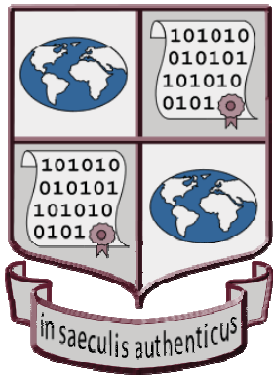
InterPARES 2 Project

International Research on Permanent Authentic Records in Electronic Systems



Description and Metadata

Presentation developed by Sue McKemmish and Joanne Evans, on behalf of the IP2 Description Cross Domain

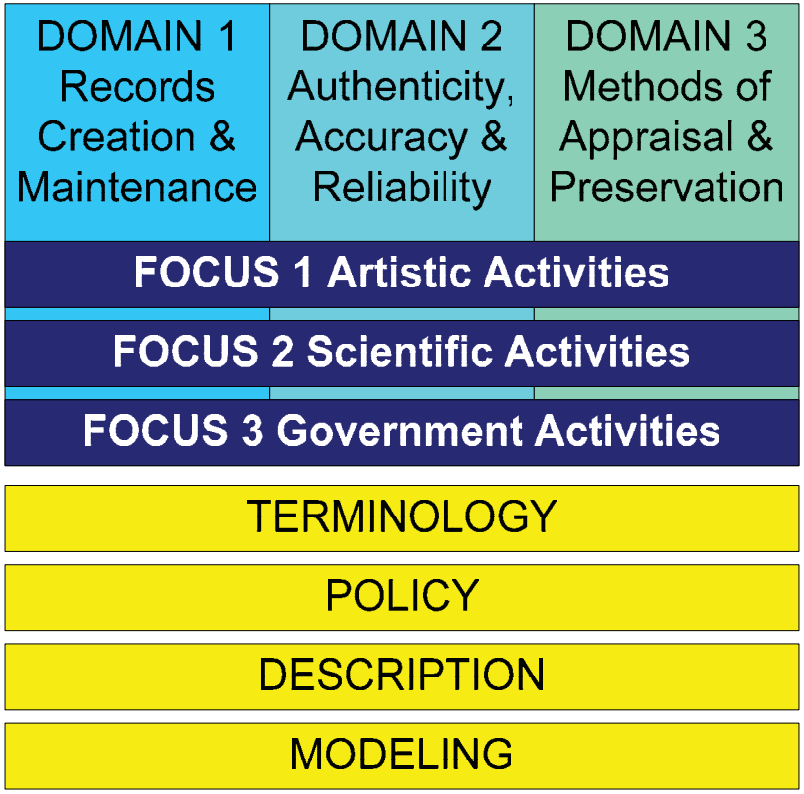
The Hague, 22 June 2005

Outline

- InterPARES 2
- Description Cross Domain
- Research Activities
  - Literary Warrant Database
  - Specification Models
  - Metatools
- MADRAS, Metadata and Archival Description Registry and Analysis System
  - Descriptive elements
  - Analysis framework
- Research Outcomes

InterPARES 2

- Collaborative international research initiative investigating issues of authenticity, reliability and accuracy of records created in digital environments resulting from artistic, scientific and government activities.
  http://www.interpares.org/ip2/ip2_index.cfm



Description Cross Domain People

Professor Sue McKemmish (co-chair) & Joanne Evans, Monash University, Australia

Professor Anne J. Gilliland PhD (co-chair), Nadav Rouche, Lori Lindberg, Holly Wang and Monique Leahey-Sugimoto, University of California, Los Angeles

Karuna Bhoday, National Archives of Australia/Monash

Professor Terry Eastwood and team of GRAs, University of British Columbia, Canada

Hans Hofman, National Archives of the Netherlands

Dr Richard Marciano, San Diego Supercomputer Center

Dr Bill Underwood, Georgia Tech, USA

Funding support for InterPARES is provided by the Social Sciences and Humanities Research Council of Canada, United States' National Historical Publications and Records Commission and the National Science Foundation

Support for Joanne Evans and Karuna Bhoday is provided through an Australian Resource Council (ARC) Linkage Grant, the Clever Recordkeeping Metadata Project, in conjunction with Monash University, National Archives of Australia, State Records Authority of New South Wales and the Australian Society of Archivist's Committee on Descriptive Standards

Premise

An essential component of electronic recordkeeping and archiving is an infrastructure to support the creation, preservation, and accessibility in and through time of trustworthy, understandable metadata, including archival description. Such an infrastructure includes metadata registries, recordkeeping metadata schemas, archival description standards, metadata brokers and metatools.

