Preservation Metadata:
Adapting or Adopting PREMIS for APSR

Bronwyn Lee, Gerard Clifton, Somaya Langley
Australian Partnership for Sustainable Repositories
and
National Library of Australia

iPRES 2006, Cornell University, October 2006

Outline

- ## APSR

- ## PREMIS Requirements Statement project (PRESTA)

- ## Project products:

  - ► List of preservation metadata elements

  - ► File formats

  - ► Tools for automated metadata collection

  - ► Gap reports

  - ► Preservation Event use cases

  - ► Profile for exchanging metadata

APSR

- Centre of excellence for the management of scholarly assets in digital format
- Focus on access continuity and sustainability
- Demonstrator repositories
- Develop skills and expertise
- National outreach and international linkages

Outline

✓ APSR

PREMIS Requirements Statement project (PRESTA)

Project products:

List of preservation metadata elements

File formats

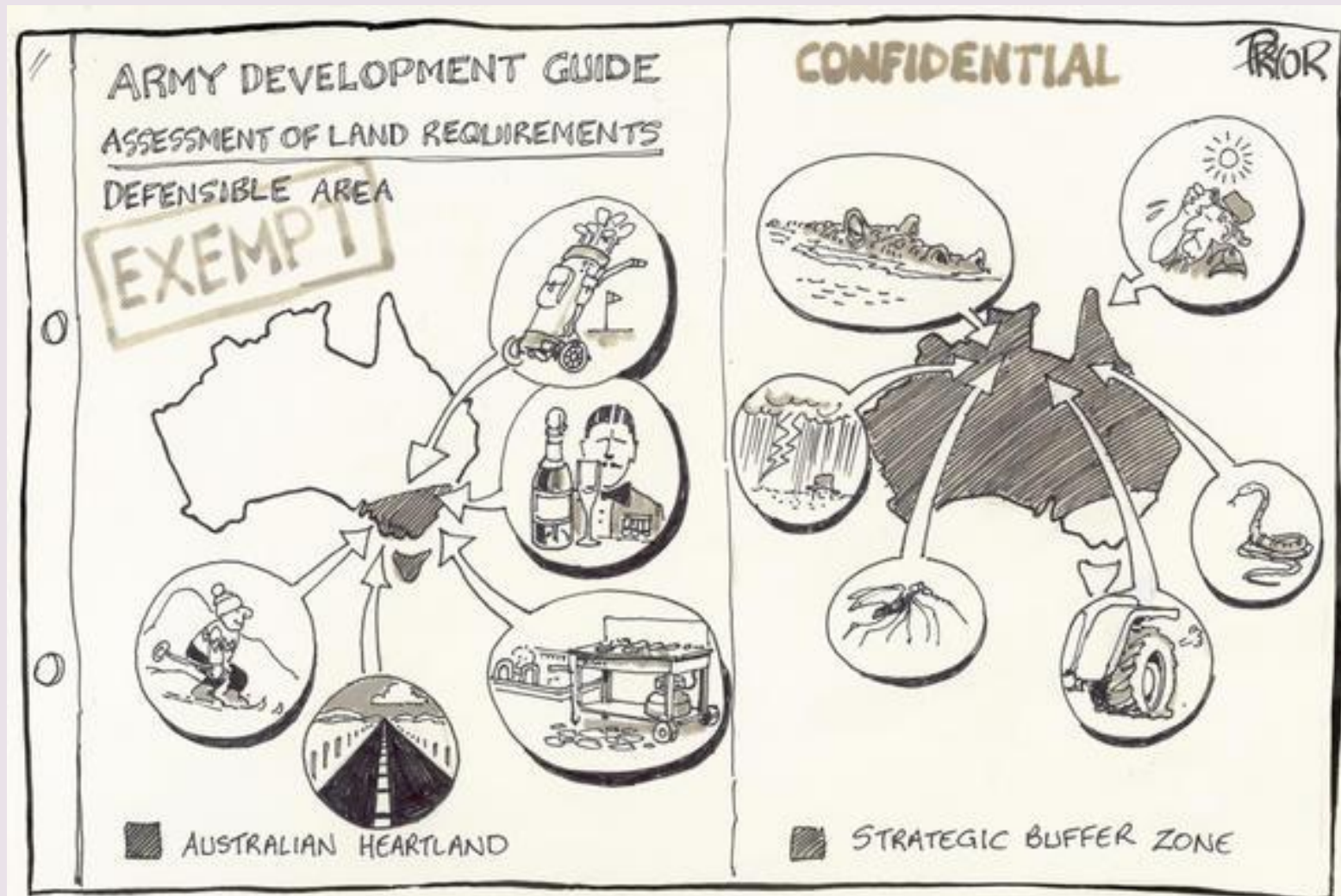Tools for automated metadata collection

Gap reports

Preservation Event use cases

Profile for exchanging metadata

Pryor, Geoff, 1944-
Army development guide, assessment of land **requirements** [picture] 1985

PREMIS Requirements Statement (PRESTA)

- Aim:
  - ▶ To specify requirements for the collection of metadata for preservation management based on PREMIS and help these be applied to selected partner repository implementations

