



InterPARES 2 Project

International Research on Permanent Authentic Records in Electronic Systems

Description Cross-domain Task Force Research Design Statement

Revised 8 November 2004

This document outlines the research design to be followed by the Description Cross-domain Task Force (DTF) of InterPARES 2. It was originally developed in accordance with the SSHRC grant proposal, taking into consideration the plans of the Focus, Domain and other Cross-domain teams.

NOTE:

- (1) The document is updated after each Description Cross-domain plenary meeting to reflect the decisions taken by the team, and the ongoing work of the Focus, Domain and other Cross-domain teams. More detailed plans and timelines are added as the research progresses, as well as ongoing assignment of researcher responsibilities for individual activities and sub-activities.
- (2) The achievement of the objectives set out in the plan is dependent on continuing funding of the research work by the various sources that currently support the research work of the Description Cross-domain.

Scope and Expected Outcomes:

Given the importance of descriptive practices for ensuring the reliability, accuracy and authenticity of electronic records, for inferring or verifying such qualities over time, and for carrying out a proper appraisal and successful preservation of any such record, DTF will evaluate existing metadata schema, descriptive standards and tools using a variety of methods and instruments as outlined in the Research Design section of this document. It will develop guidelines and recommendations relating to the development or extension of schemas, standards and tools for the intellectual control of records from the moment of their creation throughout their appraisal and preservation, including schemas, standards and tools directed to the records creators and standards directed to records preservers.

The expected outcomes of this research are scholarly evaluation, comparative discussions and recommendations relating to existing schemas, standards and tools, and an intellectual framework for the extension and development of metadata schemas, descriptive standards and metadata and tools for the records under examination. The development of new standards is not an objective of the research at this time.

Description Research Questions:

- What is the role of descriptive schemas, standards and instruments in records creation, control, maintenance, appraisal, preservation and use in current traditional recordkeeping systems in the three focus areas?
- What is the role of descriptive schemas, standards and instruments in records creation, control, maintenance, appraisal, preservation, and use in emerging recordkeeping systems in digital and Web-based environments in the three focus areas? Do new schemas, standards and tools need to be developed, and if so, what should they be? If not, should present instruments be broadened, enriched, adapted?
- What is the role of descriptive schemas, standards and instruments in addressing reliability, accuracy and authenticity requirements (including the InterPARES 1 Benchmark and Baseline Authenticity Requirements, and the requirements of the new recordkeeping metadata standard, ISO/TS 23081) concerning the records to be investigated by InterPARES 2?
- What is the role of descriptive schemas, standards and instruments in archival processes concerned with the long-term preservation of the records in question?
- Do current interoperable frameworks support the interoperability of descriptive schema, standards and instruments across the three focus areas? If not, what kinds of frameworks are needed?
- What are the implications of the answers to the above questions for traditional archival descriptive standards, systems and strategies? Will they need to be modified to enable these sectors to meet these requirements, or will new ones need to be developed? If so, what should they be?
- To what extent do existing descriptive schemas, standards and instruments used in the other sectors concerned with the focus areas addressed by the project (e.g., the geo-spatial data community) support and inform requirements such as those developed by InterPARES 1? Will they need to be modified to enable these sectors to meet these requirements, or will new ones need to be developed? If so, what should they be?
- What is the relationship between the role of descriptive schemas, standards and instruments needed by the creator and those required by the preserver to support the archival processes of appraisal, preservation and dissemination? What tools are needed to support the export/import/exchange of descriptive data between systems?
- What is the role of descriptive schema, standards and instruments in rights management and in identifying and tracking records components, versions, expressions, performances and other manifestations, and derivative works?

- Is it important to be able to relate the record of artistic and scientific activity to the associated expression, performance, product, work or other manifestation of it and, if so, in what ways can descriptive practices facilitate this?

Research Design

The Description Cross-domain Task Force proposes to apply several methods in the conduct of the following specific research activities: literary warrant analysis, metadata mapping, the development of instruments to analyse metadata schemas, and descriptive standards, conceptual modeling, activity modeling, meta-modeling and empirical instantiation. The research design is essentially sequential: it provides for activities that will be ongoing throughout InterPARES 2; activities to be conducted by DTF that are independent of the timelines of any other InterPARES 2 task force; activities that can only take place as data are received from the Focus Task Force case studies, Domains and other Cross-domain teams; and project completion activities.

Activity 1. Identify a maximum of TWO questions that should be addressed by each Focus Task Force case study to gather data for the purposes of comparative analysis by DTF. **[DONE]**

Questions:

- 1) What descriptive or other metadata schema or standards are currently being used in the creation, maintenance and use of the recordkeeping system or environment being studied?
- 2) What is the source of this metadata (institutional convention, professional body, international standard, individual practice, etc.)?