Definitions

- Recordkeeping
- Recordkeeping Metadata
- Archival Description
- Role of Metadata in Recordkeeping Systems and Archival Description
- Recordkeeping Metadata Schema
- Archival Description Standards

Description Cross Domain

- Investigating issues relating to the creation, capture, management and preservation of metadata in recordkeeping systems and archival description, and its role in assuring authentic, accurate and reliable records and archives

- Major research activities:
  - MADRAS, Metadata and Archival Description Registry and Analysis System
  - Metadata specification models
  - Metadata tools
  - Evaluation and analysis frameworks and instruments
  - Literary Warrant Database
  - Collaboration with Monash CRKM, ISO and SDSC

Objectives

Evaluate whether existing and emerging metadata schemas, descriptive standards and metadata tools meet recordkeeping metadata and archival description requirements relating to the appraisal, creation, management, preservation and use of reliable and authentic records in e-government, artistic activities and scientific activities

Recommend the development of new metadata schemas, descriptive standards and metadata tools, or the extension of existing/emerging ones

Develop metadata specification models and specifications for metadata tools to support this

Literary Warrant Database

Identifies warrant, or authoritative sources relating to metadata requirements for the accuracy, reliability, and authenticity of records and the ways in which metadata and archival description can support appraisal, creation, management, preservation and use

Metadata Specification Models

Recommendations and specification models for

- extension of existing generic and sector specific schemas and standards; or
- development of new ones

to meet recordkeeping metadata and archival description requirements, particularly in e-government, artistic and scientific communities

Metatool Evaluation and Specification

- Evaluation of usefulness of existing tools
- Specifications for metadata management tools for activities such as automatic metadata specification and extraction

MADRAS, Metadata and Archival Description Registry and Analysis System

Describes relevant existing and emerging recordkeeping metadata schema and archival descriptive standards in a standardised way:

- scope and purpose
- type of metadata/descriptive elements
- related encoding schemes
- applicability of the schemas and standards to recordkeeping and archival functions

Analyses and evaluates schemas and standards in relation to recordkeeping and archiving requirements as set down in standards

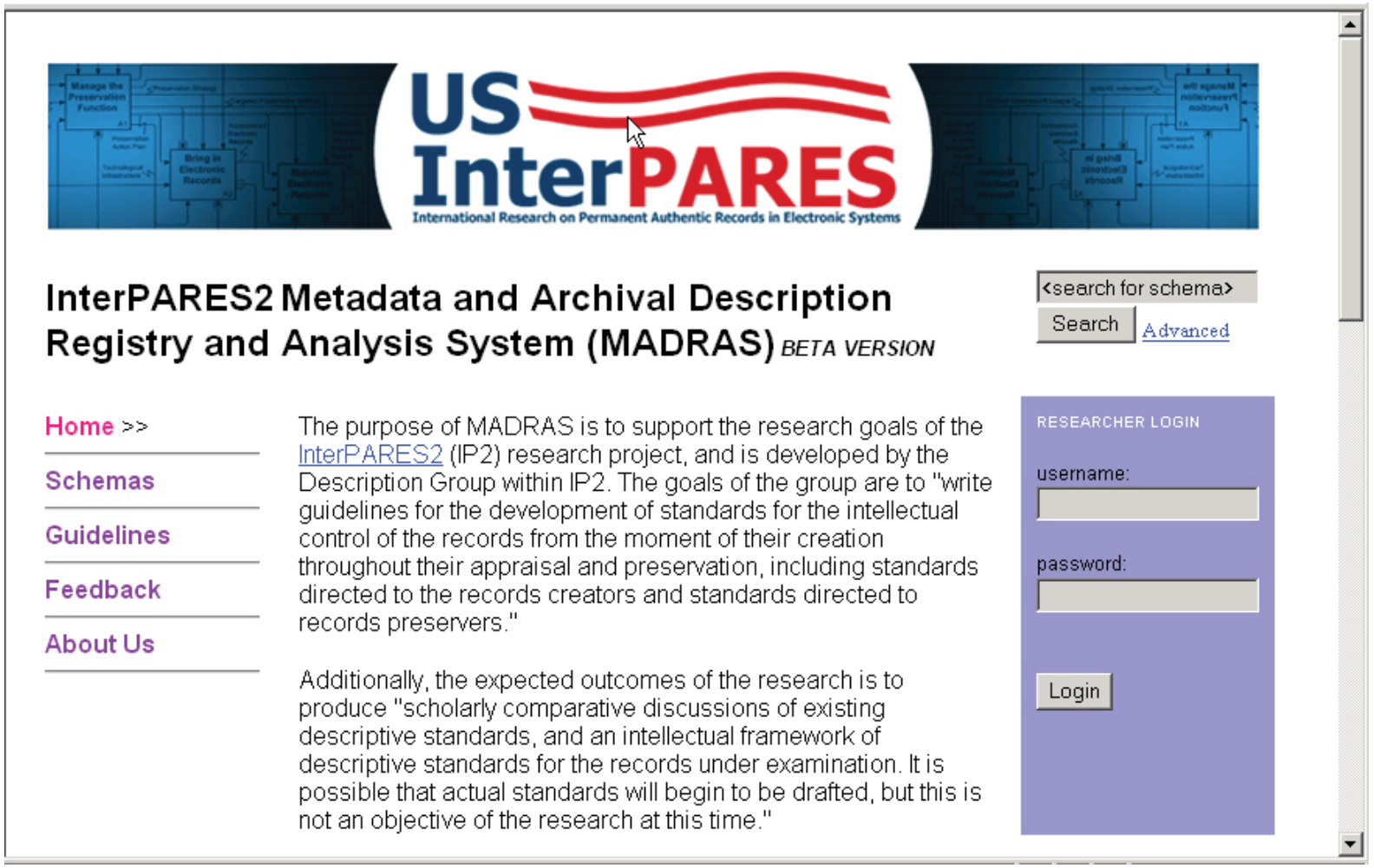
Prototype Version

- Prototype consists of:-
  - Microsoft Access database for capturing and controlling descriptive metadata about metadata standards and related entities in this initial phase – allows testing of elements, their structures and sources of values.
  - Draft XML DTD – instantiated through programmed output from the Access database
  - Analysis process - guidelines, worksheet, summary document
  - Web pages

