Internationaler Archivkongress 2004

23.-29. August - Wien - Österreich

Archive, Gedächtnis und Wissen

Smart Metadata and the Archives of the Future

Sue McKemmish

Joanne Evans

Anne Gilliland-Swetland

Nadav Rouche

Richard Marciano

Hans Hofman

Assessing the Capabilities of Current and Future Metadata Schemas to Support the Creation and Preservation of Trustworthy Records:

The InterPARES 2 Metadata Schema Registry

Anne Gilliland-Swetland
Nadav Rouche
Department of Information Studies
University of California, Los Angeles

InterPARES Rationale

- The authenticity of electronic records is threatened whenever they are transmitted across *space* (i.e., when sent between persons, systems or applications) or *time* (i.e., either when they are stored offline, or when the hardware or software used to process, communicate, or maintain them is upgraded or replaced).

- Requirements for assessing the authenticity of electronic records that are preserved over the long term are necessary to support the presumption that an electronic record is and continues to be, what it purports to be and has not been modified or corrupted in essential respects.

- Preservation processes, mechanisms and metadata need to be identified that ensure these requirements are, and continue to be met.

InterPARES 2 (IP2), 2002-2006

- Second phase continues to analyze the nature and persistency of authenticity in the digital world, but draws in additional disciplinary perspectives

- Explores issues of reliability and accuracy throughout the life of the record.

- Expands the focus to include the digital by-products of artistic and scientific, as well as government activities carried out using increasingly complex interactive, dynamic, and experiential digital technologies.

Selected IP2 Description Group Activities

- Collect data on metadata being used in case studies
- Develop a database for analyzing warrant for metadata relating to creating reliable and preserving authentic digital materials.
- Develop and compile a metadata schema registry that describes and analyzes salient features of relevant extant descriptive and other metadata schema and standards.
- Develop and test metadata specifications relating to the activity, entity and data models developed by the Modeling Group which identify the type, source and application of metadata identified in the models, and the existence of relevant metadata schemas.
- Develop specifications for metadata management tools for activities such as automatic metadata creation and extraction.
- Interface with other relevant R&D activities such as ISO 23081 development, the Clever Metadata Project and the work of the SDSC on metadata tools development.

Why a Metadata Schema Registry?

Goal: to support the development and extension of metadata schemas that will support the creation and preservation of trustworthy records

- A centralized place for people to go to in order to understand which schemes or combination of schemas will best meet their specific needs for ensuring the creation and preservation of trustworthy records (supporting the ISO 23081)

- The creation of a meta-metadata model that can be applied in the assessment of existing and development of new metadata schemas

- A tool that can continue to be developed and used beyond the completion of the InterPARES2 Project

Registry Purpose

- To register, unambiguously, relevant metadata schemes and sets.
- To evaluate each scheme or set against the InterPARES' Benchmark and Baseline requirements.
- To make recommendations for how each scheme might be extended or otherwise revised to address the reliability, authenticity and preservation needs of records created within the domain, community or sector to which they pertain.
- To provide a standardized framework by which any metadata schema or set could be assessed for its ability to address these needs, and which could be adopted by standards-setting bodies.
- To provide input into the development of a Metadata Specification Model.

Designing the Metadata Schema Registry

- Operationalising InterPARES 1 Benchmark and Baseline Requirements (primary set of conceptual conditions to be met) and ISO 23081 requirements.

- Addressing ISO/IEC 1179 Information Technology – Metadata Registry (MDR) standard guidelines re: naming of elements and attributes.

- Other design inputs:
  - Relevant elements drawn from existing recordkeeping standards and requirements, including ISO 15489, DoD 5015.2, and MoReq.

- Iterative design process
  - Developing back end and front end separately
  - Pilot analysis of selected schemas to identify and refine registry elements, attributes, values, and capabilities.

Metadata Registry Schema Structure

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.

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. location of schema specification or documentation, 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.

Sub-Elements

- The 11 elements are further broken down into approximately 110 sub-elements, going three levels deep (i.e., up to sub-sub-elements).