Sub-activity A. Provide guidance and assistance to Focus Task Forces in relation to case studies on:

- Types of metadata relevant to DTF.
- Collection and analysis of information and documentation relating to metadata and description issues. **[ONGOING]**

Activity 2: Analyze literary warrant: Conduct a thorough literature review across all areas contained in each focus area to identify warrant relating to addressing the accuracy, reliability, authenticity and preservation of records by means of metadata schema, including archival description standards.

Subactivity A. Set up database to capture standardized literary warrant analyses. **[DONE]**

Subactivity B. Develop guidelines for using database and analyzing warrant, and train researchers in their use. **[DONE]**

Subactivity C. Identify and analyse literary warrant, populating the database with the outcomes. **[ONGOING]**

Subactivity D. Liaise with similar activities being undertaken by Focus Task Forces, Domains1-3 and Policy Cross-domain teams. **[ONGOING]**

Activity 3. Develop a database registering, describing and evaluating salient features of relevant extant metadata schema, including archival description standards.

Subactivity A. Develop a standardized Metadata Schema for registering, describing and evaluating recordkeeping-related metadata schemas and archival descriptive standards. Develop an XML-based DTD for a Web-based database that can be made accessible to InterPARES 2 researchers and build a prototype. Develop a production version for the public Web site.

- Develop Registration and Description sections of Schema. **[DONE]**
- Finalise Analysis section of Schema and associated Metadata Schema/Descriptive Standard Analysis and Evaluation Instrument (which enables evaluation against the requirements of the recordkeeping metadata standard ISO/TS 23081, the InterPARES 1 Benchmark and Baseline Requirements for Authenticity and the SPIRT Australian RKMS framework). **[DEC 2004]**
- Test, cross-validate and revise Schema Analysis and Evaluation Instrument through further round of instantiation. **[DEC 2004]**
- Develop controlled vocabularies for classification of purpose of schema and standards, and for types of metadata specified in schema and descriptive standards (drawing on recordkeeping metadata ISO/TS 23081, Monash SPIRT Recordkeeping Metadata Research Project outcomes, and InterPARES Models). **[DONE]**
- Develop XML DTD for the Metadata Schema and an XML Database of the Registry. **[DEC 2004]**
- Develop prototype of Registry. **[DONE]**
- Develop production version of Registry for InterPARES public Web site, including Web interface to Registry and Analysis and Evaluation Instruments. **[FEB 2005]**

Subactivity B. Identify and analyze relevant recordkeeping-related metadata schema and archival description standards, and register them in the database.

- Develop guidelines for registering, describing and analyzing schemas and standards. **[DONE]**
- Identify relevant schema and standards via case study input and canvassing relevant communities. **[ONGOING]**
- Establish processes for nomination of schemas and standards to be included in the Registry, and for registration, description, analysis and evaluation by external parties, and for provision of feedback to relevant communities on possible modifications / extensions to schemas and standards. **[MAR 2005]**
- Continue to identify, describe, analyse and evaluate relevant schema and standards, and populate Registry. **[ONGOING]**
- Provide input to recordkeeping metadata standard ISO/TS 23081 Part 3, Evaluation of Existing Metadata Schema. **[NOV 2005]**

Subactivity C. Explore the utility of classification and other schemes that are referenced in schema and standards registered in the database. [2005]

Subactivity D. Explore how metadata should be sourced to ensure their reliability and authenticity, e.g., through recordkeeping requirements for metadata registries. [2005]

Subactivity E. Explore the boundaries between/around records and related metadata, noting that some metadata relate to the content, structure (documentary form) and business context of the record (essentially these metadata are concerned with the nature of the business transaction captured in the record), and some of the metadata relate to the recordkeeping processes that manage the record. [2005]

Subactivity F. Map and otherwise model correspondences and other relationships between schema and standards. [2005]

Activity 4. Analyse the case study findings to address the following DTF research questions:

- What is the role of descriptive schemas, standards and instruments in records creation, control, maintenance, appraisal, preservation, and use in current traditional recordkeeping systems in the three focus areas?
- What is the role of descriptive schemas, standards and instruments in records creation, control, maintenance, appraisal, preservation, and use in emerging recordkeeping systems in digital and Web-based environments in the three focus areas? Do new schemas, standards and tools need to be developed, and if so, what should they be? If not, should present instruments be broadened, enriched, adapted?
- Is it important to be able to relate the record of artistic and scientific activity to the associated expression, performance, product, work, or other manifestation of it, and, if so, in what ways can descriptive practices facilitate this?
- What is the role of descriptive schema, standards and instruments in rights management and in identifying and tracking records components, versions, expressions, performances, and other manifestations, and derivative works?