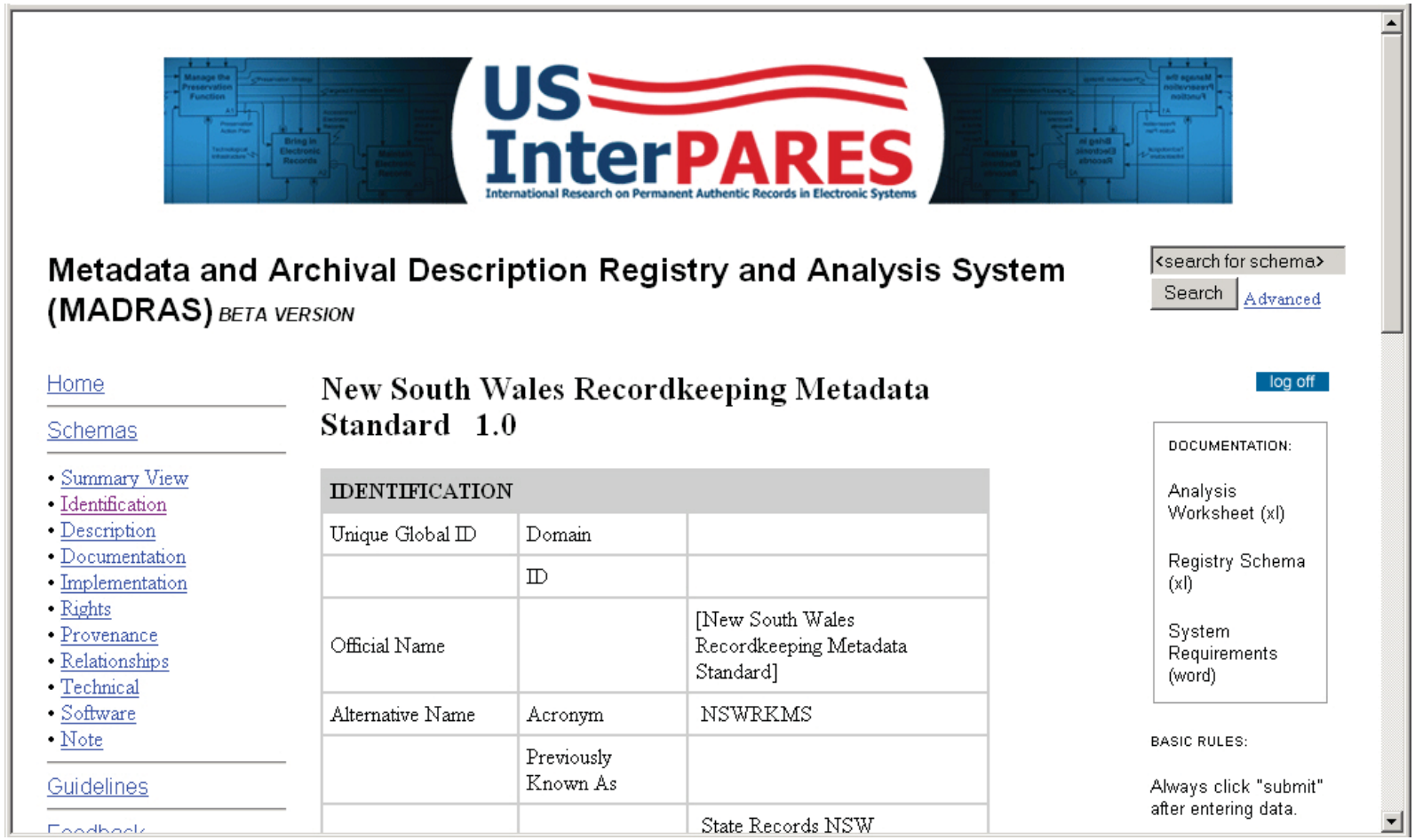
Production Version

- Production version consists of:-
  - Environment: MySQL database , PHP script, and Apache web server running on Windows.
  - The website comprises a public area where completed schemas are published, and a private area where researchers create and manage schema records.
  - The system allows for online distributed collaboration, authentication of users and basic workflow management.
  - Beta version will go live in mid-June, and includes all registry schema elements except the Analysis element.
  - Next steps:
    - Collect feedback and bug reports from users
    - Integrate the Analysis element into the system
    - Prioritize roll out of additional functionalities

Production Web Interface – Homepage



Production Web Interface – NSWRMS record



Production Web Interface – EAD close-up (partial)

Encoded Archival Description 2002

IDENTIFICATION		
Unique Global ID	Domain	
	ID	
Official Name		[Encoded Archival Description]
Alternative Name	Acronym	EAD
	Previously Known As	
	Other	
Version		2002
Status		
Publication	Publisher	Society of American Archivists
	Place	USA
	Date	19 March 2003
Duration	Start Date	19 March 2003
	End Date	

