[DRAFT]

"RECORDKEEPING METADATA"* WORKSHOP

*Structured or semi-structured information which enables the creation, management, and use of records through time and within and across domains in which they are created. Recordkeeping metadata can be used to identify, authenticate, and contextualize records; and the people, processes and systems that create, manage, and maintain and use them.

PROCEEDINGS OF THE ARCHIVING METADATA FORUM

June 5-8, 2000
The Netherlands

Wolfheze (June 5-6)
Castle of Bergh (June 7)
The Hague (June 8)

Sponsored by:
Archief School,
Nederlands Instituut voor Archiefonderwijs en Onderzoek
(Netherlands Institute for Archival Education and Research)

Proceedings produced by:
David A. Wallace
University of Michigan, USA
(September 2000)

Table of Contents

PAGE

Agenda

Monday June 5, 2000

Location: Bilderberg Hotel, Wolfheze

First working session
9.00 – 12.30

Introduction to the workshop
Further elaboration on the agenda
Discussion on objectives, perspectives and expectations
Resource persons: Hans Hofman, Peter Horsman

Second working session
14.00 – 17.00

Recordkeeping metadata within a broader metadata framework
What makes recordkeeping metadata different
Objectives of recordkeeping metadata
Resource person: Barbara Reed

Third working session
17.30 – 20.00

Recordkeeping metadata, a functional perspective
Recordkeeping metadata and recordkeeping functions
Resource person: Anne Gilliland-Swetland

Tuesday, June 6, 2000

Fourth Working Session
9.00 – 12.30

Emerging standards
Need of standards
Resource person: Wendy Duff

Fifth Working session
14.00 – 17.00

Connections between the themes
The themes revisited
Adequacy and applicability
Discovering "white spots"
Resource person: Sue McKemmish

Wednesday June 7, 2000

Location: Castle of Bergh

Sixth working session
9.00 – 12.30

Research agenda
What, how, by whom
Cooperation
Towards a network of excellence
Infrastructure for research
Resource person: Margaret Hedstrom

Seventh Working session
14.00 – 16.00

Research agenda (continued)
Follow-up
Adjournment
Resource persons: Peter Horsman, Hans Hofman

Thursday, June 8, 2000

Location: Novotel Hotel, The Hague

International Seminar: Understanding and Preserving Reliable and Authentic Recorded Information in a Digital World -- Focus on Metadata
9.30 – 17.00

Chair:
Eric Ketelaar, University of Amsterdam, Netherlands

Speakers:
Margaret Hedstrom, University of Michigan, USA
Adrian Cunningham, National Archives of Australia
Carl Lagoze, Cornell University, USA
Wendy Duff, University of Toronto, Canada
Sue McKemmish, Monash University, Australia

Participants

Adrian Cunningham, Director, Recordkeeping and Descriptive Standards, National Archives of Australia

Gabriel David, Auxiliary Professor, Department of Electrical and Computer Engineering, Faculty of Engineering, Porto University, Portugal

Wendy Duff, Assistant Professor, Faculty of Information Studies, University of Toronto, Canada

Anne Gilliland-Swetland, Assistant Professor, Graduate School of Education and Information Studies, University of California, Los Angeles, USA

Margaret Hedstrom, Associate Professor, School of Information, University of Michigan, USA

Peter Hirtle, Co-Director, Cornell Institute for Digital Collections, Cornell University, USA

Hans Hofman, Senior Consultant, Ministry of the Interior, The Netherlands

Peter Horsman, Senior Consultant, Archief School – The Netherlands Institute for Archival Education and Research

Ingmar Koch, Student, Archives School – The Netherlands Institute for Archival Education and Research

Carl Lagoze, Digital Library Scientist, University Library and Department of Computer Science, Cornell University, USA

Heather MacNeil, Assistant Professor, School of Library, Archival and Information Studies, University of British Columbia, Canada

Sue McKemmish, Associate Professor, School of Information Management and Systems, Monash University, Australia

Angelika Menne-Haritz, Director, Archivschule, University of Marburg, Germany

Barbara Reed, Principal Consultant and Director, Recordkeeping Systems Pty. Ltd.

Christina Ribeiro, Auxiliary Professor, Department of Electrical Engineering and Computing, Faculty of Engineering, Porto University, Portugal

Meg Sweet, PRO Catalogue Manager and A2A Programme Manager, Public Record Office, United Kingdom

Titia van der Werf, Senior Project Manager, Library Research Department, National Library of The Netherlands

David A. Wallace, Assistant Professor, School of Information, University of Michigan, USA

Nigel Ward, Senior Research Scientist, Distributed Systems Technology Centre, Australia

Workshop Proceedings

[A note on form: The following narrative has been written to capture the flavor and nature of the workshop's discussions. Attributions are not made to individual participants, except for individual session "resource persons" responsible for initiating and framing particular discussions, or when a graphic/illustration was offered for consideration. The author takes sole responsibility for any mis-renderings of the workshop discussions.]