PREMIS Requirements Statement (PRESTA)

- Draft work plan late 2005
- Analysis Dec 2005 - Jun 2006
  - ▶ National Library of Australia
- Report completed Jul 2006
- Selected repositories:
  - ▶ Australian National University (ANU) (DSpace)
  - ▶ University of Queensland (UQ) (Fez/Fedora)

Project approach

- ANU and UQ repositories already established with DSpace and Fedora models

- More emphasis on *what* metadata should be collected than on *how* it is collected

- Identify gaps and recommend enhancements

- Preservation event use cases and requirements *were* written as this was a significant gap.

- Preservation monitoring being addressed by AONS project

Project approach

- "Implements" PREMIS in two ways:
  - ▶ uses the Data Dictionary as a checklist against which repositories can compare their preservation metadata
  - ▶ uses the PREMIS schemas in a METS profile for exchanging preservation metadata (concrete framework for implementing PREMIS)

Outline

✓ APSR

✓ PRESTA - PREMIS Requirements Statement project

Project products:

List of preservation metadata elements

File formats

Tools for automated metadata collection

Gap reports
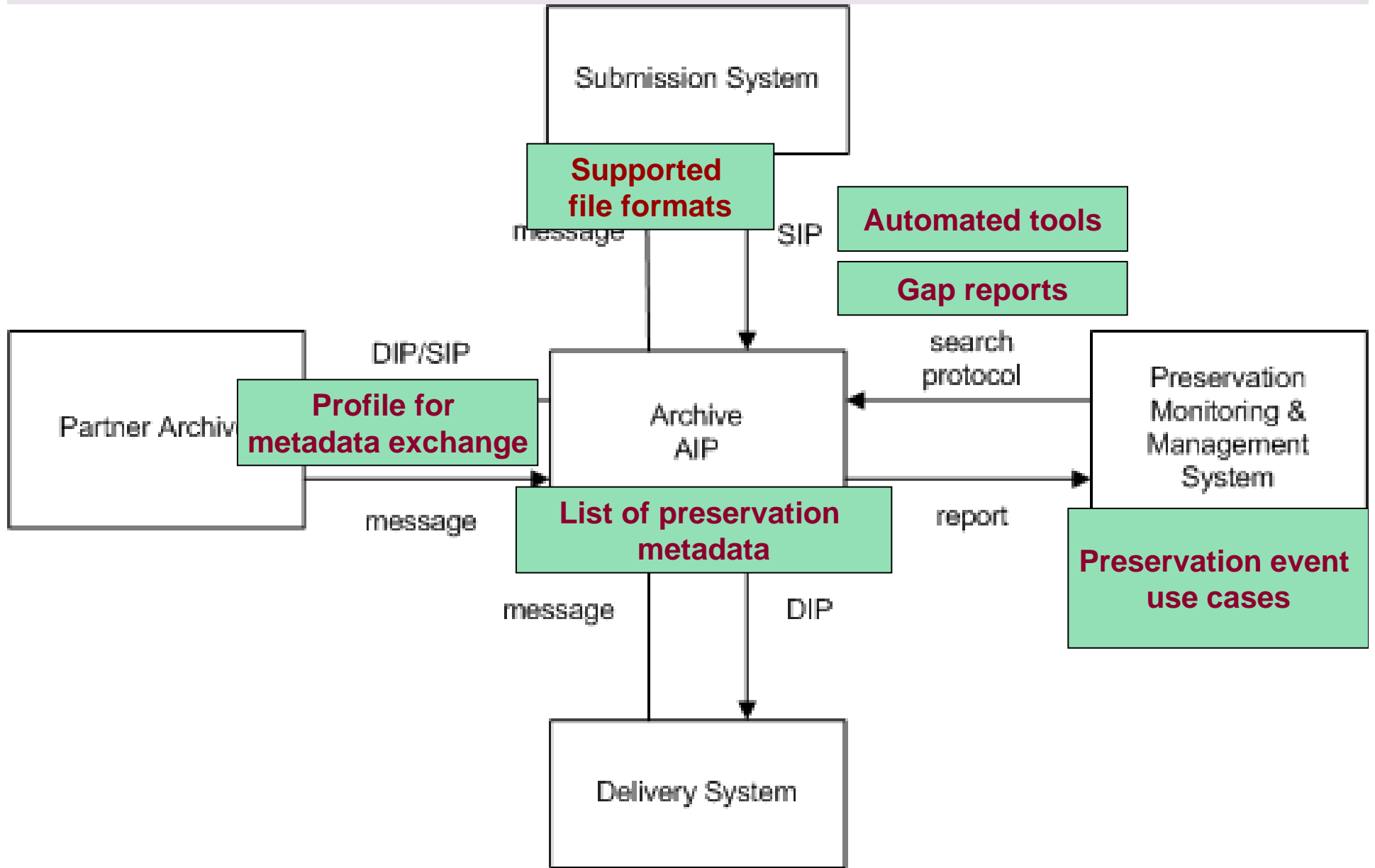
Preservation Event use cases