MADRAS: The Registry Metadata Schema

Breadth and Depth of Schema

- 120 fields organized hierarchically in an element structure
- Relatively flat structure going 3 levels deep
- The first level of the hierarchy comprises eleven elements: Registration, Identification, Accessibility, Rights, Provenance, Description, Analysis, Documentation, Relationships, Administration, and a general Note element

Registry Schema Elements 1-5

1. **Registration**
   Data elements to register metadata schema into the registry, i.e. registration number, date and action officer.
2. **Identification**
   Data elements to identify and distinguish metadata schemas, i.e. title, unique global identifier, version, publication statements etc.
3. **Accessibility**
   Data elements to capture information relating to the accessibility of a schema, i.e. hardware and software requirements, etc.
4. **Rights**
   Data elements to capture intellectual property rights associated with the use of a metadata schema.
5. **Provenance**
   Data elements to capture organisations or other bodies/agents associated with the development, publication and maintenance of a metadata schema.

Schema Elements 6-11

6. **Description**
   Data elements to capture the purpose, scope, jurisdiction, of a metadata schema including the types of entities and objects it has been designed to be used for, etc.
7. **Analysis**
   Data elements for analysing a metadata schema or data elements for capturing the results of analysis of a metadata schema against archival and recordkeeping requirements.
8. **Documentation**
   Data elements for capturing citations to the documentation of a metadata schema, e.g. specification, guidelines, etc.
9. **Relationships**
   Data elements to capture relationships amongst metadata schema and to other classification schemes.
10. **Administration**
    Data elements for the administration of the schema registry.
11. **Note**
    General note element for the whole record.

MADRAS: Analysis Instrument

- Assess existing metadata sets and archival description standards to identify what elements/descriptors are captured and what are not
- Assess against 'best practice' standards:
  - ISO 23081 Records management processes – Metadata for records
  - InterPARES Benchmark and Baseline requirements
  - Monash SPIRT RKMS
  - Classification of Recordkeeping Metadata Purposes (Kate Cumming)

Analysis – 1. General

1.1 What types of (conceptual) entities does the metadata schema/archival description standard describe?

1.2 What types of (physical) objects are described (if specified in the schema/standard documentation)?

1.3 Does schema/standard encapsulate the object which it is describing?

1.4 What is the intended method of metadata/descriptive element creation?

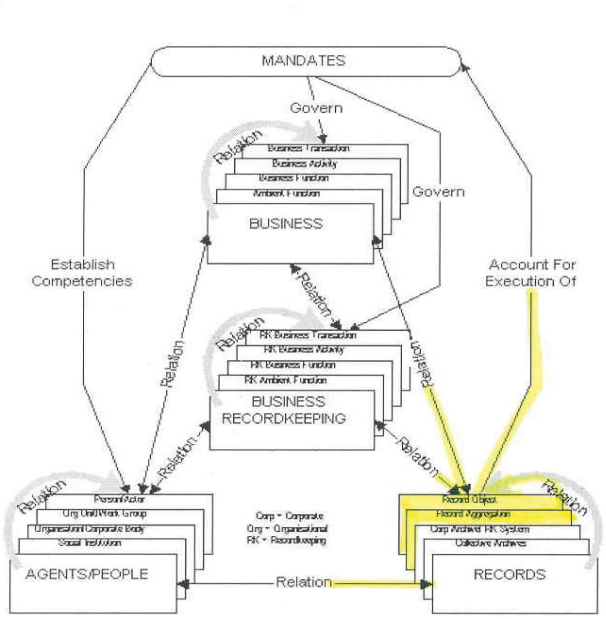
1.5 What is the expected nature of the metadata/descriptive elements?

1.6 What categories of metadata/descriptive elements does the schema cover?

1.7 What is the conceptual data model on which the schema/standard is based?

Analysis – 2. Recordkeeping

2.1 a) What recordkeeping entities could/does the schema/standard describe?



Record Object
- Record Object
- Record Aggregation
- Recordkeeping System or Corporate Archive
- Collective Archives

Business rules, policies or mandates

Agent
- Person/actor
- Org Unit/Work Group
- Organisation/Corporate Body
- Social Institution

Business Activities or Processes
- Business Transaction
- Business Activity
- Business Function
- Ambient Function

Records management/Recordkeeping business processes
- RK Business Transaction
- RK Business Activity
- RK Business Function
- RK Ambient Function

Metadata about the metadata record itself

Analysis – 2. Recordkeeping

2.1 b) What recordkeeping relationships does the schema/standard describe?

2.2 Aggregation - How are levels of aggregation dealt with?

2.3 What is the recordkeeping scope?

Relationships of Record entities to:-
· Record Entities
· Agent Entities
· Business Entities
· RK Business Entities
· Mandates
Relationships of Agent entities to:-
· Record Entities
· Agent Entities
· Business Entities
· RK Business Entities
· Mandates
Relationships of Business entities to:-
· Record Entities
· Agent Entities
· Business Entities
· RK Business Entities
· Mandates
Relationships of RK business to:-
· Record Entities
· Agent Entities
· Business Entities
· RK Business Entities
· Mandates
Relationships of mandates:-
· Record Entities
· Agent Entities
· Business Entities
· RK Business Entities

Analysis – 3. ISO 23081

3.1 Metadata/descriptive elements about records - how and to what extent does the schema/standard meet the requirements of Section 10.2?