MONDAY, JUNE 5, 2000

FIRST WORKING SESSION

Introduction to the workshop
Discussion on objectives, perspectives and expectations
Resource Persons: Hans Hofman, Peter Horsman

Peter Horsman opened the proceedings. He acknowledged the support of the Archief School in hosting this international meeting on recordkeeping metadata. This was followed by general introductions from each participant. (See Participants list)

Hans Hofman then introduced the expectations and objectives for the workshop. These were developed initially by Hofman and then added to by participants in advance of the workshop.

The expectations included examinations of:

- Metadata versus archival description. Or, what are we describing with metadata?
- The interrelationship between different disciplines dealing with metadata.
- The scope of recordkeeping metadata.
- How to assess the existing metadata standards or initiatives?
- Interoperability of metadata schema.
- Implementation of metadata standards or schema

The objectives for the conference included determining how to:

- Position recordkeeping metadata initiatives with other disciplines and communities.
- Identify metadata issues and concepts that the recordkeeping community shares with other communities.

- Articulate common research questions to come up with a research agenda to investigate the applicability specific recordkeeping standards, such as the Australian Recordkeeping Metadata Schema.

- Identify metadata requirements in context of business processes, recordkeeping, and culture.

- Explore possibility of developing a common infrastructure of research.

Participants shared their perspectives as to what encompasses "recordkeeping metadata," how it relates to and is distinct from understandings of "metadata" in general and understandings of "metadata" in other communities. Does metadata represent a new term or a new concept to the recordkeeping community? Is it something beyond traditional archival description? It was recognized that the metadata concept extends beyond recordkeeping needs and that other communities employ the term for their own purposes. It was suggested that there was a need to more clearly and systematically examine how other communities use the term and how the concept of recordkeeping metadata could be made meaningful to them. How do archivists/records managers/recordkeepers translate their meaning(s) of metadata for other communities and also operate within their evolving referential frameworks? How well do discovery and usage metadata relate to the interests of other communities?

One participant pointed out that in the library world, references to electronic texts now attempt to associate more metadata with an object than has been done with traditional library cataloguing. Metadata for discovery, technical specifications, preservation management, and other areas are being considered, as are issues of inheritance and aggregation (e.g., parts of a book versus the entire work). One participant underscored that metadata could be understood from the level of an individual document/object as well as for aggregates of documents/objects. Understanding and articulating inheritance relationships between individual items and aggregates was seen as central to recordkeeping control and parallel to fundamental archival descriptive principles.

One participant pointed out that one characteristic feature of recordkeeping is the issue of time and its passage and that metadata should be seen as something that accrues to a document/record/object throughout its lifecycle. It was argued that *how* the record travels through time really is a critical aspect of its metadata from the recordkeeping perspective. It provides for richer and more informative documentation that does metadata associated solely with a record at creation.

As the session developed, it was pointed out that was confusion between the participants on the relationship between "recordkeeping metadata" and "archival description," and even on how archivists applied the term "metadata" It was generally acknowledged that there were differences between the participants over what they included within "recordkeeping metadata" and "archival description." In spite of this, there was a strong sense that there was no immediate need to arrive at a consensus position as to what exactly recordkeeping metadata meant. What was more important during this opening working session was to use the concepts of metadata and recordkeeping metadata as a platform for communication. What was agreed upon was that there was a difference in metadata's meaning both between professional communities and within the recordkeeping community.

BREAK
After the break, Hofman offered the following model of an idealized recordkeeping regime/function in order to orient discussion specifically on recordkeeping metadata.