Profile for exchanging metadata

Sawyer, Harry G. (Henry George), 1903-
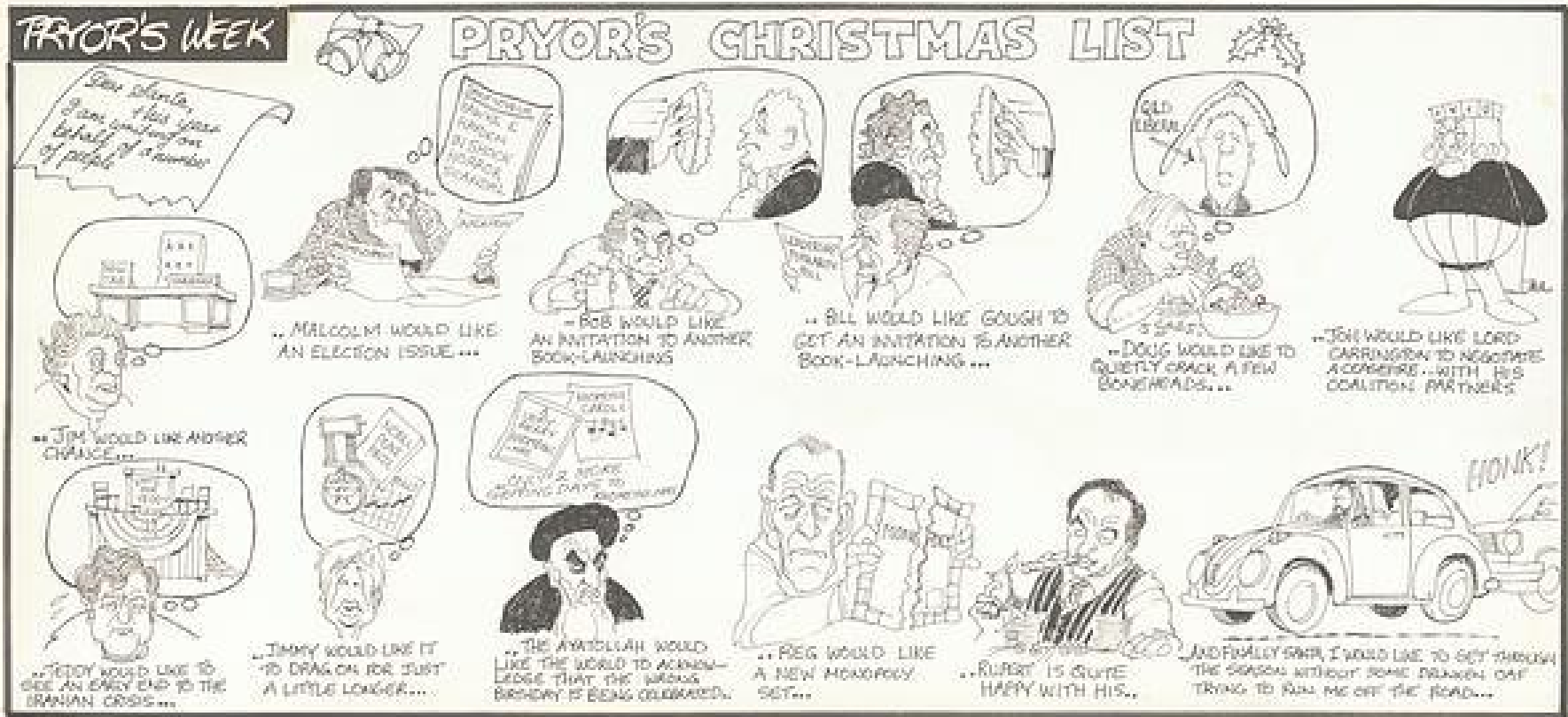My Watkins **Products** van, [Boronia, 1940s?] [picture]

