Archival Science (2004) **4:** 315–334
DOI 10.1007/s10502-005-2598-4

Utilizing Systems Development Methods in Archival Systems Research: Building a Metadata Schema Registry

JOANNE EVANS[1] and NADAV ROUCHE[2]
[1]*Caulfield School of Information Technology, Monash University, Melbourne, Australia; (E-mail: jeeva2@student.monash.edu.au)* [2]*Graduate School of Education and Information Studies, Department of Information Studies, University of California, Los Angeles, USA; (E-mail: nadavr@ucla.edu)*

Abstract. This paper discusses the use of systems development research methods in the context of ongoing research that is formulating a specification for a metadata schema registry. The paper extrapolates from this discussion how such methods can be utilized in archival systems research and how such an approach differs from commercial systems engineering. In particular, the paper suggests that adopting a user-centred prototyping approach in a research context allows for exploration of the interplay between theory and practice, advancing the practice, while also offering new insights into theoretical concepts. It argues that these research methods are of increasing interest to the archival profession in order to conceptualise and realise the tools necessary to support recordkeeping and archival processes in digital environments.

Introduction

Archival systems, like other information systems, are undergoing radical change as the impacts of digital and network technologies on recordkeeping and archival processes are grappled with. Accustomed to dealing with mature systems and technology, the field of archival science is at a point where archival research needs to encompass methods that investigate how emerging theories are operationalized through systems development. Systems development research methods allow exploration of the interface between theory and practice, including their interplay with technology. Not only do such methods serve to advance archival practice, but they also serve to validate the theoretical concepts under investigation, challenge their assumptions, expose their limitations, and produce refinements in the light of new insights arising from the study of their implementation.[1]

[1] Frada Burstein, "Systems Development in Information Systems Research", in Kirsty Williamson (ed.), *Research Methods for Students, Academics and Professionals: Information Management and Systems*, 2nd edition, (Wagga Wagga, New South Wales: Centre for Information Studies, Charles Sturt University, 2002), pp. 147–158.

This paper explores the use of systems development research methods in the context of ongoing research that is formulating the specification of a registry of metadata schemas supporting recordkeeping requirements. It illustrates how the development of a prototype system has allowed researchers to gain familiarity and insight into the domain under investigation, the supporting technologies, and the requirements and role of such a registry in practice.

Background on the Development of the Metadata Schema Registry

Archival descriptive practices encompass the creation, capture or integration, and maintenance of metadata associated with records, in order to support the verification over time of their reliability, accuracy and authenticity. Within the InterPARES 2 research project, the Description Cross Domain research team is seeking to identify descriptive schemas and other instruments used in both the traditional and emerging digital recordkeeping practices of artistic, scientific and e-government focus areas.[2] The researchers aim to examine the role these metadata schemas play in records creation, control, maintenance, appraisal, preservation and use and whether they can support metadata requirements for ensuring the reliability, accuracy and authenticity of records through time.

Metadata schemas provide semantic and structural definitions of metadata, including the names of metadata elements, their structure, and their meaning.[3] Standards for archival description and specifications for archival control systems are examples of metadata schemas in the archival domain. Their purpose has been to identify and define the metadata necessary for the archival control of records. The development of XML as a structural encoding language has seen a proliferation of initiatives to define metadata schemas for information objects. Each initiative reflects the conceptual viewpoint of the community responsible for its construction with the structure and semantics of a metadata schema reflecting the purposes to which the

[2] For more information on the InterPARES 2 research project and the Description Cross Domain, see the website at http://www.interpares.org/ip2/ip2_index.cfm. A more complete discussion of the rationale underlying the registry development can be found in Anne Gilliland-Swetland, Nadav Rouche, Joanne Evans and Lori Lindberg, 'Towards a Twenty-First Century Metadata Infrastructure Supporting the Creation, Preservation and Use of Trustworthy Records: Developing the InterPARES 2 Metadata Schema Registry', forthcoming in *Archival Science*.

[3] Joanne Evans, Sue McKemmish and Karuna Bhoday, "Create Once, Use Many Times: The Clever Use of Recordkeeping Metadata for Multiple Archival Purposes", forthcoming in *Archival Science*.

metadata will be put. Metadata schemas may also be represented in a variety of ways for both human and machine processing.