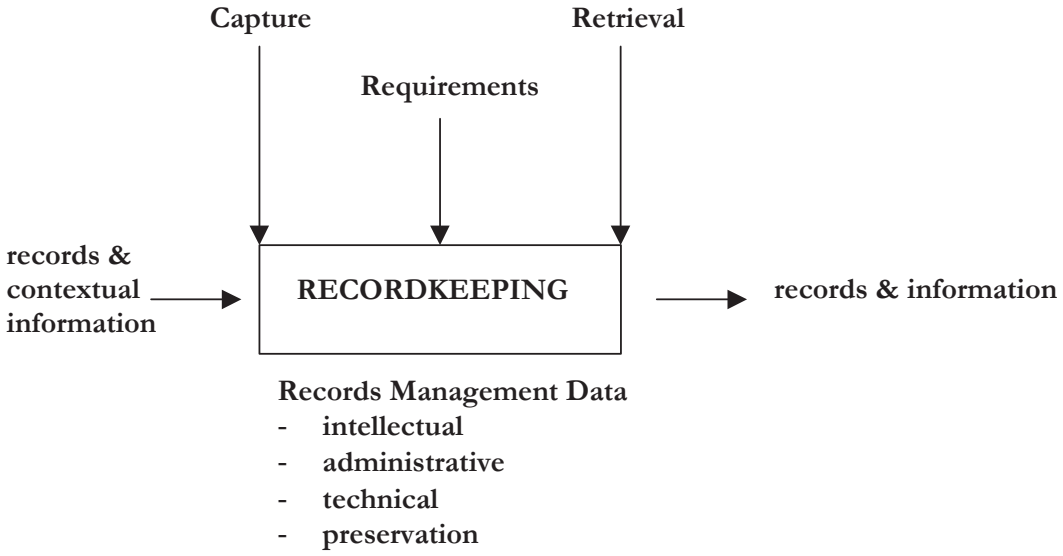

Figure 1. Recordkeeping Regime/Function (H. Hoffman)

This model could be viewed from the perspective of controlling records or from the perspective of recordkeeping processes and functions. The participants appreciated the utility of the model. Discussion considered how this model could be used to explore the scope of recordkeeping metadata: how it gets populated and what its outputs might be. Does recordkeeping metadata concern itself with formal archival description? Does it also include the idea of "warrant" (statements from laws, regulations, case law, information technology standards, auditing standards and best practices that either mandate or suggest particular recordkeeping behaviors)? Developing a shared frame of reference to discuss these issues proved challenging.

It was recognized that the concept of recordkeeping metadata could be fruitfully examined on many levels. Metadata will be need for the records while they are in current usage as well as for when they are retired. Metadata is required for records in creation and use, for archival control, and for preservation. And since electronic records will live in different systems (for example, they can be fixed in one environment and manipulated in another environment), metadata will also be needed for long term control and authentication. It will be important to parse the metadata accruing to records across their existence -- e.g., to be able to distinguish metadata associated with the original transaction that created the record from metadata for long-term archival control. Not doing so opens up the risk of diluting or losing attributes of authenticity, such as when the record becomes fixed as evidence of a particular transaction. It was additionally noted that much of the metadata that was implicit in the active environment was made explicit via post accessioning archival description – such as provenance and scope and content. It was suggested that some of this formerly implicit data needed to be made more explicit during the active stage in a world of distributed networks.

While authenticity and fixity are critical recordkeeping concerns, it was pointed out their meanings were evolving in the networked environment. Authentic documents can be dynamically assembled "on the fly" for viewing, or generated in during solely machine to machine e-commerce transactions. It was also noted that while the recordkeeping profession has spent great of effort on modeling how records should be created and managed, it was unclear how well these reflected the recordkeeping realities and practices of networked organizations and their actors. Within networked organizations, recordkeeping professionals are finding that the stable recordkeeping practices they had for paper are just not holding up. Several participants expressed that there is a pressing need to adapt professional recordkeeping practices to meet the challenges of this emerging environment.

BREAK

SECOND WORKING SESSION

Recordkeeping metadata within a broader metadata framework
What makes recordkeeping metadata different
Objectives of recordkeeping metadata
Resource person: Barbara Reed

During this session, participants developed a mutually-agreed upon definition of "recordkeeping metadata":

> **"Structured or semi-structured* information which enables the creation, management, and use of records through time and within and across domains in which they are created. Recordkeeping metadata can be used to identify, authenticate, and contextualize records; and the people, processes and systems that create, manage, and maintain and use them."**

* Structured metadata, such as the Dublin Core, provides a fixed scheme for organization. Semi-structured metadata, on the other hand, does not require a fixed structure. XML is an example of semi-structured metadata. It is an extensible set of tags that can accommodate unique configurations.

This definition was derived from a reworking of the definition for recordkeeping metadata created by State Records, New South Wales, Australia: "information which facilitates the management, accessibility and meaning of records through time by identifying, authenticating and contextualising records and the people, processes and systems that create and keep them."

While developing and dissecting the definition, participants generally agreed that it needed to be broad and robust enough to encompass both the active recordkeeping environment and traditional archival description (e.g., recognizing a finding aid as a particular type of value-added container of recordkeeping metadata). A question remained though. Could this definition be used to define, relate, and distinguish the recordkeeping domain from other disciplinary views on metadata? How do emphases on context, time, and inheritance distinguish recordkeeping metadata from other forms of metadata?

One non-archivist participant noted that by considering attributes of context and change

across time and domains, archivists were "light years" ahead of other communities' metadata orientations.

There was concern, though, over how this definition did/did not fit into non-recordkeeping perspectives on metadata and how it might be translated to other information management professions. One computer scientist participant suggested that other communities are likely to be unclear over what exactly is meant by "records" and recordkeeping "processes." Another noted though that many of the ideas expressed in the definition, such as processes, were relevant to information systems designers. The problem becomes modeling processes in such a manner that they are apparent and understandable to later users. A third computer scientist participant pointed out for information systems modeling it is becoming increasingly difficult to determine where the document is and what the archival object is and that systems are becoming more complex and implicit in this regard. And another participant contributed that metadata is being defined along a broad range of contexts and scales: for particular information formats; for usage (resource discovery and retrieval); and, for specific functions and processes (e.g., preservation).