10.2.1 Structural and storage metadata at point of capture
a) Date and time of record creation
b) Agents involved in record creation
c) Record structure
d) Record form
e) Chemical and physical properties
f) Technical characteristics and dependencies
g) Relationships between data or format elements comprising the record
h) Requirements for making available, reproducing or rendering the record
I) Relationships between record and business transaction or activity that generated it
m) Links between records, or between an individual record and the broader record aggregate of which it is a part
10.2.2 Process metadata after capture
Changes to metadata regarding logical and physical structure and technical attributes of the record
Changes to relationships with other records and aggregates
10.2.3.1 Accessibility Metadata at point of capture
a) Identity of records or aggregations
b) Identify and document the aggregation
c) Record location information - logical and physical, multiples (e.g. home and current)
d) Identify and document links between records agents and processes
e) Descriptive information to facilitate use and understanding - subject classification, title, keyword indexing, etc.
10.2.3.2 Accessibility process metadata after capture
Changes in accessibility metadata
Additional descriptive metadata to explain explicitly the business context of the records
10.2.4.1 Metadata supporting the security of records at point of capture
a) Identify access rights and restrictions
b) Document time limitations for access rights and restrictions
10.2.4.2 Metadata supporting the security of records after record capture
Changes to access rights/restrictions arising from administrative or personnel changes

Worksheet

3.1 Metadata about records

	Structural/Storage	Accessibility	Security
Point of capture:	Adequate	Comprehensive	Adequate
Process:	Adequate	Comprehensive	Adequate

The metadata schema meets most of the requirements for metadata about records in section 10.2. It allows for metadata at point of creation and the accumulation of process metadata with the Event History element providing an audit trail of records management and other processes carried out on the record. The Relation element allows for the capture of relationships to agent and function elements through time.

Summary document e.g. NSW RKMS

Analysis – 3. ISO 23081

3.2 Metadata/descriptive elements about business rules, policies and mandates - How and to what extent does the standard/schema meet the requirements of Section 10.3?

10.3.1 Metadata about business rules/policies and mandates at point of capture
a) Identify the specific metadata schema/scheme used in organisational business systems
b) Business rules or other system controls that regulate records creation and management
c) Business rules or other system controls that regulate metadata creation and management
d) Business rules or other system controls that regulate records management operations
e) Business rules or other system controls that regulate access to, and rights of records
f) Mandate or other regulatory requirement for record creation and management
g) Mandate or other regulatory requirement for record retention, security or destruction
h) Relationships between mandate information and the records and/or records management processes
10.3.2 Metadata about business rules/policies after point of capture
Changes to business rules, mandates or regulatory requirements
Changes to relationships between mandates, records and record management processes

Worksheet

3.2 Metadata about business rules, policies and mandates

Point of capture:	Adequate
Process:	Adequate

All entities have a mandate element. Record entity has Business Rules component to Preservation, Retrieval, Disposal, Control, Access, Use and Event History elements. Function entity has Business Rules element.

Summary document e.g. NSW RKMS

Analysis – 3. ISO 23081

3.3 Metadata/descriptive elements about agents - how and to what extent does the schema/standard meet the requirements of ISO 23081 Section 10.4?

10.4.1 Agent Metadata at point of capture
a) Agents (persons, workgroups, organisations) involved in records creation (identity, position, relationship to other agents)
b) Agents involved in records management processes or authorisation, e.g. identities and position of agents making decisions relating to records disposition
c) Identify agents authorised to access records (security metadata)
10.4.2 Agent Metadata after point of capture
Agents associated with use and management of records over time
Changes in roles and authorisations for accessing records and performing records management operations

Worksheet

3.3 Agent Metadata

Point of capture:	Comprehensive (except for security)
Process:	Comprehensive (except for security)

Agents associated with record and function entities can be identified and described and relationships between agents, functions and records captured through time. However the ability to capture security metadata relating to agents is lacking.

Summary document e.g. NSW RKMS

Analysis – 3. ISO 23081

3.4 Business process metadata/descriptive elements -

If schema/standard describes business entities then how and to what extent does it meet the requirements of Section 10.5?

10.5.1 Business Process Metadata at point of capture
a) Business functions, activities and transactions documented by the records
b) Relationships between records, agents and the business functions, activities and transactions
c) Agents or participants in transactions
d) Security and access rules for business processes and transactions
f) Classification of business functions, activities and transactions
h) Date and time of transaction
10.5.2 Business Process Metadata after point of capture
Business functions, activities and transactions associated with the use and management of records over time

Worksheet

3.4 Business process metadata

Point of capture:	Comprehensive (except for security)
Process:	Comprehensive (except for security)

The function entity can be used to identify and describe business processes associated with agent and record entities and relationships between agents, functions and records captured through time. However the ability to capture security metadata for function entities is lacking.