Project products

1. List of preservation metadata elements
2. Supported file formats
3. Tools for automated metadata collection
4. Gap reports for ANU and UQ (May 2006)
5. Preservation event use cases and requirements (history logging)
6. Draft METS profile for exchanging metadata

Submission System

Supported file formats

message

SIP

Automated tools

Gap reports

DIP/SIP

search protocol

Profile for metadata exchange

Partner Archive

Archive AIP

Preservation Monitoring & Management System

message

List of preservation metadata

report

message

DIP

Preservation event use cases

Delivery System

Outline

✓ APSR

✓ PRESTA - PREMIS Requirements Statement project

✓ Project products:

List of preservation metadata elements

File formats

Tools for automated metadata collection

Gap reports

Preservation Event use cases

Profile for exchanging metadata

Pryor, Geoff, Pryor's Christmas **List** [Malcolm Fraser, Bob Hawke, Bill Hayden, Doug Anthony, Joh Bjelke-Petersen, Ted Kennedy, Jimmy Carter, Ayatollah Khomeini, Rupert Murdoch] [picture] 1979

List of preservation metadata elements

- Scenarios - "just in case" metadata?
- All the metadata required for long term sustainability and meaningful access
    - ▶ PREMIS Object, Event, Agent
    - ▶ Descriptive metadata (Intellectual Entity)
    - ▶ Structural metadata
    - ▶ Format specific metadata e.g. image, audio
    - ▶ Access rights metadata
- Collect as much as possible automatically
- Accept that some metadata unobtainable

"Mandatory" PREMIS metadata

- Things a repository should know about *every* archival object

- Doesn't specify how metadata is stored

- If not stored explicitly for each object it should be documented explicitly somewhere e.g. in policy or procedures

"Mandatory" PREMIS metadata

- Object
  - ▶ to metadata already mandatory in PREMIS e.g. objectIdentifier, preservationLevel we added some others e.g. size, fixity

- Event
  - ▶ ingest
  - ▶ when an archival object is changed

"Mandatory" PREMIS metadata

- Agent
  - ▶ If an Event changes an object, record an agent e.g. the software used
- Rights
  - ▶ Agreement with depositor - formal agreement or documented standard conditions or policies

Outline

- ✓ APSR
- ✓ PRESTA - PREMIS Requirements Statement project
- ✓ Project products:
  - ✓ List of preservation metadata elements