Discussion next turned to the role of "forensic" metadata. In the absence of explicitly captured and managed "recordkeeping metadata" do existing systems create metadata which could be exploited by the recordkeeping community to lend attributes of "recordness" to information that was not originally viewed or managed that way? Also, How can recordkeepers use extant metadata to re-purpose records beyond the domain they were originally generated within?

Given these challenges, one participant asked that the objectives for this definition, as well as for the workshop itself, be more clearly articulated. Is it to develop mechanisms and incentives that encourage organizations to implement recordkeeping systems to manage their records? Is it to develop a model for recordkeeping to ensure that archives will be able to ingest records from organizational systems? Is it to develop a means to communicate with other metadata communities? While these are not mutually exclusive goals, it was felt that there was a need to be clearer on these points in order to focus and sharpen discussion.

BREAK

After the break, the session's resource person, Barbara Reed, noted that there remained a number of unresolved issues, especially in regard to the types of systems the participants were advocating. Were these systems for organizational purposes (primary usage) and/or for archival purposes (secondary usage)? Also unresolved was a clear sense as to what exactly was included within and excluded from notions of recordkeeping metadata.

As a means of clarifying these issues, Reed offered language on records management principles and scope from the International Standards Organization's draft international standard for records management (ISO/TC46/SC11):

Draft International Standard for Records Management (ISO/TC46/SC 11), November 10, 1999

4.3 Records management general principles

Records are created, received and maintained in the conduct of business activities. To support the continuing conduct of business, satisfy applicable legal requirements, and provide necessary accountability, organizations should create and maintain authentic, reliable and usable records, and protect the integrity of those records. Documents which are captured in records systems provide evidence and information about business activities. To do this, organizations should institute and carry out a comprehensive records management program which includes:

- determining what records should be created, what information needs to be included in the records, and what level of accuracy is required;
- deciding in what form and structure records should be created and captured;
- determining requirements for retrieving and using records and how long they need to be kept to satisfy those requirements;
- deciding how to organise records so as to support requirements for use;
- ensuring that records are created and maintained in accordance with these requirements;
- preserving the records and making them accessible over time, in order to meet business and community requirements;
- complying with legal and regulatory requirements, applicable standards and organizational policy; and
- ensuring that records are retained for as long as required.

Rules for creating and capturing records should be incorporated into the procedures governing all business processes for which there is a requirement for evidence of that activity.

The regulation of records management policies and procedures ensures that appropriate protection is given to all records, and that evidence and information contained in records can be more efficiently and effectively retrieved, using standard forms of identification and retrieval procedures.

4.4 Scope of records management

Records management includes:

- both records professionals/managers and records users;
- set of authorized policies, assigned responsibilities, delegations of authority;
- procedures, user guidelines and other documents which are used to promulgate the records management policies and to implement strategies and training;

- specialised records systems used to manage the records;
- software and other data files;
- hardware and other equipment, stationery and other supplies; and
- the records themselves.

One participant offered that many communities would be interested in the contents and requirements of the draft standard's language, even if they are not formally or consciously oriented towards "records."

One participant found the draft standard's language narrow in two respects. First, its sole focus was on organizational records. What about personal records? In addition, the emphasis on accountability ignored emerging tools and concepts such as data mining, knowledge management, and organizational learning. Finally, the draft was seen as providing limited value for establishing requirements to manage records across time and across domains (where records are used beyond their creating context and for purposes other than which they were created). Another participant was uncomfortable with the draft's position that recordkeepers should be involved in the creation of records (such as determining what records should be created and what they should consist of). And yet another participant felt that the question over what exactly constitutes a record (in reference to its content and associated metadata) remained unresolved. The draft standard defines records as: **"**Documents created, received, and maintained as evidence and information by an agency, organization, or person, in pursuance of legal obligations or in the transaction of business." In reference to the scope of recordkeeping metadata, the draft standard offers that for a record to accurately reflect what was communicated, decided, or done, it needed to contain the following metadata along with the record's actual informational content:

- the structure of a record, that is , its physical and logical format and the relationships between the data elements comprising the record, should remain physically or logically intact;
- the context in which the record was created, received and used within business should be apparent in the record (including the business process of which the transaction is part, and the participants in the transaction); and
- the links between documents, held separately but combining to make up a record, should be present.

Looping back to the definition for recordkeeping metadata developed by the group, it was generally agreed that it was a definition participants were comfortable with. One participant pointed out that its emphasis on context was consistent with the "archival bond" concept which was also focused on the processes surrounding records.

