PREMIS Deep Dive – Personal View
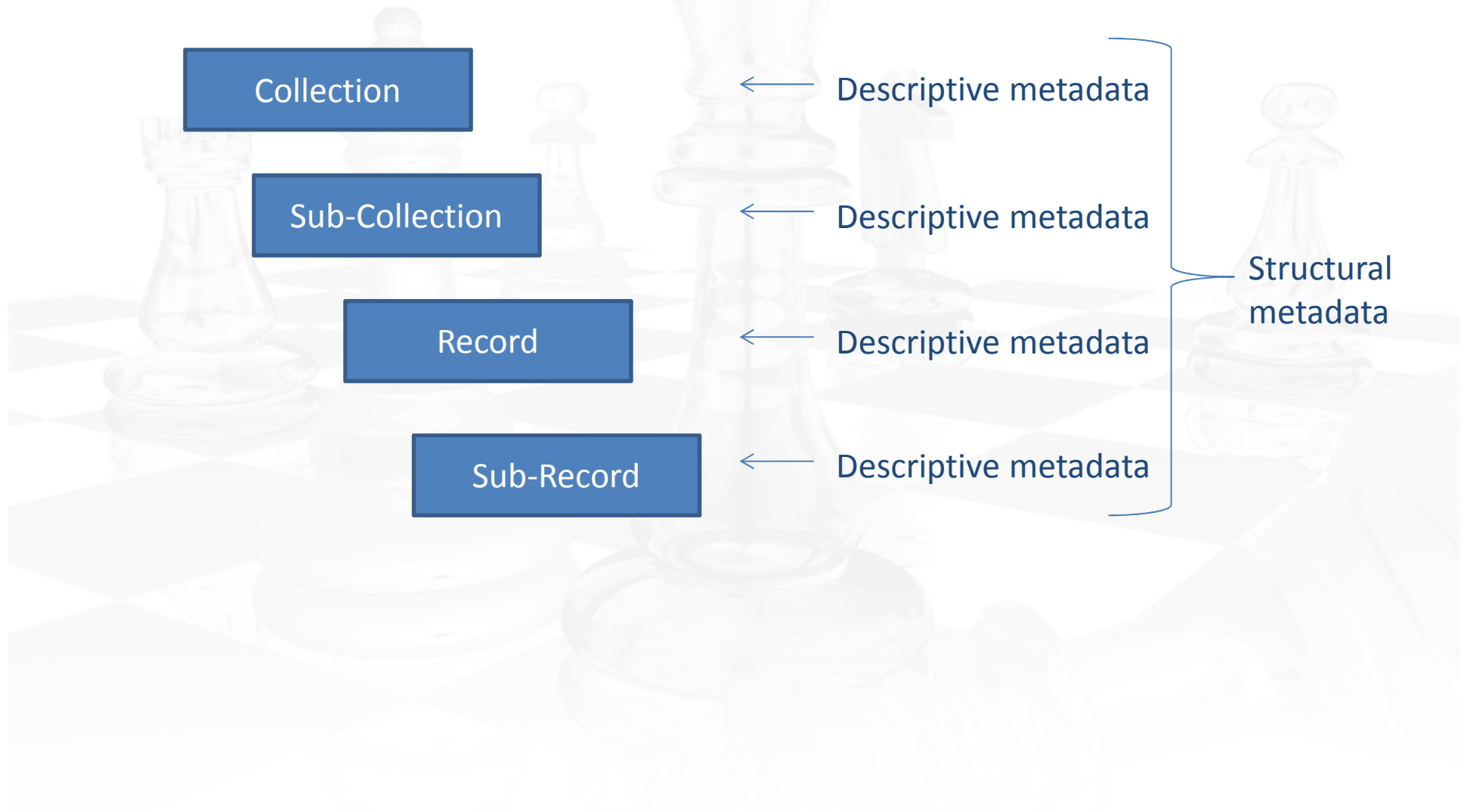
Robert Sharpe, Tessella

23rd April 2013

What is PREMIS?

- digital PREservation Metadata
- PREMIS is an Information Model:
  - "The information a repository uses to support the digital preservation process"
  - "Things that <u>most</u> working preservation repositories are <u>likely</u> to need to know to support digital preservation"

Tessella

Contents

- Basics of PREMIS
- PREMIS Conformance
- PREMIS practicalities
- Where next?

Tessella

PREMIS Basics - Always had intellectual entities

Collection ← Descriptive metadata

Sub-Collection ← Descriptive metadata

Record ← Descriptive metadata

Sub-Record ← Descriptive metadata

Structural metadata

Tessella

PREMIS Basics - But now have digital files…

Records of XYZ Committee

1995 – Word Perfect	← Technical metadata
2000 – Word 97	← Technical metadata
2005 – Word 2002	← Technical metadata
2010 – Word 2010	← Technical metadata

- Fixity
- Size
- Format
  - Version
- Rendering software
- Embedded images:
  - Fixity
  - Size
  - Format
    - Version
- …

Tessella

PREMIS Basics - And need to do more things...