Registration Element (partial):

Item	Element	Sub-element	Sub-sub-element
1	Registration		
2		Number	
3		Date	
4		Action Officer	Personal Name

Definition, Purpose and Comments

- Additional fields to capture definitions, purpose, comments and suggestions during iterative process of development.
- Provide for future supplementary documentation, such as Frequently Asked Questions (FAQ) and User Guidelines (abridged data).

Registration Element (partial):

Item	Definition	Purpose	Comments
1	Data to register a schema into the registry		
2	System assigned sequential number	To provide a system ID for the schema in the registry	As opposed to the Unique Global Identifier subelement which is created using a meaningful formula
3	The date of registration using ISO 8601 (yyyy-mm-dd)	To record the date of registration of the schema in the registry	Automatically assigned by the system
4	The name of the Action Officer		If the system will require log in, the Action Officer data entry will be automated using the user account information

Schema and Interface Rules, 1

- **Schema Rules:**
  - MS – Mandatory for Submission (minimum set of fields for submitting a schema)
  - MC – Mandatory for Completion (minimum set of fields for a complete schema record)
  - R – Repeatable
  - D – Default value
  - N/A – Not Available
- **Interface Rules:**
  - V – Viewable (e.g., Administrative Element)
  - ED – Editable by InterPARES Description Group members
  - EI – Editable by all InterPARES members
  - EN – Editable by non-InterPARES users

Data in Process of Being Analyzed

- Metadata schemas and sets identified in the course of the case studies of scientific, government, and arts settings/application/systems undertaken by the InterPARES 2 Focus Groups as well as by specialized studies undertaken by the Description Group.

- Other relevant Focus-specific metadata schemas and sets identified by Focus Groups or by the Description Group.

- Archival description rules, schemas, and related practices (e.g., ISAD(G)/ISAAR, EAD/EAC/DACS, RAD, and the Australian Series System).

Anticipated Outputs of the Registry

- Terms and proposed definitions relating to metadata and archival description submitted for inclusion in InterPARES 2 Glossary and Thesaurus.

- Recommendations for the extension of existing archival descriptive rules, schemas and practices to address requirements for authenticity, reproduction, and preservation of electronic records.

- Recommendations for the extension or revision of Focus-specific metadata schemas and sets.

- Analytical data on the adequacy of existing metadata schemas and sets that can be fed into the working group on the draft ISO 23081 Recordkeeping Metadata standard.

- Other policy and standards recommendations.

Metadata Specification Model

- To identify an overall set of metadata requirements that specify what metadata needs to be created, how, and by whom at all points within the Chain of Preservation and Records Continuum activity and data models being developed by InterPARES.

- To test and validate the metadata model through implementations of prototype metadata using case studies and scenarios