BREAK

THIRD WORKING SESSION

Recordkeeping metadata, a functional perspective
Recordkeeping metadata and recordkeeping functions
Resource person: Anne Gilliland-Swetland

During this session, participants broke up into two groups. One examined metadata issues in the "continuum" model of recordkeeping. The other looked at metadata through the lens of the "lifecycle" model.

The continuum model recently emerged from Australia as an "integrated regime of management processes for the whole of the records existence." Here records are not seen as "passive objects to be described retrospectively, but as agents of action, active participants in business processes and technologies." In the continuum, appraisal and description begins at or even before records are created and is continually generated throughout the record's existence. The lifecycle model concept emerged in the United States in the 1940s. It prescribes a stricter delineation of roles for records manager (active and semi-active records) and archivists (inactive records). Within it, appraisal and description are viewed solely as the realms of the archivist. Here, records are described only after they have been accessioned by the archives. The main distinction between the continuum and the lifecycle lies in the area of roles and responsibilities – for the creator, records manager, and archivist. The continuum sees involvement by all three entities during a record's primary existence within its creating organization. The lifecycle, on the other hand, holds the creator and records manager responsible for the record up to the point where it is transferred to an archival body, when it crosses the "archival threshold." At that point, responsibility for preserving and documenting records shifts to the archivist.

Continuum group

This session's resource person, Anne Gilliland-Swetland, offered a series of questions to frame the discussion. What are recordkeeping functions? What are the roles of records? Up to this point, discussion has revolved on examining people, processes, and systems, as well as accuracy, reliability, and authenticity. How do we then move on to identifying the types of metadata that support recordkeeping functions (such as during creation, capture, maintenance, use, preservation, and description)? Who will be responsible for each type of metadata identified? How much interoperability is needed between recordkeeping metadata and metadata deployed by metadata communities? How do we extend and relate recordkeeping needs to other metadata communities (such as convincing systems designers to build recordkeeping metadata into new systems).

This group assessed the records continuum thinking advocated by Australia and the SPIRT (Strategic Partnership with Industry – Research & Training) project based out of Monash University, Australia. SPIRT used the records continuum as a frame of reference for its research efforts to develop a draft Australian Recordkeeping Metadata Schema (RKMS). The RKMS provides a:

- "standardized set of structured recordkeeping metadata elements;
- framework for developing and specifying recordkeeping metadata standards;

- framework for reading or mapping metadata sets in ways which can enable their semantic interoperability by establishing equivalencies and correspondences that can provide the basis for semi-automated translation between metadata schemas."[1]

SPIRT's conceptual framework concerns itself with four main classes of entities: business, agents (people), records, and business-recordkeeping. The following graphic provides a high level view of this model:[2]

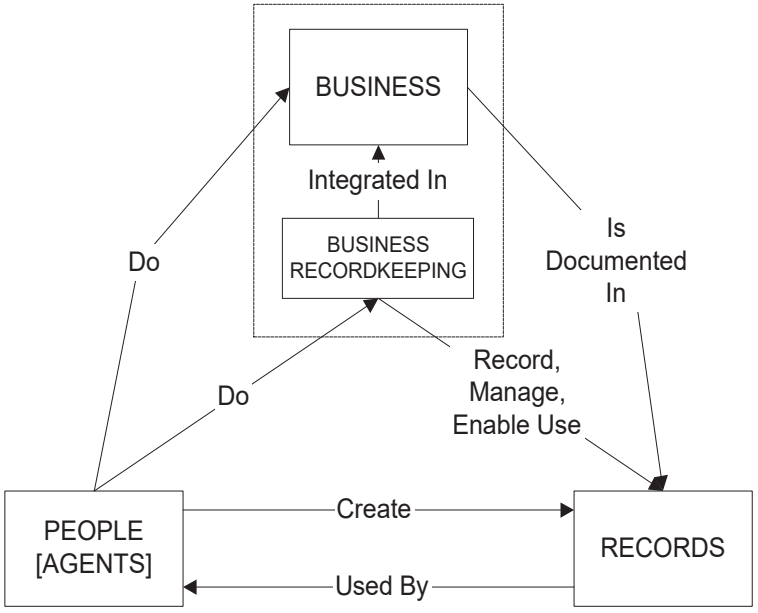

Figure 2. Conceptual Model for Recordkeeping
(S. McKemmish, G. Acland, N. Ward, B. Reed)

At the broader societal level, the RKMS seeks to identify socio-legal requirements, expectations and opportunities for recordkeeping. At the organizational level it provides for a functional analysis to identify context specific recordkeeping requirements that can be developed in to an organizational recordkeeping regime. Ideally, the system can be audited for internal as well as external purposes.

Based on this examination of the SPIRT work, participants identified criteria for framing a discussion on recordkeeping metadata.