Records of XYZ Committee

Original Representation

Migrated Representation

1995 – Word Perfect

2000 – Word 97

2005 – Word 2002
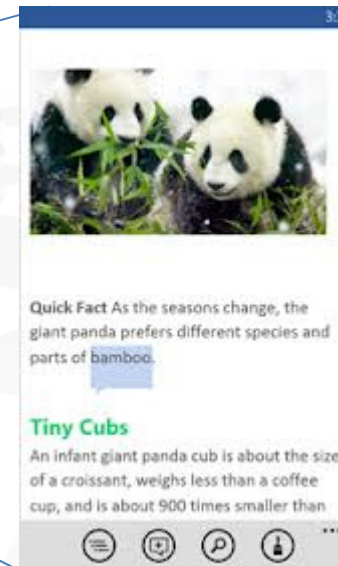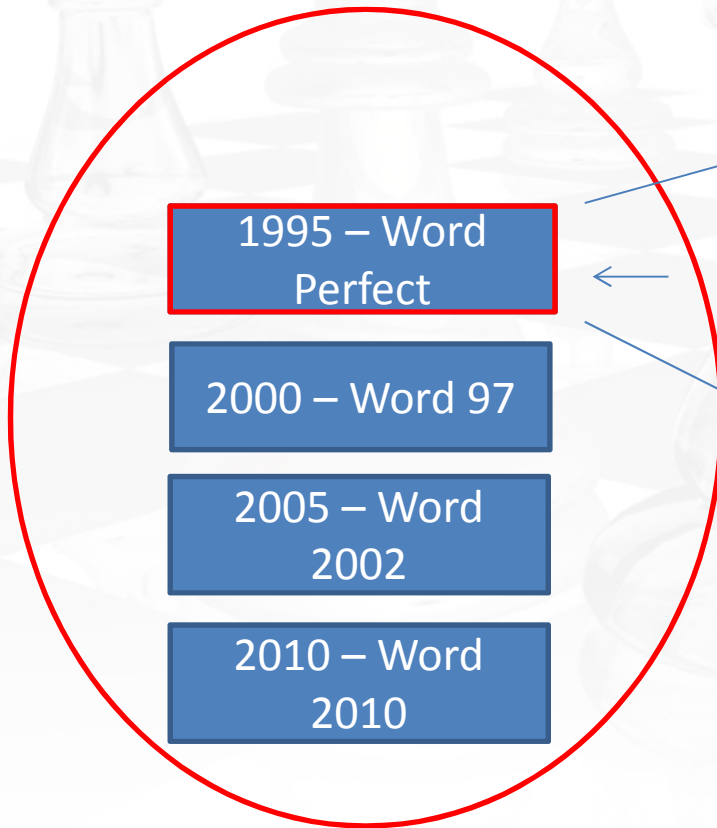
2010 – Word 2010

1995 – PDF/A

2000 – PDF/A

2005 – PDF/A

2010 – PDF/A

Tessella

PREMIS Basics - 3 types of Object

Representation

1995 – Word Perfect

2000 – Word 97

2005 – Word 2002

2010 – Word 2010

File

Bistream

Quick Fact As the seasons change, the giant panda prefers different species and parts of bamboo.

Tiny Cubs
An infant giant panda cub is about the size of a croissant, weighs less than a coffee cup, and is about 900 times smaller than