Summary document e.g. NSW RKMS

Analysis – 3. ISO 23081

3.5 Metadata/descriptive elements about records and archives management processes - how and to what extent does the standard/schema meet the requirements of Section 10.6?

10.6.1 Metadata about Records Management Processes at point of capture
b) Disposition metadata applied to records
c) Identify and document agent authorisations or permissions for records management activities
d) Time limitations for user authorisations or permissions
e) Document the security and access metadata applied to records
f) Classification of business functions
h) links between records and their aggregations and between records
10.6.1 Metadata about Records Management Processes after point of capture
Ongoing records management processes

Worksheet

3.5 Metadata about records management processes

Point of capture:	Minimal
Process:	Minimal

The schema adopts a record centric view of documentation for records management processes with metadata captured as elements of record entities. In particular this approach does not allow for agent authorisations or permissions to be captured within the schema.

Summary document e.g. NSW RKMS

Analysis – 3. ISO 23081

3.6 Metadata about the metadata/descriptive record - How and to what extent does the schema/standard allow for metadata about the metadata/descriptive record to be captured? (Scheme derived from Section 9)

Metadata about creating/altering metadata
Date/time of metadata creation/alteration
Agent (9.3.9.2)
Activity (9.3.9.2)
Metadata about security and access restrictions for metadata (9.3.8)
Metadata about metadata
Rules and policies (9.3.1, 9.3.9.2, 9.3.5, 9.3.7)
Structures (9.3.3, 9.3.7, 9.4)

Worksheet

3.6 Metadata about the metadata record

Point of capture:	Minimal
Process:	Minimal

The metadata schema only has an element to capture metadata about the process by which metadata is attributed to record entities.

Summary document e.g. NSW RKMS

MADRAS - Outcomes

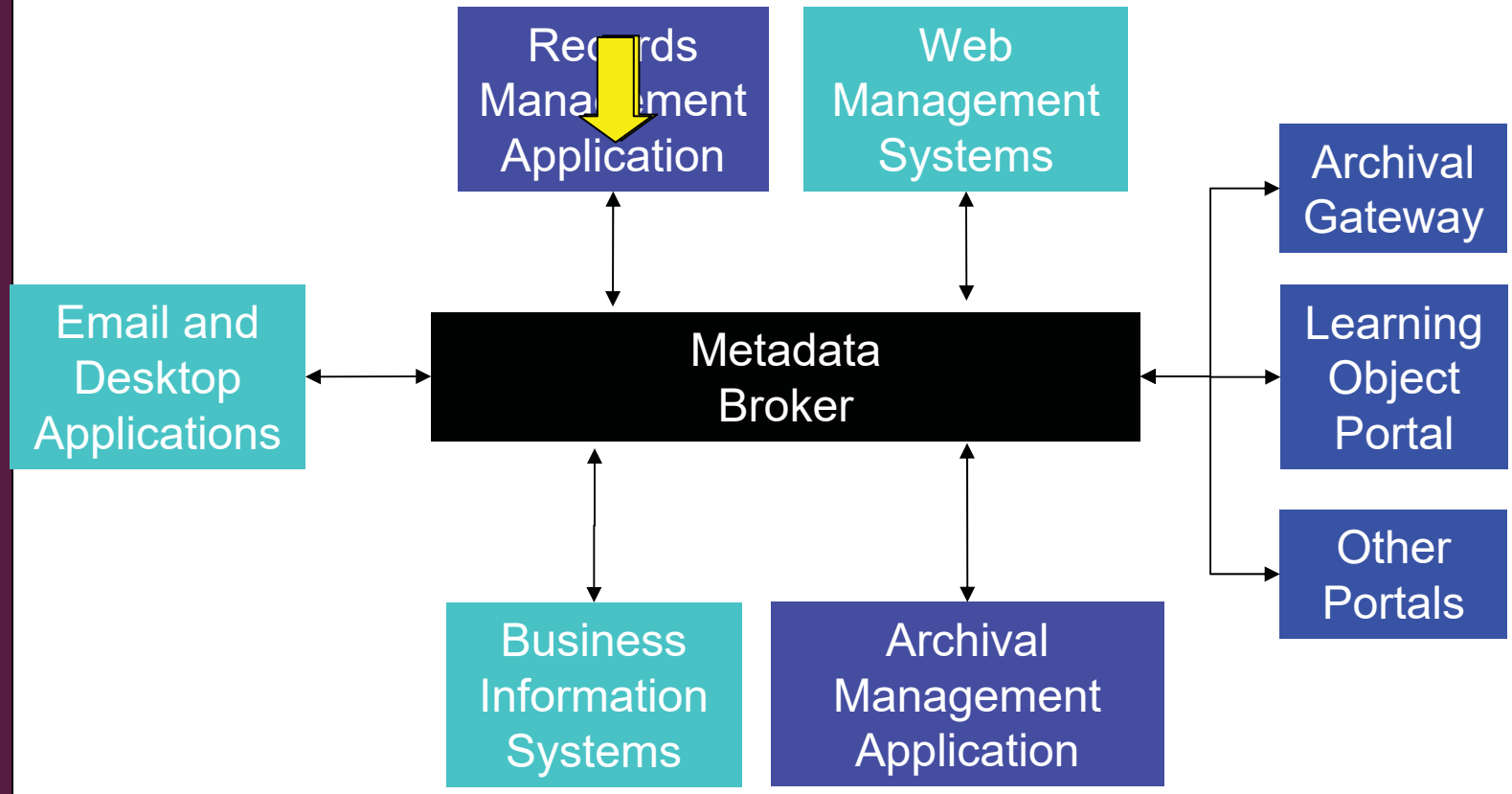
- Database which describes, analyses and evaluates schemas/standards with reference to ISO, InterPARES 1 Benchmark and Baseline Requirements, and Monash SPIRT RKMS

