transformation of an object creating a version more conducive to preservation
Virus check	the process of scanning a file for malicious programs
Compression	Code data for transmission or save space
Deaccession	Remove data from system

# PREMIS Agent



Agents are actors that have roles in events and in rights statements

Agents can be people, organizations, or software applications.

# PREMIS Rights

Copyright Information

Licence Information

Statute Information

Rights Granted

# Packaging with other metadata

Preservation metadata requires careful packaging for storage and linkage to descriptive (resource discovery), administrative, and structural metadata



# PREMIS tools

Extracting technical metadata from objects

Converting extracted metadata into the PREMIS XML schema elements

Generating a METS object with appropriate slots for PREMIS metadata

Converting output from other tools (eg JHOVE) to PREMIS elements

Recording events and outcomes (e.g. format validation, fixity check, etc.)

## PREMIS Implementation Registry

To have your projects added to the PREMIS registry, or to edit an existing registry entry, please submit the requisite information directly to the Network Development and MARC Standards Office at [ndmso@loc.gov](mailto:ndmso@loc.gov).

### Search the PREMIS Implementation Profiles

[View the entire PREMIS registry](#) OR:

Browse the Index by Project Name

- Select a Project --
- Archivematica
- Carolina Digital Repository
- Creating a digital repository at the Swedish National Archives using PREMIS
- Digitaal Magazijn
- Digital Data Archive (DDA) Project
- Digital Library of Castile and Leon/Biblioteca Digital de Castilla y León
- Digital Library of Castile-La Mancha/Biblioteca Digital de Castilla-La Mancha
- Digital Library of the Royal Academy of History
- Digital Library of the Royal Academy of Pharmacy
- Electronic Records Archives (ERA) Project
- European project SHAMAN
- FCLA Digital Archive and DAITSS
-  Galiciana: Biblioteca Digital de Galicia/Digital Library of Galice
-  HathiTrust shared digital repository
-  Hub and Spoke Architecture for Repository Interoperability and Preservation
-  Iuris Digital
- Keeping Emulation Environments Portable (KEEP)
- Kramerius
- MathArc

[Contact Us](#)

Next....

A practical exercise in preservation  
metadata

but first:  
any questions?