To support their research, the Description Cross Domain researchers have identified the need to develop 'a database registering and describing salient features of relevant extant descriptive and other metadata schema and standards'[4]. The purpose of such a registry of metadata schemas is to act as a data collection and analysis tool to support comparative studies of the descriptive schemas. The results of such studies will feed into the guidelines for the development of standards for the intellectual control of records, a major task for the description team within the InterPARES 2 project.[5]

Beyond the specific needs of InterPARES 2 researchers, a secondary application of the metadata schema registry will be to provide a framework for the selection and evaluation of schemas by the recordkeeping community at large. The registry could be used to support the implementation of recordkeeping and archival systems by providing services supporting the evaluation and comparison of schemes according to defined requirements. It would enable users to discover existing metadata standards that may meet their needs and to also be aware of their recordkeeping capabilities or limitations. It could also foster standardization in identifying common requirements in order to reduce the proliferation of 'competing' schemas which make interoperability and interchange more difficult.[6] Thus, the registry could encourage the re-use of existing metadata schemas by supporting decision-making processes in the selection of a schema as a whole, or for the selection of groups of elements within a schema to support particular functions, such as administration, discovery, or rights management functions, to name a few. Furthermore, the selection and comparison process between schemas and the extraction of specific sets of elements within schemas could be partially automated.

With no existing blueprint for such a registry, the first task of the research team was to conceptualise the system and establish its requirements. In so doing several key questions are raised including: – What are the salient features of metadata schemas that need to be documented for the purposes outlined above? How are these realised as elements?

[4] Description Cross-domain Research Team, Research Design Statement, 30 September 2003, http://www.interpares.org/rws/display_file.cfm?doc = ip2_desc_research_design_methods(20030930).pdf.

[5] *ibid.*

[6] Michael Day, "Integrating Metadata Schema Registries with Digital Preservation Systems to Support Interoperability: a Proposal", *DC 2003: Supporting Communities of Discourse and Practice – Metadata Research Applications, Seattle, Washington (USA), September 28th–October 2nd, 28th September–2nd October 2003*, http://www.siderean.com/dc2003/101_paper38.pdf.

To what entities do they relate? From where are their values to be sourced? How is an assessment of recordkeeping and archival capabilities to be determined? And what workflow processes must the registry support? In order to address these questions, the research team looked at utilizing systems development as an exploratory research approach.

Systems Development Research Methods

The use of systems development as a research method has been discussed at some length in the information systems literature, although it is conspicuously lacking in the archival literature. Nunamaker, Chen and Purdin argue the case for the inclusion of systems development as a pivotal part of 'a multimethodological approach to IS research'.[7] Developing a system in a research context can serve as 'proof-by-demonstration' of the underlying theory, as well as producing an artefact which can form the basis of ongoing and expanded research. They state that 'building a system in and of itself does not constitute research' but that the process of expressing new concepts or technologies in 'a tangible product' can afford validation and insight into its theoretical underpinnings. Burstein elaborates on the process for such a systems development research approach, suggesting three major iterative stages – concept building, system building and system evaluation (see Figure 1).[8] The concept building phase involves the identification and development of the research questions and investigation of the system requirements and functionality, incorporating relevant ideas and approaches from other disciplines. The system building phase involves constructing the system using systems development techniques and the systems evaluation phase involves analysing and assessing the system.

Burstein emphasizes both the iterative and recursive nature of the method, where research 'can be conducted as a sequence of related projects, where each complements the others in a full cycle of theory development and testing throughout the system development.'[9] Burstein also addresses the crucial issue of the differences between conventional systems development and systems development research:-

[7] Jay F. Nunamaker, J.R. Minder Chen and Titus D.M. Purdin, ''Systems Development in Information Systems Research'', *Journal of Management Information Systems*, vol. 7, no. 3, (Winter 1990–1991), pp. 89–106.

[8] Burstein, op. cit., p. 153.