- **Appraisal**
  - identifying recordkeeping requirements
  - determining what should be created

- providing for disposal and retention

- **Control**
  - creation and registration
  - classification
  - arrangement
  - description
  - authentication
  - metadata management
- **Preservation**
  - migration
  - refreshing
  - storage
- **Retrieval**
  - rendering
  - presentation
  - representation
- **Access**
  - terms
  - conditions for use
  - permissions
- **Use**
  - users
  - use history

- **Export / Transfer / Disposal**

Any modeling of these criteria would require a feedback loop. This would be used to audit the system in action and also to monitor and evaluate whether or not it was satisfying its requirements. Recordkeeping metadata would be created by and about each of these functions and activities as they occurred. Some of this metadata would be generated automatically by the host system, some created by the record creator(s), and some created by recordkeeping professionals. Some of the system and record creator generated metadata could likely serve recordkeeping needs even though it was not specifically created to do so (akin to the discussion on "forensic" metadata discussed earlier in the day).

In regards to expectation for user-created metadata, it was argued that what is not automatically generated by the system needed to be "thin" enough from a users input perspective so as not to become a barrier to capture. This belief was based on earlier writings on recordkeeping metadata which recognized that the sheer physical volume of records being generated in the computerized environment made the physical capture/creation of recordkeeping metadata beyond human capacities.

The next proposed step for dissecting these criteria would be determining how and by whom the recordkeeping metadata associated with each area elements would be captured.

A second graphic for evaluating recordkeeping metadata within this framework was offered by Hofman:

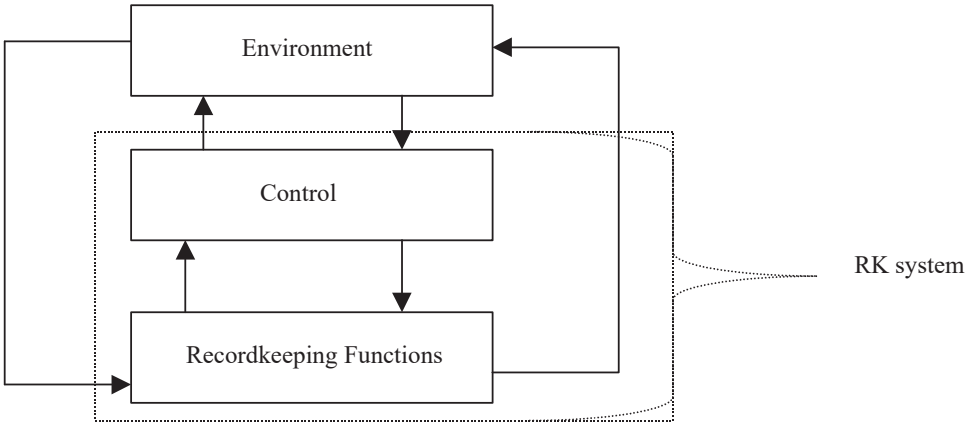

Figure 3. Conceptual Model for Recordkeeping (H. Hofman)

Finally, the continuum group examined the draft ISO international records management standard's "records management operations" to see how it might speak to recordkeeping metadata issues:

8. Records management operations
    8.1 Determining documents to be captured into a records system
    8.2 Determining how long records are required to be kept
        8.2.1 Decisions about records
    8.3 Records Capture
    8.4 Registration
    8.5 Classification
        8.5.1 Classification of business activities
        8.5.2 Classification schemes
        8.5.3 Vocabulary controls
        8.5.4 Indexing
        8.5.5 Allocation of numbers and codes
    8.6 Storage and handling
    8.7 Access and Retrieval
    8.8 Tracking
        8.8.1 General
        8.8.2 Action tracking
        8.8.3 Physical tracking systems
    8.9 Applying disposition authorities
    8.10 Documenting records management processes

It was the sense of the continuum-subgroup that the ISO draft standard was not robust enough to cover the scope implied by the continuum model and the criteria develop above. The standard was seen less integrative than the continuum model and too narrow to embrace the scope of recordkeeping metadata definition developed by the participants earlier in the day.

Lifecycle group

The lifecycle group focused on the components of an "archival framework" and how records move across time and domains throughout their existence.

Within the "archival framework," records are created, maintained, and preserved within the active environment by their creating bodies. Archival control during the active phase would include: creating a classification scheme; developing retention schedules; and, defining access principles. At

some point the records would become inactive and cross the "archival threshold" where they are then arranged, described, and made accessible to different populations of users.

The lifecycle focuses on how records move across time and domains. As a means of articulating and graphically representing this process, the lifecycle group developing the following illustration:

	Time 1	**Time 2**
Domain A	**Creator (active records)**	**Creator (active / inactive records)**
Domain B	**Other users (active records)**	**Archives / other users (inactive records)**

Figure 4. Time & Domain matrix (M. Hedstrom)

The general impression was that the recordkeeping profession has spent considerable thought and effort on the Domain A/Time 1 (vis a vis records management) and the Domain B/Time 2(vis a vis archives) quadrants. Less well understood are the Domain B/Time 1 and the Domain A/Time 2 quadrants. There was a sense that the recordkeeping profession needed to expend more effort on understanding processes and activities across all of these quadrants.