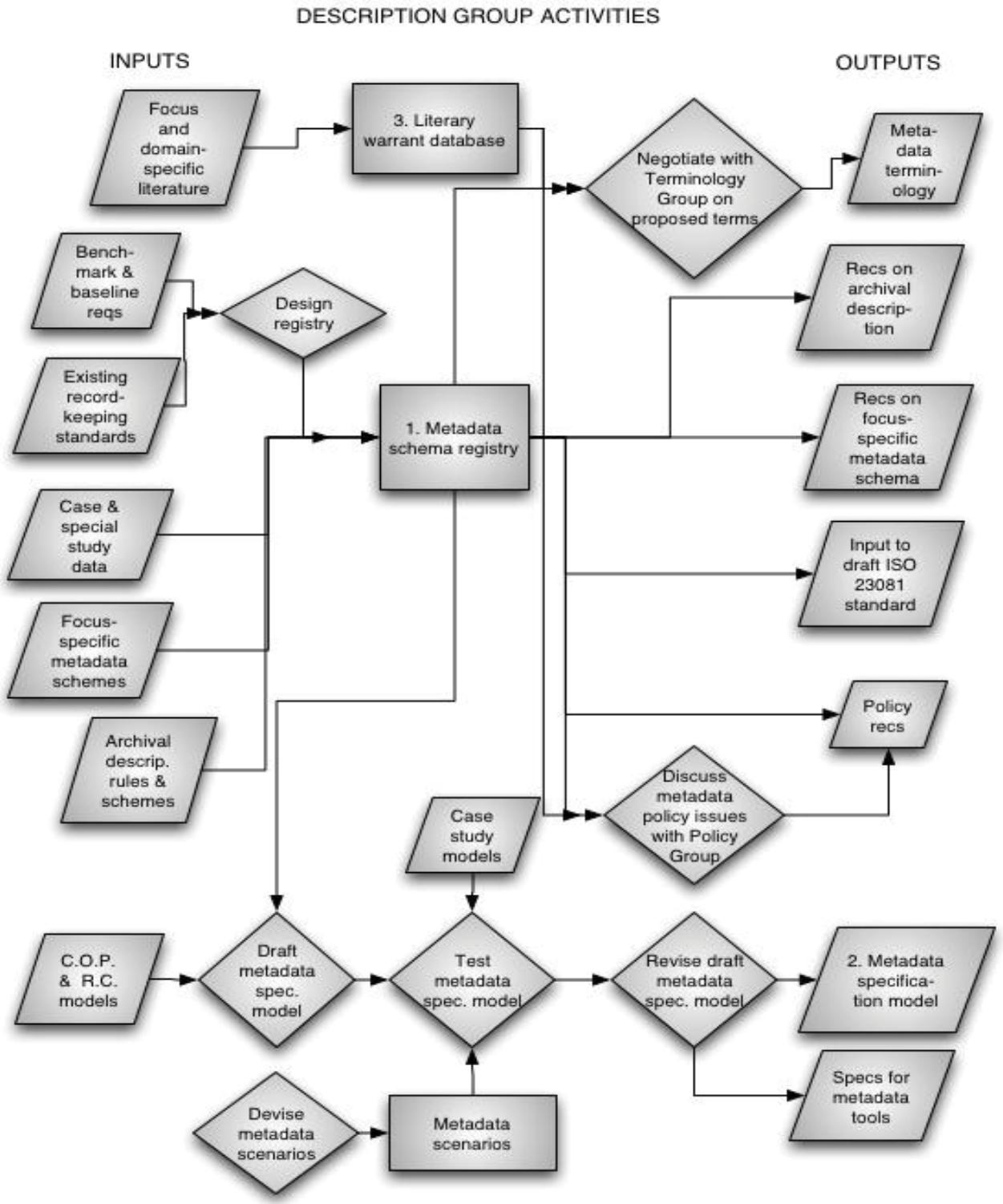
- To develop a set of specifications for automated tools that can be used to assist with the creation, capture, management, preservation and use of essential metadata for active and preserved records.

Inherent Considerations Needing to be Addressed by this Research

- The role of metadata in the creation and preservation of trustworthy records.

- Contested issues:
  - Delineation of the concept of archival description in relation to diverse notions of "metadata."
  - Life cycle vs. continuum views on archival roles.

- The creation and preservation of trustworthy metadata.


DESCRIPTION GROUP ACTIVITIES
INPUTS
OUTPUTS
Focus and domain-specific literature
3. Literary warrant database
Negotiate with Terminology Group on proposed terms
Meta-data termin-ology
Bench-mark & baseline reqs
Design registry
Recs on archival descrip-tion
Existing record-keeping standards
1. Metadata schema registry
Recs on focus-specific metadata schema
Case & special study data
Input to draft ISO 23081 standard
Focus-specific metadata schemes
Policy recs
Archival descrip. rules & schemes
Discuss metadata policy issues with Policy Group
Case study models
C.O.P. & R.C. models
Draft metadata spec. model
Test metadata spec. model
Revise draft metadata spec. model
2. Metadata specifica-tion model
Specs for metadata tools
Devise metadata scenarios
Metadata scenarios