[9] *ibid.*

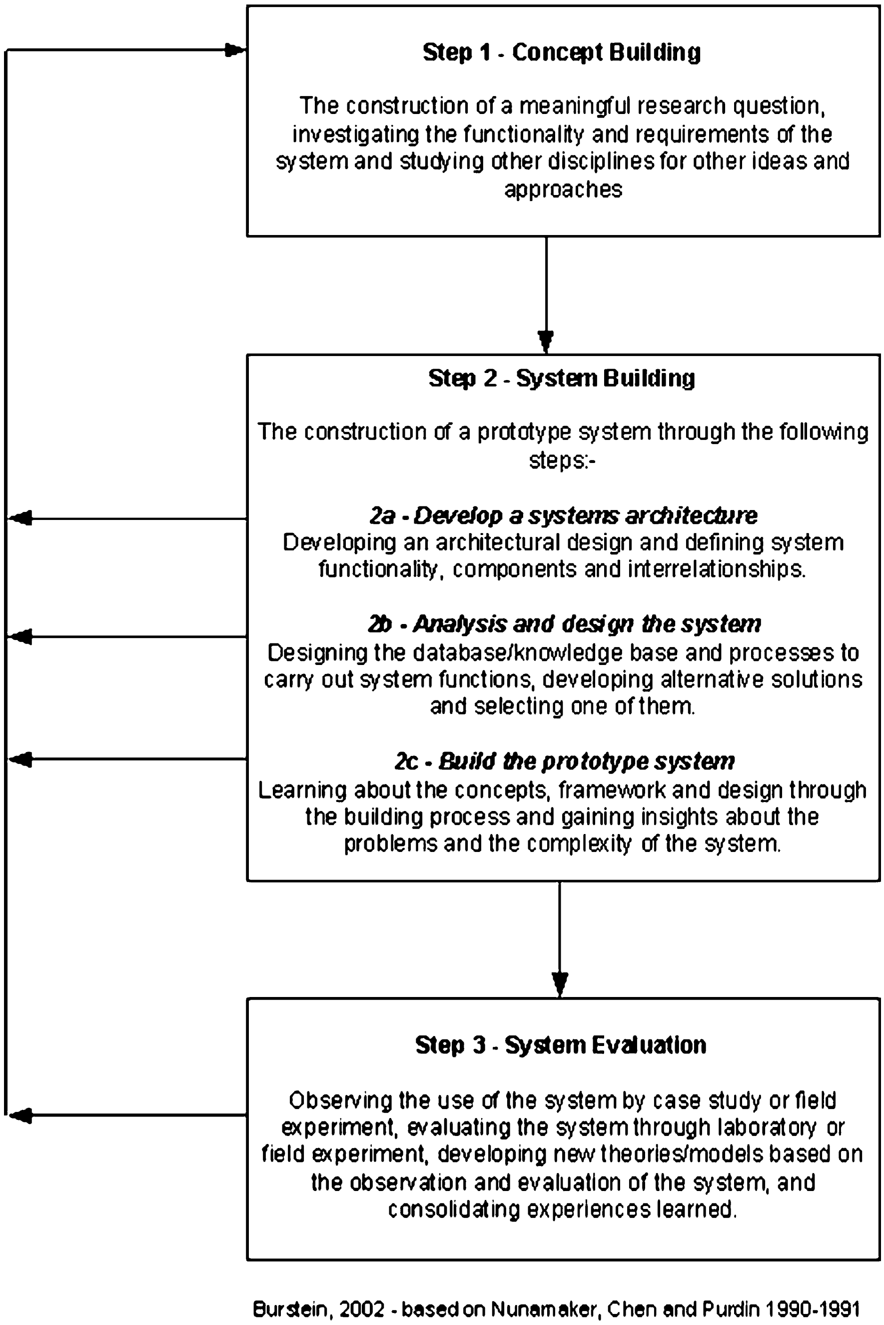

Figure 1. The systems development method.

> The major difference between this approach as a research method and conventional systems development is that the major emphasis is on the concept that the system has to illustrate, and not so much on the quality of the system implementation. At the beginning of such a project the implementation has to be justified in terms of whether there is another existing system capable of demonstrating the features of the concept under investigation. The evaluation stage of the systems development method is also different from the testing of a commercial system. It has to be done from the perspective of the research questions set up during the concept-building stage, and the functionality of the system is very much a secondary issue.[10]

In utilizing systems development as a research approach, all stages of development must reflect this focus on the concept that the system is to illustrate. It must be shown in the initial justification of the research approach that no existing system could be used to explore the concepts under investigation. In the system building stage, adequate documentation of the design, analysis and build processes must be collected to show how understandings and insights into the research domain are gained. And in the system evaluation stage, the evaluation should be from the perspective of how the systems development process and the resulting artefact address the research questions. By contrast in conventional systems development, commercial decisions justify the approach taken, and evaluation and testing are concerned with measuring performance and functionality – issues of a lesser importance when using systems development methods in a research context.

Traditional techniques for systems engineering involve first developing a specification and then building the system from that specification. The implication of employing such techniques is that there is, *a priori*, a comprehensive understanding of both the requirements and their interaction with the technology. However, such techniques may not be suitable for a research process where the focus is on developing understandings of what the requirements might be and on exploring what is possible with the technology. This raises the question of whether in such situations it may be more appropriate to employ emerging user-centred rapid application development techniques, with their short cycles of design, implementation and evaluation, to iteratively gather the necessary data in order to delineate the system and technological requirements.