- Centralised repository of information to assist organisations and archival institutions in identification of relevant schemas/standards/combinations of schemas to meet their specific needs

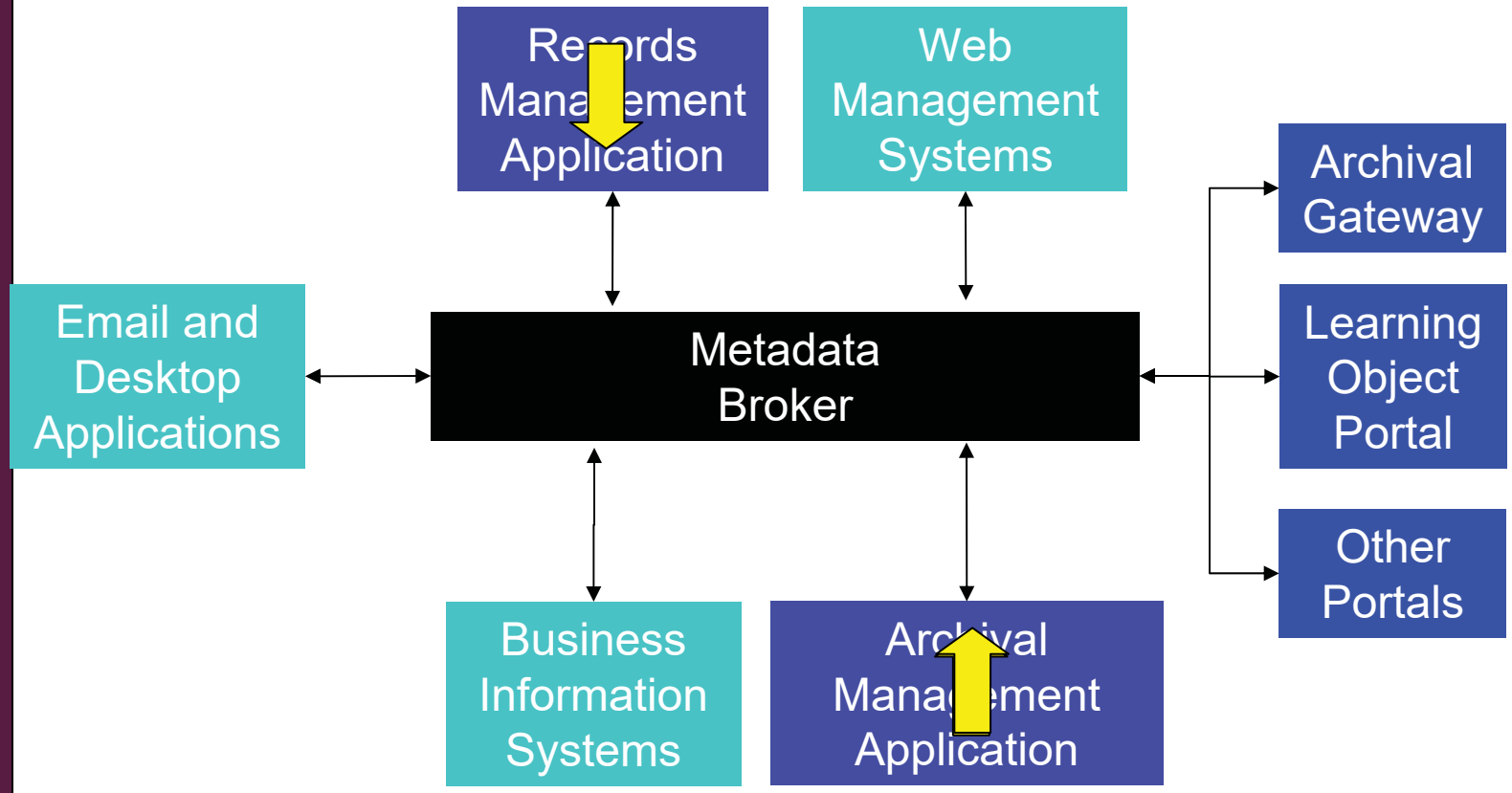
MADRAS - Outcomes

- Recommendations and model specifications for extension of existing generic and sector specific schemas and standards or development of new ones to meet recordkeeping metadata and archival description requirements, particularly in e-government, artistic and scientific communities
- Standardised framework for assessing functionality of schemas and standards, e.g. by standards-setting bodies (Note: link to ISO 23081 Part III)
- Findings on comprehensiveness and adequacy of standards and benchmarks
- Component in a metadata broker service that enables automatic or semi-automatic re-use and repurposing of metadata for business, recordkeeping and archiving purposes