Tessella

PREMIS Basics - Objects

- PREMIS Objects:
  - Representation
  - File
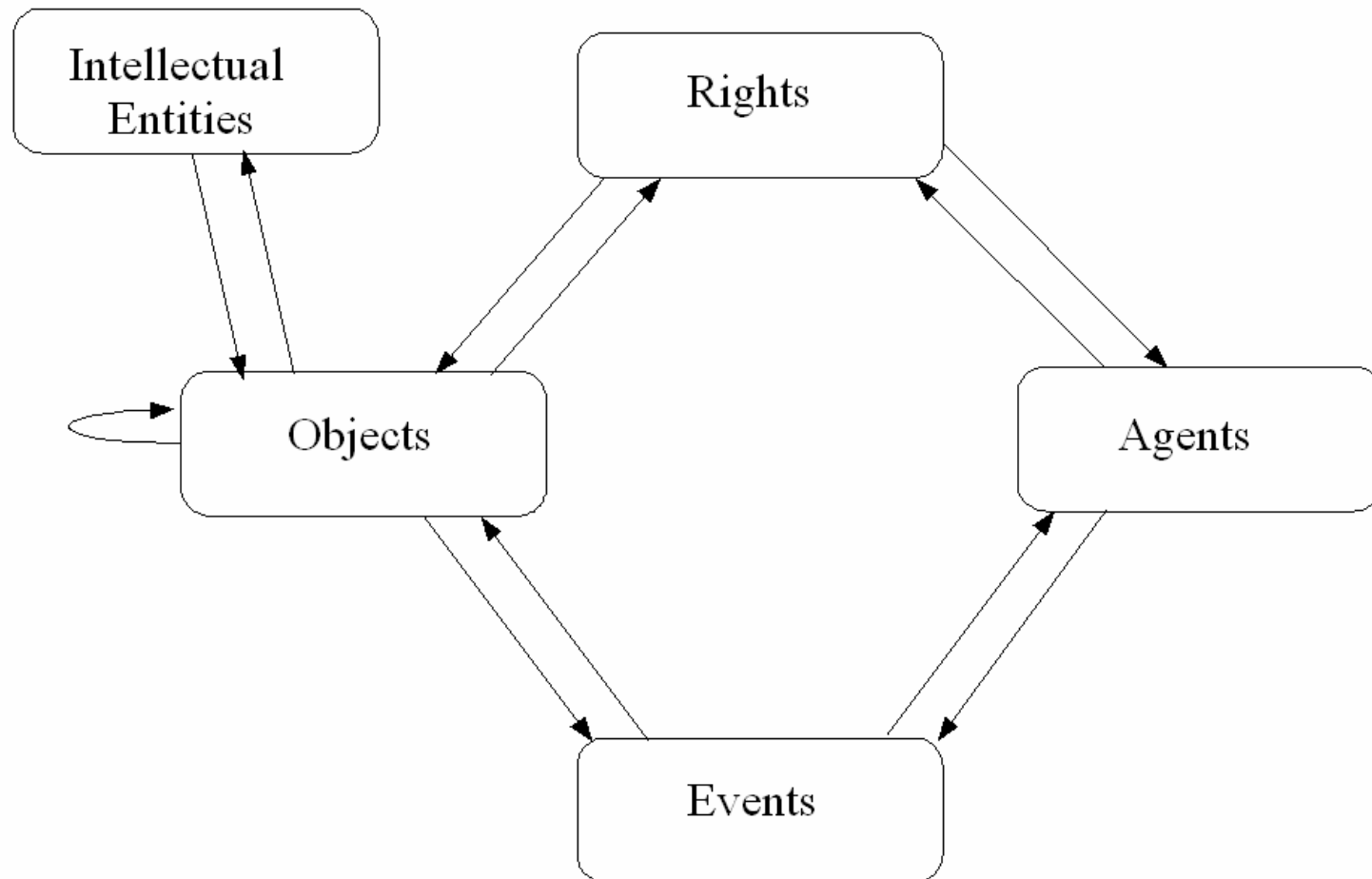 - Bitstream
- Define common properties

Tessella

PREMIS Basics - Object properties 1/2

- Identifier
- Category (Representation, File, Bitsteam)
- Preservation level
- Significant properties:
    - Type (e.g., page count)
    - Value (e.g., 7)
- Characteristics:
    - Fixity
    - Size (bytes)
    - Format (Designation, Registry, Note)
    - Creating application
    - …
- Original name
- Storage

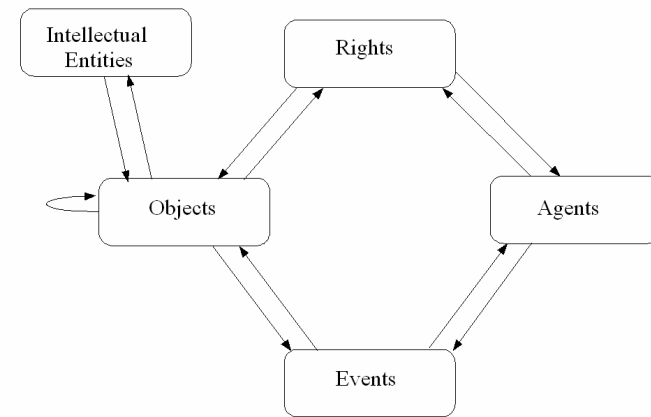Tessella

PREMIS Basics - Object properties 2

- Environment:
  - ...
    - Software
    - Hardware
  - ...
- Signature
- Relationship
- Linked events
- Linked intellectual entity
- Linked rights statement