[10] *ibid.*

Metadata Registries

Metadata registries are a burgeoning area of research. In the computer science and software engineering disciplines, metadata registries have evolved from data dictionaries and are seen as key parts of component based system architectures. The international standard, *ISO 11179 Information Technology – Metadata Registries*, has been developed specifying the types of semantic metadata about data elements that should be stored in metadata registries.[11] The electronic business community has also developed a registry specification as part of their ebXML suite of standards to support electronic commerce.[12] Their focus is on capturing and storing syntactic and administrative metadata about elements in XML schemas, as well as specifying the functionality or the methods other systems components can use to access and interact with this data. While some useful background and ideas may be garnered from these initiatives, these types of registries are not suitable for the research needs of the Description Cross Domain. Their focus is on metadata elements rather than metadata schemas as a whole, and on capturing data for machine processing rather than aiding human understanding. Furthermore they are limited to managing metadata in time rather than through time, a key requirement for a registry that is to be deployed for archival purposes.

The information management community, particularly in the digital libraries area, is another discipline involved with metadata registry research. They are exploring the use of metadata registries to facilitate the interchange of metadata about information resources for discovery and retrieval and as part of preservation frameworks. Systems development research approaches have been adopted in this area with a number of prototypes developed. The DESIRE, SCHEMAS and CORES projects are examples of successive metadata registry developments using exploratory, evolutionary prototypes to further understanding

[11] *ISO/IEC 11179, Information Technology – Metadata Registries (MDR)* is a six part standard, previously known as *ISO/IEC 11179–1, Information Technology – Specification and Standardization of Data Elements*. It is made up of: – Part 1 Framework, Part 2 Classification for Administered Items, Part 3 Registry Metamodel and Basic Attributes, Part 4 Formulation of Data Definitions, Part 5 Naming and Identification Principles, and Part 6 Registration.

[12] OASIS/ebXML Registry Technical Committee, *OASIS/ebXML Registry Information Model v2.5*, OASIS, June 2003, http://www.oasis-open.org/committees/regrep/documents/2.5/specs/ebrim-2.5.pdf, and OASIS/ebXML Registry Technical Committee, *OASIS/ebXML Registry Service Specification v2.5*, OASIS, June 2003, http://www.oasis-open.org/committees/regrep/documents/2.5/specs/ebrs-2.5. pdf.

and elucidate requirements.[13] In 2001–2002 a number of researchers in this area came together as a Working Group on Registries under the umbrella of the DELOS Network of Excellence on Digital Libraries to produce a white paper articulating a shared set of principles for the construction of metadata registries. This report aimed at developing 'a rough consensus' on common understandings and terminology based on their experiences in pioneering registry activities, to serve as the basis for further research and development.[14] The focus in these projects has been on capturing metadata about elements that are part of metadata sets, defined in formal schemas or as application profiles. Their aims are to foster extension and evolution of existing metadata sets and to capture the relationships between data elements across metadata sets to support automated translation. In these prototypes the metadata captured at the schema level is limited. It mainly consists of identifying the schema with a url/uri, naming it, and briefly describing it, as well as identifying the responsible authority.

Thus, while there has been a lot of research activity in the metadata registry area, none has explored requirements for describing metadata schemas from a recordkeeping perspective. However, much of the research to date has involved adopting systems development methods to create prototypes in order to gain understanding of the domain and to assess different approaches. This suggests that a similar approach for eliciting the requirements for a registry addressing the recordkeeping and archival qualities of metadata schemas would be fruitful.

Although optimisation of functionality and performance of a research prototype is of lesser concern than it might be for an operational system, once the conceptual stage and the research rationale for the development of the registry have been established, the research project still needs to take into account budget, resources and timelines as it moves into the implementation stage of development, when

[13] DESIRE Metadata Registry Framework (1998 – 2000) – implementation of a prototype metadata registry to 'enable authoritative information about metadata schemes to be declared and thus support the extensibility and evolution of element sets and provide some basis for interoperability.' The prototype registry is available at http://desire.ukoln.ac.uk/registry/. SCHEMAS: Forum for Metadata Schema Designers and Implementers (2000–2001) – As part of this forum for metadata schema designers a registry of metadata projects based on the DESIRE prototype has been established to facilitate the 'disclosure of new and emerging metadata standards' and to 'support development of best practice usage of metadata schemas.' See http://www.schemas-forum.org/. CORES – A Forum on Shared Metadata Vocabularies (2002–2003) – a further iteration of a metadata schema registry 'to encourage the sharing of metadata semantics' through developing consensus on the expression of standard definitions of terms, local usage and adaptations in a machine-readable way.

[14] DELOS Working Group on Registries, *Principles of Metadata Registries*, white paper, DELOS Network of Excellence on Digital Libraries, 2002, http://delos-noe.iei.pi.cnr.it/activities/standardizationforum/Registries.pdf.

functionality of the system and the granularity of the schema come to the fore. Taking into account these considerations, the research development team needs to implement a system that will support and ease the researchers' work by streamlining the process of registering and comparing schemas in an efficient way, while also taking into account the additional time and effort invested in the development of the registry itself. Otherwise, establishing some form of workflow management between researchers using spreadsheets and word documents would better serve the needs of the research, and save time and resources.