  File formats

  Tools for automated metadata collection

  Gap reports

  Preservation Event use cases

  Profile for exchanging metadata

National Library of Australia   nla.pic-vn3411612-v

Sievers, Wolfgang, 1913-

Tape **file** cabinet [at] Remington Rand, South Melbourne, [Victoria], 1968 [picture]
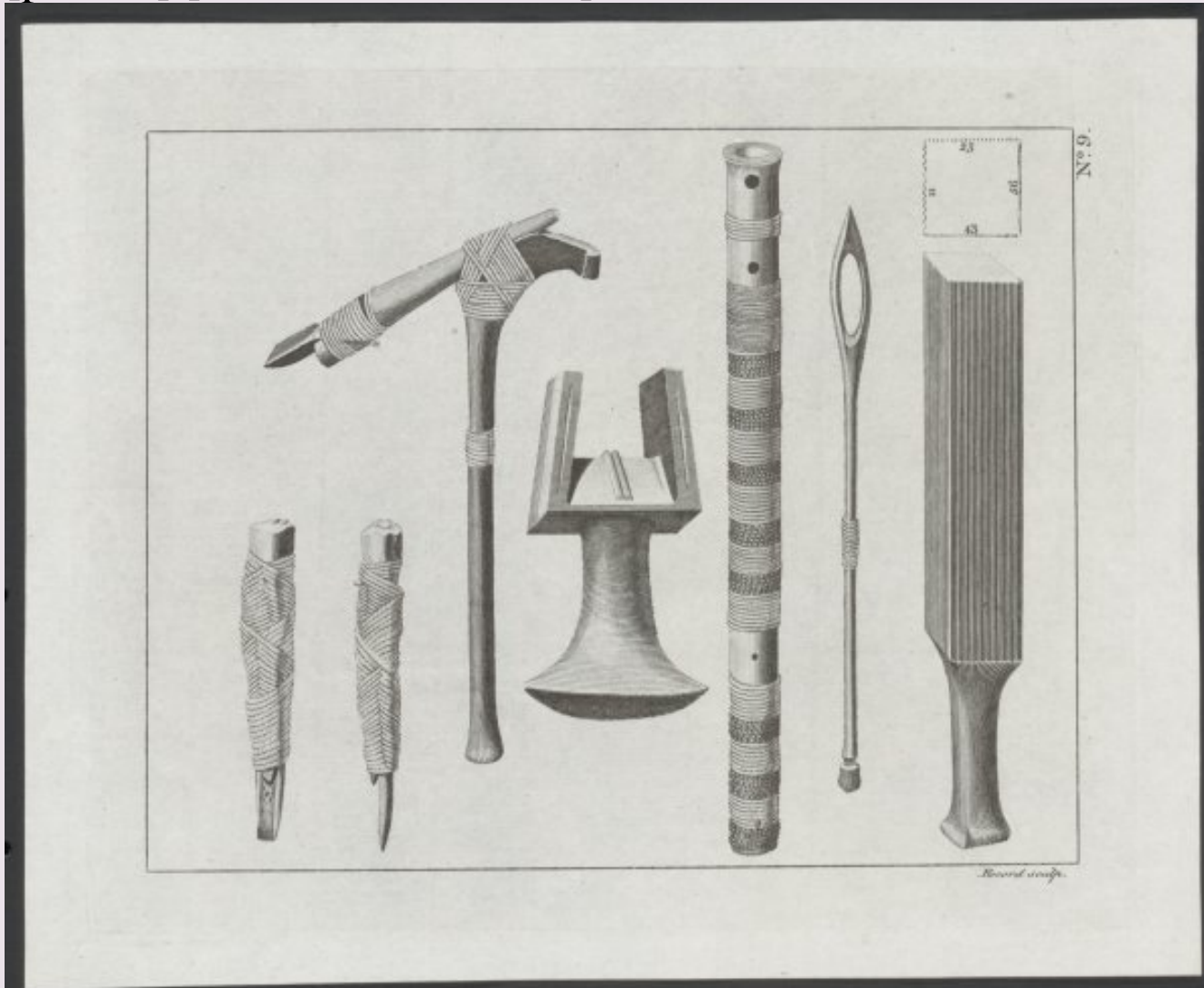
File formats

- Comments on images, audio, video, text
- Brief comments on databases, PDF, websites, multimedia, others
- Recommended archival formats
- Formats in common usage likely to be accepted e.g. digital camera formats
- Unsupported formats (should be converted to archival format)

Outline

- ✓ APSR
- ✓ PRESTA - PREMIS Requirements Statement project
- ✓ Project products:
  - ✓ List of preservation metadata elements
  - ✓ File formats