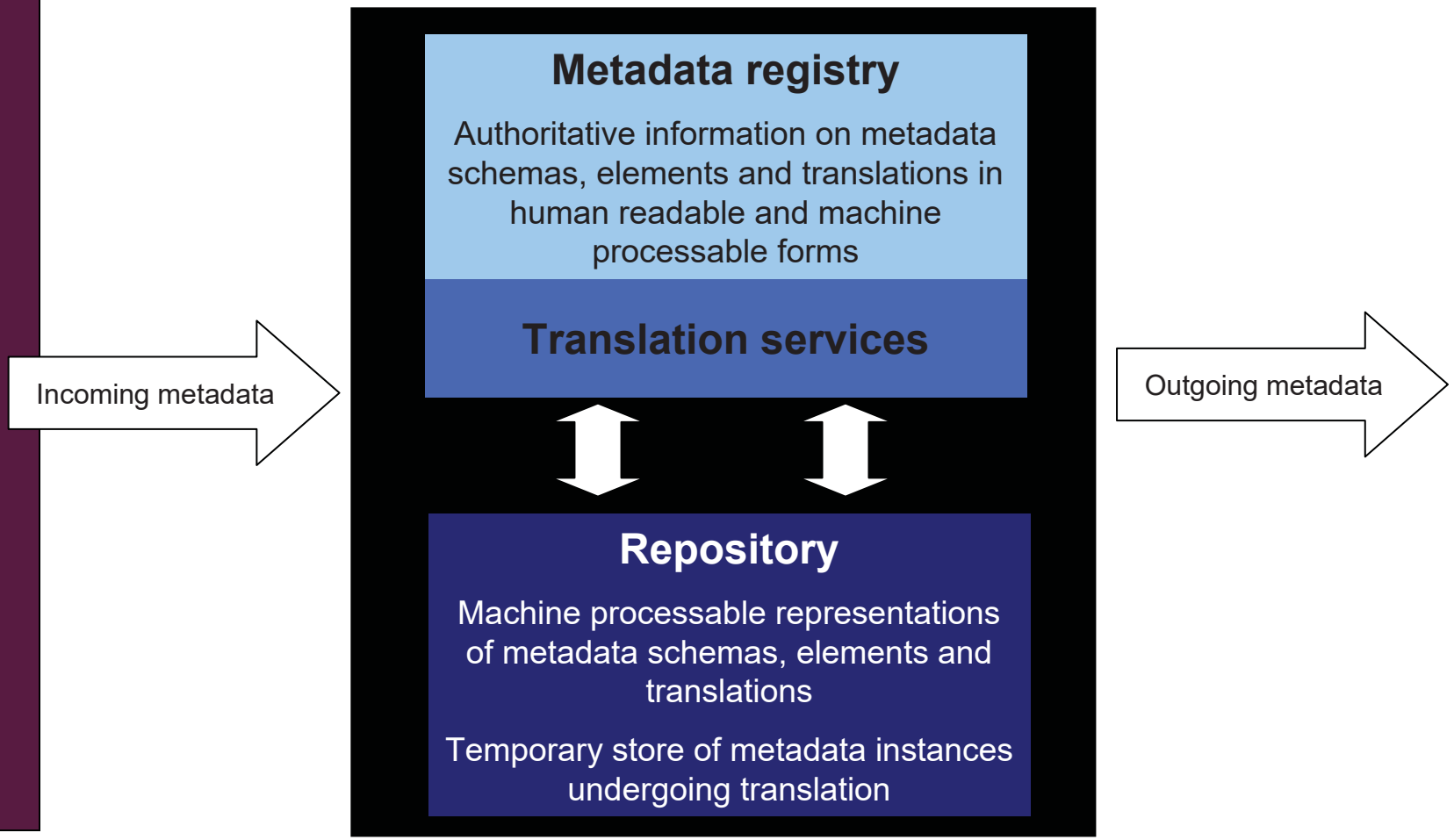
Clever Recordkeeping Metadata Project (CRKM): Re-use Scenarios



CRKM Metadata Re-use Scenarios



Clever Recordkeeping Metadata Broker



InterPARES Project Description Cross Domain
Sue McKemmish & Joanne Evans
School of Information Management and Systems, Monash University

MONASH University
Information Technology

Other Research Outputs

- Other policy and standards recommendations (Policy Cross Domain; ISO; Standards Australia)
- Literary Warrant Database
- Specifications for Metadata Tools
- Contribution of new knowledge, understandings, models and tools to wider metadata community
- Collaborations/sharing research outputs with other R&D initiatives, including ISO, Monash CRKM project, SDSC projects
- Publications and presentations

References - URLs

- InterPARES2 site: www.interpares.org
- US InterPARES site: http://www.gseis.ucla.edu/us-interpares/
- Clever Recordkeeping Metadata Project: http://www.sims.monash.edu.au/research/rcrg/research/crm/
- San Diego Supercomputer Center: http://kbis.sdsc.edu/index.html
- International Standards Organisation, (2004) AS ISO 23081: *Information and Documentation – Records Management Processes – Metadata for Records, Part1: Principles*, 2004