Developing a Specification for a Metadata Schema Registry

As with any systems development project, the process begins with an expression of the scope of the work and an initial outline of requirements. In this case, this document was produced by an expert group of researchers within the Description Cross Domain and consisted of a set of purposes for the registry and identification of descriptive features grouped by categories. An overview of this document is presented in Table I.[15]

The Description Cross Domain is made up of recordkeeping researchers, computer scientists and representatives from the three areas under investigation – artistic, scientific and governmental – and thus acts as a corpus of diverse expertise to guide, assess and evaluate the metadata schema registry development. A decision was made early in the research to develop an Extensible Markup Language (XML) encoding of the descriptive schema for this metadata about metadata schemas. XML provides the mechanism to encode structure in a universal format and also allows for the development of the registry as a web application.

Requirements to Specification

From the initial requirements document, the research team then undertook a process of progressive refinement in order to produce a detailed specification for a production version of the registry. The process began by breaking down the descriptive features into a detailed hierarchy of elements and sub-elements using a spreadsheet

[15] Initial purpose and requirements document for Registry of Metadata Schema Relevant to Recordkeeping and Archival Functions developed by Anne Gilliland-Swetland, Hans Hofmann, Bill Underwood and Sue McKemmish.

Table I. Initial Requirements for a Metadata Schema Registry Relevant to Recordkeeping and Archival Functions

Purposes	• To describe relevant metadata schemas and their features in a standardized way • To provide an overview of existing and emerging schemas • To provide an overview of the applicability of the schemas to recordkeeping and archival functions • To describe the scope and purpose of the schemas • To specify what type of metadata they cover • To identify related schemes (e.g. schemes that control data values, schemes that provide structure for metadata elements)
Descriptive features	
Registration	Data elements to register metadata schemas into the registry, i.e. registration number, date and action officer.
Identification	Data elements to identify and distinguish metadata schemas, i.e. title, unique global identifier, version, publication statements, etc.
Accessibility	Data elements to capture information relating to the accessibility of a schema, i.e. hardware and software requirements, etc.
Rights	Data elements to capture intellectual property rights associated with the use of a metadata schema.
Provenance	Data elements to capture organisations or other bodies/agents associated with the development, publication and maintenance of a metadata schema.
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.
Analysis	Data elements for analysing a metadata schema or data elements for capturing the results of analysis of a metadata schema against recordkeeping requirements.
Documentation	Data elements for capturing citations to the documentation of a metadata schema, e.g. specification, guidelines, etc.
Relationships	Data elements to capture relationships amongst metadata schema and to other classification schemes
Administration	Data elements for the administration of the schema registry.

(Table II), along with additional columns for definitions, descriptions of purpose and researchers comments (Table III, representing the same items as in Table II). Capturing this data was essential in facilitating an iterative collaboration process and also providing the basis

Table II. Hierarchy of Elements and Sub-elements

Hierarchy				
Item	Element	Subelement	Sub-subelement	Sub-sub-subelement
1	*Registration*			
2		Number		
3		Date		
4		Action Officer	Personal Name	

for future supplementary documentation such as user guidelines and frequently asked questions.

While the initial descriptive features requirements provided a top-down approach to hierarchy development, the iterative process of elaborating and refining the hierarchy sub-elements provided a bottom-up feedback loop into the definitions of the initial requirements themselves.

As researchers reviewed and provided feedback on the registry structure, additional columns were included to capture interface and schema rules for each element (see Table IV).

Although the schema structure was still a work in progress, it appeared to be a useful exercise to consider interface and schema rules at this early stage of development in order to provide an additional and refreshing perspective of the elements, their values, and their purpose as it pertains to identified user groups. Elements were tagged as mandatory, optional, and/or repeatable and default values were identified where applicable. Considering interface rules early on in the process forced the researchers to clearly identify and define target users (such as InterPARES members and non-InterPARES users), and by doing so, better understand what kind of elements would be of use to them and in what ways.

Another benefit of this exercise was that instead of focusing on abstract concepts, it helped the designers to visualize a system and think in terms of its requirements. For example, as user groups were identified and fields were tagged as mandatory or optional, workflow management requirements came to the fore. These in turn helped shape the ADMINISTRATIVE element to include workflow management sub-elements, such as submitted, completed, and approved. Additionally, initial data types were identified. Finally, it also laid the basis for future interface design and system development requirements.