Tools for automated metadata collection

Gap reports

Preservation Event use cases

Profile for exchanging metadata

Record, John, fl. 1768-1780. [**Tools** and instruments from the Society Islands]
[picture] [London : s.n., 1773]

National Library of Australia   nla.pic-an9184910-v

Automated tools

- Tools for identifying file formats and automatically extracting metadata
- JHOVE, DROID, NLNZ Preservation Metadata tool
- Evaluation of capabilities
- Examples of output from the tools are linked to the report

Outline

✓ APSR

✓ PRESTA - PREMIS Requirements Statement project

✓ Project products:

- ✓ List of preservation metadata elements
- ✓ File formats
- ✓ Tools for automated metadata collection

Gap reports

Preservation Event use cases

Profile for exchanging metadata

Howard, Bruce, 1936-, Bridging the **gap** : the Gisborne By-Pass Road under construction 40km north of Melbourne, 1989 [picture], [Melbourne] : Herald and Weekly Times, 1989.

National Library of Australia   nla.pic-vn3068149-v

Gap reports for ANU, UQ

- Level of support (at May 2006) for core preservation metadata (i.e.PREMIS)
- Most significant gaps:
  - ▶ recording of preservation events
  - ▶ recording of structural relationships
  - ▶ file format validation (ANU)
  - ▶ checksum generation (UQ)

Outline

- ✓ APSR
- ✓ PRESTA - PREMIS Requirements Statement project
- ✓ Project products:
  - ✓ List of preservation metadata elements
  - ✓ File formats
  - ✓ Tools for automated metadata collection
  - ✓ Gap reports

  Preservation Event use cases

  Profile for exchanging metadata

Seselja, Loui, 1948- Scoreboard, equestrian individual three day **event** jumping, Horsley Park, Sydney 2000 Olympic Games, 22 September 2000 [picture]

Preservation Event use cases

1. Performing an action on an object which doesn't change the object e.g. error checking

2. Performing an action on an object which transforms the object into a new object e.g. migration to a newer format

3. Deleting an object

Preservation Event use cases

4. Updating the content of an object (not preservation but included for clarification)

5. Updating metadata about an object

Event use case example:
Performing an action on an object which
transforms an object into a new object

Base course: A new object is created and the old
object *is* kept.

1. Preservation Monitor or Workflow System
   alerts the Event Manager that an action to
   change an object needs to be performed.

2. The Event Manager schedules the event.

Event use case example

3. The Event Manager takes a copy of the object, and modifies it to create a new object.

4. The Event Manager submits the new object to the Repository along with details of the event which created it, including its relationship to the old object.

5. The Repository ingests the new object, records the relationship between the new and old objects, applies version information and assigns a unique identifier to the new object.

Event use case example

6. The Repository stores relevant preservation metadata about the new object.

7. The Repository ensures descriptive, rights and and any other relevant metadata from the old object are associated with the new object.

8. The Repository records details of the event which created the new object and associates the event with the new and old objects.

Event use case example

9. The Repository records the event which ingested the new object.

Alternative course: A new object is created and the old object is *not* kept.

Outline

- ✓ APSR
- ✓ PRESTA - PREMIS Requirements Statement project
- ✓ Project products:
  - ✓ List of preservation metadata elements
  - ✓ File formats
  - ✓ Tools for automated metadata collection
  - ✓ Gap reports
  - ✓ Preservation Event use cases
  - Profile for exchanging metadata

Hurley, Frank, 1885-1962. Eating gum tips, koala **profile** [picture] : [Taronga Park Zoo, Sydney, New South Wales] [between 1910 and 1962]

National Library of Australia   nla.pic-an23817478-v

Draft profile for exchanging metadata

- METS profile for transferring custody of an object (needs most metadata)