Interestingly, the lifecycle group came to the conclusion that, contrary to much of the debate in the professions between the lifecycle and the continuum, there was no significant incompatibilities between these two approaches. This represented an erosion of the view that sees strict distinctions between the lifecycle and continuum. Instead of focusing on how to draw distinctions between these two outlooks, participants thought that the recordkeeping profession should instead concentrate on understanding the four quadrants and how responsibility for recordkeeping shifts across time and domains.

Tuesday, June 6, 2000

Fourth Working Session
Emerging standards
Need of standards
Resource person: Wendy Duff

Resource person Wendy Duff opened the second morning by re-visiting points that arose several times over the preceding day: How can recordkeeping professionals exploit existing metadata standards outside of the profession and how can we represent our metadata approaches to other domains? How and when can metadata crosswalks – the mapping of data elements across different metadata standards – be exploited? Since there is not always a one-to-one relationship between standards, what levels of integration and granularity can be accommodated/tolerated in these crosswalks? Duff suggested starting first at examining other standards to see how they may/may not fit with the criteria developed the day before:

- **Appraisal**
  - identifying recordkeeping requirements
  - determining what should be created
  - providing for disposal and retention
- **Control**
  - creation and registration
  - classification
  - arrangement
  - description
  - authentication
  - metadata management
- **Preservation**
  - migration
  - refreshing
  - storage

- **Retrieval**
  - rendering
  - presentation
  - representation
- **Access**
  - terms
  - conditions for use
  - permissions
- **Use**
  - users
  - use history
- **Export / Transfer / Disposal**

One Australian participant pointed out that the Australian records management standard – AS 4390[3] -- represented a dramatic shift in thinking about the concept of appraisal in that country. Two key features of this shift is the process of determining what records should be created and in what form; and, appraising functions over appraising records. A Dutch attendee pointed to the Dutch PIVOT model where even before they are created, records have been appraised and the basis for their description has been developed. And a Canadian noted that the National Archives of Canada has been working more aggressively to link its appraisal data with its descriptive data.

Related to the issue of functional analysis and appraisal, one participant pointed out that records creating organizations are also aggressively pursuing functional analysis for their own strategic management purposes. It was suggested that recordkeeping initiatives could attempt to develop links to these. Useful metadata will almost certainly come from the active environment as well being generated by traditional archival description. And over the long term, useful metadata will derive from the actual use(s) of the records themselves.

Linking to other disciplines own metadata standards was seen as essential. It was recognized that there was a pressing need to examine and use taxonomies and classifications from other domains and professions. Taxonomies and classifications such as the "Nursing Intervention System" which provides a classification system of 336 interventions further sub-divided into 6 domains and 26 classes, each comprised of a label, a definition, a set of activities, and background readings.[4] Such tools, however, tend to be very domain specific and recordkeeping professionals need to be aware of

the relevant standards and classifications from those domains whose records they oversee. How can the profession take advantage of the classification schemes used in specific domain when they ingest records? How can the profession "inherit" the classification schemes already in place and used within organizations instead of trying to redevelop aspects of them later for archival descriptive purposes?

One participant cautioned against relying too much on the creator. The archivist provides a unique perspective on why the records were created and how they were used. The creator is not necessarily interested in this perspective. This participant argued that what was also needed was strategies to reconstruct the original character and shape of the collections accessioned. Another participant related that the National Archives of Switzerland constructs a classification scheme for each agency, which is then used by the agency to manage their own records. When the archives then accessions the agency's records they add additional metadata at the file and series level.

As all participants agreed taxonomy and classification systems from other domains were relevant to the recordkeeping endeavor, Duff pressed the group to address more specifically what actions the recordkeeping profession should take to exploit this possible advantage.

One participant noted that what was needed was a model/framework to import information from other communities and their taxonomies into a system that helped the profession satisfy recordkeeping concerns. The kind of metadata that archivists will need to/be able to create will be contingent upon the metadata captured by organizations for their own records. Following this thread, one participant highlighted that many communities are using XML (Extended Markup Language) to create and develop document type definitions (DTDs) to exchange information in standard ways. This presents one concrete avenue for exploiting metadata within other domains. Another participant highlighted the work being done by the U.S. National Archives (NARA) and the San Diego Supercomputer Center (SDSC) to use DTDs as part of a strategy to automatically ingest data objects (such as word processing documents, electronic mail messages, databases, etc.) and metadata about them into an archive of persistent digital objects.[5]

Summarizing the session's discussion to that point, a computer scientist participant noted that they fell along three fronts:

- Do other domains use metadata models that archivists can also exploit?
- Do other domains have vocabulary metadata schemas that archivists might also use?
- How can metadata be shared across domains?