Table III. Definition, Purpose and Comments

Item	Definition	Purpose	Comments
1	Data elements to register metadata schemas into the registry		
2	System assigned sequential number	The system ID for the schema in the registry	As opposed to the Unique Global Identifier sub-element 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 will be populated automatically using the user account information

Table IV. Schema Rules, Interface Rules, Data Type, and Instantiation

	Schema Rules		Interface Rules						
Item	Mandatory	Repeatable	Default Values	Viewable by all users	Editable by all InterPARES members	Editable by non InterPARES users	Editable by InterPARES description domain members only	Data Type	Instantiation-ISO 19115: Geographic Information Metadata
1									
2	MS		N/A	X				integer	1
3	MS		N/A	X				date	2004.04.23
4	MS		N/A	X	X	X	X	string	John Doe

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).
N/A – Not Available.

Prototyping and Instantiation

With unresolved questions arising from the refinement process, it was apparent that there was a need to move to an instantiation of the registry with real world examples as part of the development of the specification. While the ultimate aim is to create a web-based application, development of a prototype at this stage, using available expertise within the research team, was seen as a way to foster understanding of the research domain and how the research needs could be met with technology. This is typical of how prototyping techniques can be used in systems analysis and development to ensure that all involved in the process – those commissioning a system, those developing a system, and those using a system – can learn about a problem or solutions to a problem.[16] Prototyping is characterised by a high degree of user involvement (in this case the users are the researchers), and establishes a learning process for both users and developers so that common understandings about the system and the impacts of technology on the system are reached.[17]

The role of the prototype is, therefore, to explore the instantiation of the proposed requirements for a metadata schema registry and to develop common understandings across a distributed research team. In this case the research team is both geographically distributed and its members are from different disciplines. The goal is to gain understanding of the concepts under investigation through the process of creating the prototype, as well as having an artefact as the basis for further study. Prototyping in this way also helps to make explicit the compromises that occur when moving from a system design to an actual system. Choices as to which compromises to make can be placed in the hands of the users, in this case the researchers, rather than the technologists. All parties can see the impact of a decision and evaluate the consequences – intended and otherwise – and adjust the specification and their thinking accordingly. Often in traditional or commercial systems development projects, these decisions would be shielded from the user. In a research environment it is important that these decision-making processes are transparent as they can often lead to new insight into the underlying concepts.

[16] Bill Underwood, "Systems Analysis and Modeling for Archival Scientists" Version 1.0, November 1999.

[17] Mahil Carr and June Verner, "Prototyping and Software Development Approaches", Working Paper Series, 97/04, Information Systems Department, City University of Hong Kong, 1997, http://www.is.cityu.edu.hk/Research/WorkingPapers/paper/9704.pdf, p. 7.

The prototyping for the registry involved using 'lightweight' or 'agile' methods for the rapid development of a system in which instantiations of the descriptive features could be captured and explored. The appeal of agile practices is that they favour iterative production of working systems that can respond to change and foster collaboration over more bureaucratic processes involving a design stage followed by a 'contract to build' stage.[18] They also favour people over processes and tools, encouraging the use of available expertise and technology.[19] In this case, the existing relational database and web application development experience and skills within the research team were utilized to build an exploratory and ultimately 'throw away' prototype in order to aid in the development of a detailed specification of the registry.

Building the prototype as a relational database introduced an alternate model of the descriptive features proposed for the metadata schema registry. The view adopted in the requirements-to-specification-refinement process is one of a hierarchy of elements, whereas a relational database view is based on a model of entities, attributes and relationships. The decision to develop the prototype as a relational database was one of expediency. Skills existed within the project team to build such a system quickly. Metadata about metadata schemas could begin to be captured and the proposed elements and their structures and values tested. Additionally interface issues of how best to support the data entry could be explored. At the same time, it was also felt that the specification process would benefit from developing a draft XML DTD of the registry schema as part of the refinement of the specification, so that the requirements and their interaction with the XML technology could be explored. Hence methods of translating the data from the relational database into the XML representation were also devised.

Prototyping and Instantiation Outcomes

Controlling complexity and enabling flexibility

The prototyping and instantiation process aided the development of the specification in a number of ways. One was in controlling

[18] Martin Fowler, 'The New Methodology', 2003, http://www.martinfowler.com/articles/new-Methodology.html.

[19] Robert Martin, *Agile Software Development: Principles, Patterns, and Practices*, (Upper Saddle River, New Jersey: Pearson Education, 2003), 529 pp.

complexity and enabling flexibility. When prototyping began, the specification was already quite complex, with a proliferation of very specific elements. A major aim of relational database design is to eliminate redundancy, so modeling the elements in such a way helped to identify those which could be captured using more generic structures. For example, in the provenance section, the initial proposal was for separate OWNER, MAINTENANCE AGENCY, STANDARDS AGENCY and PUBLISHER elements. However, the view taken of this when modelled in the relational database was that these are all roles that agents play in relation to a metadata schema. Hence it was instantiated as an AGENT table with provenance – that is the relationship of an agent to a metadata schema – captured in a separate table (PROVENANCE) with a field to capture the role the agent plays in relation to the metadata set (see Figure 2).