Subactivity A. Review the DTF research questions in light of the case study findings so far. [FEB 2005]

Subactivity B. Develop a process and timeline for analysing the case study data to address the above DTF research questions, including analysis of final reports, raw data if required, case study models, and case study Focus Task Force sessions. [FEB 2005]

Activity 5. Assist the Modeling Team to develop the metadata and archival description related aspects of the activities and entities modelled in the Chain of Preservation Model and the Business Process Model, identifying the types of metadata referenced, as well as where, when and how they would be used, and indicating which existing metadata schema, descriptive standards and metadata tools could be specified as controls and resources in the models. [2005-2006]

Note: this is a joint activity of the Modeling and Description Cross-domain Task Forces.

Subactivity A. Identify the points in the Chain of Preservation and Business Process models where metadata, including archival descriptive elements should be created or captured; as well as their sources, and points of application.

- Provide input to inclusion of metadata and description processes in models as they are developed.
- Provide input into evolving models as they relate to specifying and modeling metadata, including archival description requirements (where metadata are captured, stored and applied, and how they are sourced, identifying the policies/rules/controls that relate to them) based on analyses and evaluations of schemas and standards in the Registry.
- Provide input to definition of metadata, including archival description related concepts and terminology in the evolving models.
- Work with Modeling Cross-domain in formally scheduled meetings at each research workshop, and develop timelines and assign responsibilities in liaison with Modeling team.

Subactivity B. Work with the Modeling team to identify how metadata might be represented in the modeling of case study results.

Activity 6. Testing and validation of metadata-related aspects of the models: test and validate metadata, including archival description aspects of the models by participating in the walkthroughs of the models based on scenarios developed by the Modeling team from selected case studies, e.g. Obsessed Again/NASA/Irish ROS/Alsacien Land Registry, with particular attention to the extent to which the metadata schema and descriptive standards applicable in the case study jurisdiction meet the needs identified in the models. **[2005-2006]**

Note: this is a joint activity of the Modeling and Description Cross-domain teams.

Subactivity A. Provide input relating to metadata, including archival description, to the development of a range of scenarios from a selection of case studies that reflect different disciplinary domains and jurisdictional and cultural settings.

Subactivity B. Identify existing metadata and related schemas and standards, in those scenarios.

Subactivity C. Assess the degree to which existing metadata, and related schemas and standards, address the requirements identified in the walkthroughs.

Subactivity D. Identify the ways in which existing metadata, related schemas and standards, could be enhanced to meet the requirements and gaps identified in the walkthroughs.

Subactivity E. Modify the metadata, including archival description-related aspects of the models to reflect the outcomes of the walkthroughs.

Activity 7. Undertake special studies.

Subactivity A. Variable Media Project relating to the description process and the preservation of descriptive records relating to artistic works using unstable and variable media (James Turner). **[FEB 2005]**

Subactivity B. Survey of News Archives and their preservation practices linked to work of PREMIS, the OCLC/LC committee on preservation implementation strategies (Vicky McCargar, LA Times). **[DEC 2005]**

Activity 8. Develop specifications for metadata management tools for activities such as automatic metadata creation and extraction and feedback into Activity 6. **[MID 2005-2006]**

Subactivity A. Identify needs for specific metadata tools that arise out of case study data, literary warrant analysis and evaluations of metadata schemas and descriptive standards.

Subactivity B. Identify and describe the functionality of existing metadata tools.

Subactivity C. Develop recommendations relating to the extension or development of metadata tools that might play a key role in supporting the creation and preservation of reliable and authentic records.

Activity 9. Establish formal relationship with Australian Research Council Linkage Grant Project “Create Once, Use Many Times – The Clever Use of Metadata in eGovernment and eBusiness Recordkeeping Processes in Networked Environments” to exchange research findings regarding metadata requirements. **[DONE]**

Activity 10. Establish formal relationship with ISO Committee working on Recordkeeping Metadata Standard to exchange findings regarding metadata requirements. **[IN PROGRESS]**

Subactivity A. ISO/TS 23081 Metadata for Records: Principles (issued early 2004) has informed development of Metadata Schema Registry (especially by providing framework for analysis and evaluation; and input to development of encoding schemes). **[DONE]**

Subactivity B. ISO/TS 23081 Metadata for Records: Evaluation of Metadata Schema (Part 3 of the Standard) will draw on InterPARES analysis instrument in InterPARES Metadata Schema Registry. **NOVEMBER 2005**

Activity 11. Development of DTF Vocabulary. **[ONGOING]**

Subactivity A. Build set of terms and definitions, drawing on Literary Warrant Analysis Database, Metadata Schema Registry, metadata tool analysis and other DTF documents.