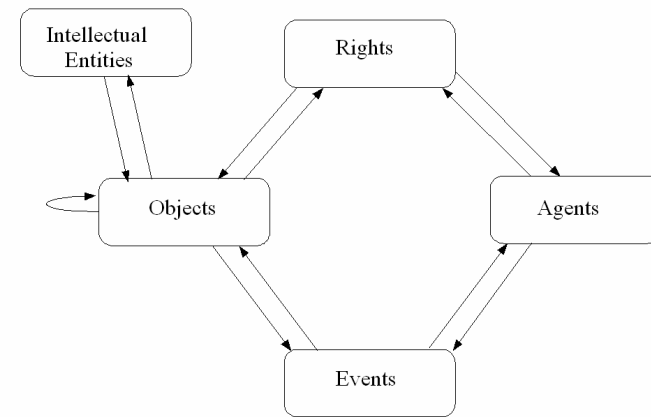Tessella

PREMIS Basics - More than Objects

Tessella

PREMIS Basics - Events

- Identifier
- Type
- Date Time
- Detail
- Outcome Information
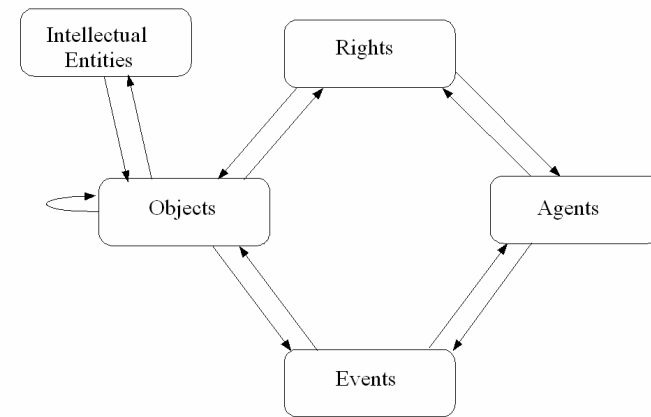- Agent Identifier
- Object Identifier

PREMIS Basics - Rights

- Identifier
- Basis
- Copyright Information
- Licence Information
- Statute Information
- Rights granted
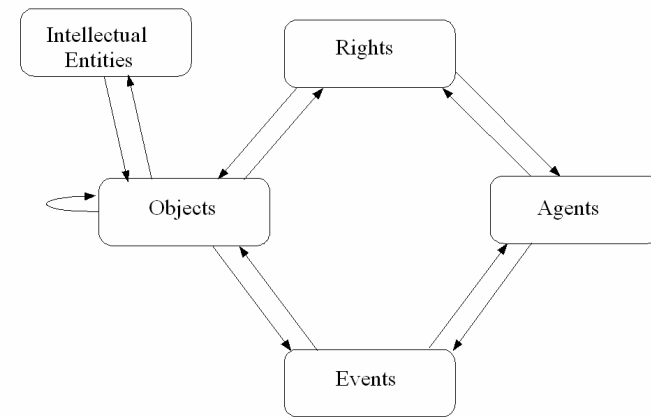- Linking object
- Linking agent

Tessella

PREMIS Basics - Agents

- Identifier
- Name
- Type

PREMIS Basics - Summary

- Objects:
  - Representation
  - File
  - Bitstream
- Events
- Rights
- Agents

PREMIS Conformance

- PREMIS is an Information Model:
  - "The information a repository uses to support the digital preservation process"
- Designed to be independent of:
  - Archiving system
  - Preservation strategy
  - Metadata management process
- There is a PREMIS XML Schema but do NOT have to use this to be CONFORMANT.

Tessella

PREMIS Conformance

- To be conformant:
    - Implement the semantic units in the Data Dictionary
    - If share name, must share definition
    - If not share name, must map definition (if mandatory)
    - Can add constraints but not remove them
- Internal Conformance:
    - Conformant within repository
- External Conformance:
    - Produce conformant export from the repository

Tessella

PREMIS Practicalities: e.g., SDB

- Pre-dates PREMIS (Representation = Manifestation)
- Need to respond quickly: Add extra fields to entities
- Need more entities:
  - Intellectual Entities broken down more than via cataloguing
  - Complex relationship with Representations
  - Use for automated migration and validation
- Don't want to hold lots of repeated information:
  - Use PRONOM PUIDs so Registry implied for every Format
- Hold Storage separate from immutable metadata
- Do have option to export to PREMIS XML schema:
  - Make implicit information explicit

Tessella

Where next? - Governance

- "Self-governed" by community
- PREMIS Editorial Committee
- If you get involved..
  - Likely to get invited on!

- Does react
- But it is everyone's part-time job!

- PREMIS 3.0:
  - Adds Intellectual Entity
  - Adds Environments
  - Allows for less verbosity
  - …

Tessella

CONCLUSIONS

- PREMIS:
  - Information Model for digital preservation
  - Allows for implementation variations
  - Allows for extensions (low conformance barrier)
  - Reacts to community
  - JOIN IN!

Tessella