In order to deal with the complexity of the real world, therefore, the specification was modified to incorporate a sub-element of the PROVENANCE element which could capture roles using an extensible vocabulary, rather than embedding the roles into the schema structure. Such an approach ensures flexibility where any number of different roles that agents play in relation to metadata sets can be captured without requiring a structural alteration. In the initial population of the registry, variation in the values for ROLE can be allowed and the decision may then be taken at a later stage to constrain the values with a controlled vocabulary. This design principle can then be applied to other elements in the specification.

Establishing Feasibility, Applicability and Sources of Values

Empirical instantiations also enabled the testing of the feasibility and applicability of the proposed elements and determination of the

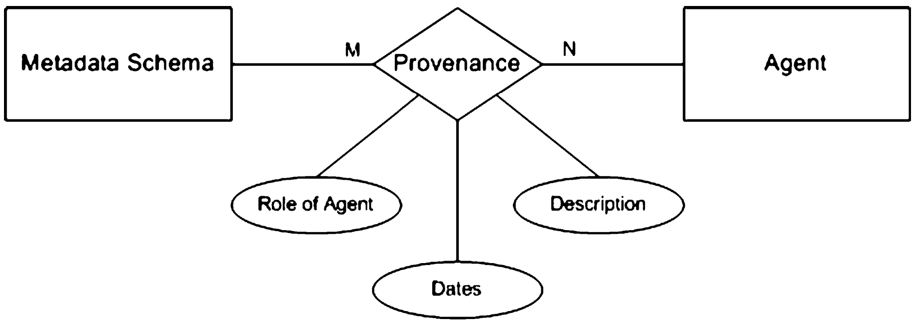

Figure 2. ER diagram of provenance.

sources of metadata values. Upon population of the database, it immediately became apparent that variation in the way metadata sets and standards are documented would have an impact on the ability to source some metadata values. Elements were proposed by members of the research team based on expectations that the data would be available across the wide variety of different types of schemas that the registry is expected to cover. Instantiating at the specification stage enabled those expectations to be tested and the viability and usability of elements for the research purposes to be assessed. One of the areas where this was the case was in the ACCESSIBILITY section. The question arose as to whether meaningful data could be captured in this section about such technical aspects as hardware, software, and encoding for metadata standards, or whether these features instead apply to implementations of a standard. If the latter is the case, and further instantiations would need to be collected before a definitive answer could be reached, the specification would need to be modified accordingly.

Another aspect of the elements evaluation is to assess what useful metadata, that is, information about a metadata schema, is not being provided by their publishers. One of the elements proposed in the documentation section is a link to a metadata schema's data model. A data model is a representation of the view underpinning a metadata set, delineating which objects or entities, relationships and attributes are represented and why. A data model is essential for understanding a metadata set, and making judgements as to its applicability. It is also necessary when considering mappings between metadata schemas. What is surprising in the initial population of the registry is that this information has not been readily available and often locating it requires a detailed examination of all the documentation, if it is provided at all.

Analysing Recordkeeping and Archival Capabilities

Analysis of the recordkeeping and archival capabilities of metadata schemas is a key purpose of this registry. Hence in setting out their descriptive requirements, the Description Cross Domain researchers identified the need for an analysis section and outlined possible elements. When it came time to instantiate this section in the prototyping process, the key question became one of how the values for the proposed elements would be determined. Unlike in other sections, this metadata would not be available in the schema documentation, thus there was a need to develop an analytical instrument that could be used to derive values in an authoritative manner.

Before even considering any possible element structures, therefore, the analysis process had to be defined. For this, the research team looked to available literary warrants, that is, authoritative sources of professional best practice,[20] in order to determine how the archival and recordkeeping capabilities of metadata schemas could be assessed. The warrants used in the analysis process are the emerging ISO standard, *ISO/TS 23081–1:2004 Information and documentation – Records management processes – Metadata for records – Part 1: Principles*[21] and the Benchmark and Baseline Requirements for assessing and maintaining the authenticity of electronic records of the initial InterPARES project.[22] The definition of the process revealed a complexity that the researchers had not considered in their original outline. The process involves the completion of an analysis worksheet that maps a metadata schema against recordkeeping and archival requirements as expressed in the above instruments. Results from this process are then summarised in a document where judgements are made as to the degree to which the requirements are met. At this stage, these documents are recorded in the documentation section of the registry, with the aim of determining what useful metadata structures could be derived from them at a later date.[23]

Documenting Schemes

A major purpose of the registry is to identify the encoding schemes associated with metadata sets and standards. Schemes can be used to control data values by providing a controlled vocabulary for the metadata content, or to provide the format or the structure of metadata elements. For example, many schemas point to *ISO 8601 Data elements and interchange formats – Information interchange –*

[20] Wendy Duff, "Harnessing the Power of Warrant", *American Archivist*, vol. 61, (Spring 1998), pp. 88–105.