Subactivity B. Develop thesaurus showing relationships between terms.

Subactivity C. Provide input to Terminology Cross-domain.

Activity 12. Progressive Dissemination of Research Outputs via InterPARES Web site, conference papers and journal articles.

Subactivity A. Develop internal Web site—interface and link to research plan, Literary Warrant Analysis Database, Metadata Schema Registry, DTF Vocabulary, Reports and Presentations, Related Project sites, Papers, etc. Note potential use of Joanne Evans' network/ontological mapping tool. **[ONGOING]**

Subactivity B. Develop publication schedule relating to DTF outcomes—a session at ICA Vienna (Aug 2004) focussed on the work of the DTF and linked projects and two of the papers will be published in *Archival Science* in late 2004; a paper by Evans, Lindberg, and Chia-Ning was presented at the Shanghai DC 2004 Conference (October 2004); a paper by Evans and Rouche on the research design and methods being used in the DTF Metadata Schema Registry work will be published in the special issue of *Archival Science* on archival research design and methods (early 2005); possible papers in conferences in a range of domains will be investigated, e.g., International Cartographic Association, Medical Librarians Association, IASSIST, ACMLA, Museums and the Web, ARLIS, CODATA. **[ONGOING]**

Subactivity C. Prepare slides with overview of DTF work for use by other InterPARES researchers. **[ONGOING]**

Subactivity D. Develop public pages relating to work of DTF. **[ONGOING]**

Activity 13. Draw together the outcomes of Activities 2-8 to address the following DTF research questions: **[2006]**

- What is the role of descriptive schemas, standards and instruments in records creation, control, maintenance, appraisal, preservation, and use in emerging recordkeeping systems in digital and Web-based environments in the three focus areas? Do new schemas, standards and tools need to be developed, and if so, what should they be? If not, should present instruments be broadened, enriched, adapted?
- What is the role of descriptive schemas, standards and instruments in addressing reliability, accuracy and authenticity requirements (including the InterPares1 Benchmark and Baseline Authenticity Requirements, and the new recordkeeping metadata ISO/Ts 23081) relating to the records being investigated by InterPARES 2?
- What is the role of descriptive schemas, standards and instruments in the archival processes concerned with the long-term preservation of the records in question?
- Do current interoperable frameworks support the interoperability of descriptive schema, standards and instruments across the three focus areas? If not, what kinds of frameworks are needed?
- What are the implications of the answers to the above questions for traditional archival descriptive standards, systems and strategies? Will they need to be modified to enable these sectors to meet these requirements, or will new ones need to be developed? If so, what should they be?

- To what extent do existing descriptive schemas, standards and instruments used in the sectors concerned with the focus areas addressed by the project (e.g., the geo-spatial data community) support and inform requirements such as those developed by InterPARES 1? Will they need to be modified to enable these sectors to meet these requirements, or will new ones need to be developed? If so, what should they be?
- What is the relationship between the role of descriptive schemas and instruments needed by the creator and those required by the preserver to support the archival processes of appraisal, preservation and dissemination? What tools are needed to support the export/import/exchange of descriptive data between systems?

Subactivity A. Using the outcomes of Activities 2-8, evaluate the extent to which existing and emerging metadata schemas, descriptive standards and metadata tools meet the metadata and archival description requirements identified, in particular those requirements relating to the creation, management, appraisal, preservation and use of records in the three focus areas, e-government, artistic activities and scientific activities.

Subactivity B. Identify gaps and make recommendations relating to the development of new metadata schemas, descriptive standards or metadata tools, or modifications / extensions to existing schemas/standards/tools to meet the unmet requirements, e.g., extensions to archival descriptive standards like ISAD(G)/ISAAR, or modification of recordkeeping metadata schema to include metadata that supports the creation and preservation of reliable and authentic records, and the long-term preservation and management of the metadata and archival descriptions themselves.

Subactivity C. Develop scholarly evaluation, and comparative discussions of existing, emerging and proposed metadata schemas, descriptive standards and metadata tools.

Subactivity D. Develop an intellectual framework that contextualizes the use of metadata, including archival description within the domains and focus areas being examined by InterPARES 2 and provides for the extension and development of relevant metadata schemas, descriptive standards and metadata tools.

Activity 14. Prepare DTF Final Report **[DEC 2006]**