- Common profile for different systems to map to

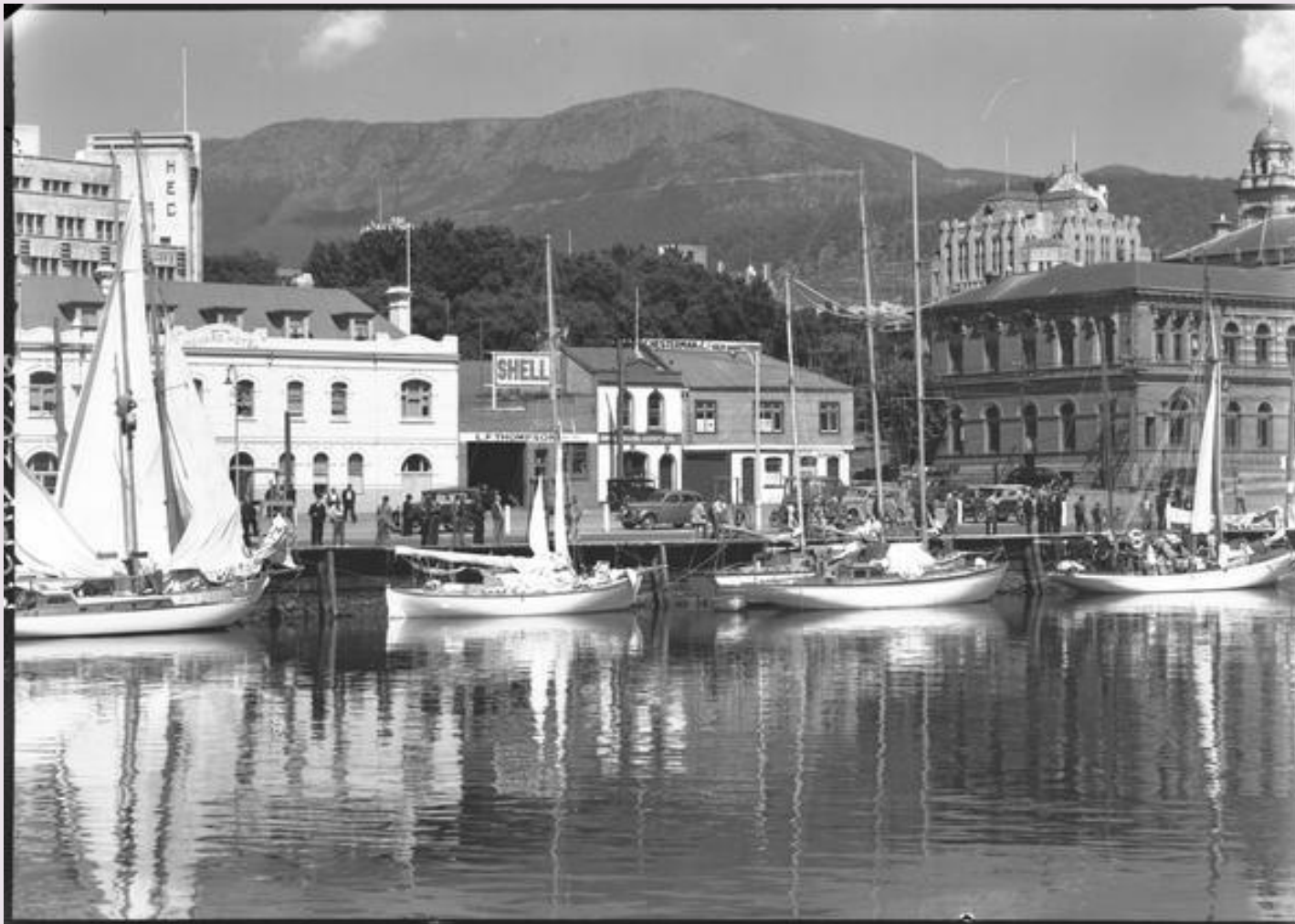- Uses MODS, PREMIS, MIX, other format specific schemas

Draft profile for exchanging metadata

- Using the draft profile ANU and UQ have built a demonstrator which transfers a digital object between DSpace and Fedora.

- Demonstrated that the receiving repository is able to understand, store and use the imported metadata.

- Draft profile in report is early thinking

- It is being revised in the light of ANU and UQ feedback and what others in PREMIS and METS groups are doing

Conclusion

- 14 recommendations
- Report published on APSR website - follow the Publications link
- http://www.apsr.edu.au/publications/PRESTA/
- Repositories and our thinking are continuing to evolve

Hurley, Frank, 1885-1962. [Sea-going yachts in Constitution Dock at the **conclusion** of the annual Sydney to Hobart yacht race] [picture] : [Hobart, Tasmania] [between 1910 and 1962]

National Library of Australia   nla.pic-an23504326-v