[21] *ISO/TS 23081–1:2004 Information and documentation – Records management processes – Metadata for records – Part 1: Principles*, (International Organization for Standardization, 2004), 18 pp.

[22] Benchmark and Baseline Requirements from the InterPARES Authenticity Task Force, "Requirements for Assessing and Maintaining the Authenticity of Electronic Records", *The Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project*, Appendix 2, http://www.interpares.org/book/interpares_book_k_app02.pdf.

[23] For more details of the description and analysis process see Joanne Evans and Lori Lindberg, "Describing and Analyzing the Recordkeeping Capabilities of Metadata Sets" in *DC-2004: Proceedings of the International Conference on Dublin Core and Metadata Applications, Shanghai, China, October 11–14 2004.*, Shanghai Scientific and Technological Literature Publishing House, Shanghai, China, 2004, pp. 75–80.

Representation of dates and times as the format for date elements. Instantiations revealed the need to consider allowing for relationships among schemes to be recorded – for example where one standard derives a scheme for an element from that used in another standard – and, therefore, also to consider including functionality in the registry to support a 'schemes view' of the metadata. This would enable the registry to derive relationships between schemas which may not be explicit in the documentation based on their use of schemes. In terms of the overall research aims, it suggests that investigation of mechanisms to manage metadata about these schemes through time and space may also be required as part of preservation frameworks.

Discovering Existing Crosswalks

Crosswalks are mappings of one metadata element set to another. A key issue is how crosswalks can be efficiently generated to enable the interchange of metadata between element sets. The consensus is that manual creation of crosswalks is a resource intensive and ultimately unsustainable process, but more accurate and authoritative than semi-automated mechanisms.

The instantiation process revealed that crosswalks to related metadata sets are often supplied as part of the documentation of a schema. There also exist a number of crosswalks generated by third parties. This proliferation of competing crosswalks may in part be caused by lack of discoverability, suggesting that documenting their existence within the metadata schema registry would be useful for both the research purposes and for any broader community of users of the system. Hence capture of metadata about crosswalks is being trialled in the prototype registry and the draft XML DTD, with the knowledge that the specification may have to be revised once the required data structures and functionality have been agreed upon. Again it points to the need to view the metadata in the registry from the perspective of the crosswalk.

Conclusion

Having moved through the development process outlined above, the researchers have found that the view of the registry is no longer one of just metadata about metadata schemas. Prototyping and instantiation activities have pointed to the potential of supporting views of the metadata from the perspective of agents, schemes, documentation, implementations and the associations between them. The cycles of implementation and evaluation involved in the user-centred prototyping approach have enabled the research team to assess the practicality of the emerging specification as well as to gain a deeper understanding of the requirements and how technology might be used to realise them.

In addition, the consideration of interface rules at this development stage has helped to gain further insight into the element hierarchy, purpose and use. By defining user groups, the purpose and potential uses of elements have become more focused as to how they relate to these groups and their needs. An additional benefit of this approach is to lay the foundation for user interface and system requirements for the production version of the registry. Visualizing and thinking in terms of a system in this way has provided new perspectives on the metadata requirements. Moving from an abstract design approach into a more concrete one has revitalized the conceptual work.

Exploration of the interaction between theory and practice is a crucial part of archival systems research, especially in the pursuit of requirements for digital recordkeeping within ever-evolving technological frameworks. Systems development methods may play an increasing part in research regarding this interaction as we conceptualise and build the tools needed to support recordkeeping processes in digital and network environments. The advantage of using systems development as a research approach is that it not only develops the practice, but it also serves to deepen theoretical understandings and ultimately ensure that new technologies can be made to serve archival science.

Acknowledgements

The authors acknowledge the funding support of InterPARES by the Social Sciences and Humanities Research Council of Canada, the United States' National Historical Publications and Records Commission and the National Science Foundation. Financial support for Joanne Evans has been provided through an APA(I) Scholarship as part of the *Create Once, Use Many Times – The Clever Use of Metadata in eGovernment and eBusiness Recordkeeping Processes in Networked Environments* Australian Resource Council (ARC) Linkage Grant 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.