The above discussions helped the participants give shape to their commentaries. Could an archives become a repository of metadata schemas that simultaneously serve archival and creating organization purposes? What are the types of recordkeeping metadata needed by recordkeeping professionals that is not created elsewhere by others? If we want the creator to capture certain types of metadata, what are the incentives for getting them to do so? Just how useful is the metadata developed by traditional archival descriptive practice? Do users really want metadata created by archivists and do they use it?

It was pointed out that archival systems manage and add value to records systems which have been physically removed from their point of origin. As such, archivists are required to take custody of systems as they exist and then work to supplement them with archivally-generated metadata to make them accessible to users. One participant likened this movement of records and information about them from the creator through the archivist to the user as a "knowledge transfer" process. Archivists are in essence intermediaries between the creators and the users whose role it is to transfer knowledge between the two. Such knowledge transfer requires that the profession conduct more systematic studies on users of archives as well as strive to make some of the tacit aspects of this process more explicit. Just how well do archival metadata tools translate to our users?

In regards to incentives for getting records creators to add archivally relevant metadata, it was the impression of the participants that creators and users will not add/provide metadata unless they see a clear purpose for it. As a general rule, archivists cannot expect creators and user to add archivally relevant metadata for future purposes. Within many governments, for example, there exists formalized recordkeeping rules and regulations which dictate specific recordkeeping behavior. However, most of these frameworks provide no sanctions for violation and are seen by many as paper tigers. One strategy to confront this reality is to help agencies meet their own needs for recordkeeping. One participant pointed to the National Archives of Australia (NAA) Recordkeeping Metadata Standard for Commonwealth Agencies,[6] which builds upon the work of the AS4390 Standard and the SPIRT project. This standard – which consists of 20 elements, 8 of which are mandatory -- was developed to meet a real need within government agencies who were looking for advice on managing the explosion of computer-generated. Staff at the NAA are now targeting systems designers to develop systems that capture much of the requisite metadata automatically. Future steps include determining how the archives can ingest metadata from these systems into archival control systems. In tandem with this initiative, the NAA launched a government-wide campaign on the value of good recordkeeping; focusing on issues such as improving corporate memory reducing litigation exposure.

It was pointed out that the degree to which metadata capture could be mandated relied in part on the degree of control possible in different communities. Some domains are more rule-based and enforcement oriented than others. Within these there is a greater opportunity to require and establish metadata capture routines. Less rule-based domains are likely to require more automatic metadata capture as users have lower incentives for adding metadata. It was unclear how metadata could be captured from these relatively uncontrolled communities.

One participant offered that given the volume of electronic records currently being created by most organizations, tools such as data mining, metadata mining, artificial intelligence, text summarization, and the like are the profession's only hope. The notion that humans can manage organizational information systems via use of standards mandating particular metadata creation just will not scale. It was believed by this participant that standards will not be a large part of the solution sought by the recordkeeping profession.

BREAK

After the break, resource person Wendy Duff directed that participants spend the rest of the session's time focusing in on a few specific topics instead of skimming over many areas in a superficial manner. She suggested that the group examine the following: What metadata do

recordkeeping professionals need that no one else is creating? What is the purpose of recordkeeping metadata?

The group re-examined the definition for "recordkeeping metadata" developed during the preceding day's discussion:

- Structured or semi-structured information which enables the creation, management, and use of records through time and within and across domains in which they are created. Recordkeeping metadata can be used to identify, authenticate, and contextualize records; and the people, processes and systems that create, manage, and maintain and use them.

One computer scientist participant pointed out that one of the focuses he sees in the above definition is the emphasis on processes and events. Other metadata communities modeling processes and events include the INDECS and HARMONY projects.

INDECS (Interoperability of Data in E-Commerce Systems) is an international collaborative project that is developing a metadata framework to support electronic commerce over computer networks. Tracking transactions and events over time and ensuring authenticity of resources are two features of the project.[7]

The HARMONY project is examining the relationships (including temporal, spatial, structural and semantic) between resources composed of text, image, audio and video objects. A basic tenet of the project is that descriptions of these objects requires an accounting of these relationships. Objective for HARMONY include:

- Collaborate with metadata communities to develop and refine developing metadata standards that describe multimedia components.

- Investigate a conceptual model for interoperability among community-specific metadata vocabularies. Such a conceptual model should be able to represent the complex structural and semantic relationships in multimedia resources.

- Investigate mechanisms for expressing such a conceptual model, including technologies currently under development in the W3C (XML, RDF, and their associate schema mechanisms).

- Develop mechanisms to map between community specific vocabularies using such a conceptual model.[8]

It was asked whether the HARMONY project might serve as a model for mapping metadata standards in a more complex manner. Instead of performing a straight field to field match between standards, the project seeks to recognize the relationships between content, events, and descriptions. This model is currently at the conceptual stage (see figure 5). HARMONY investigators are working with the Consortium for the Interchange of Museum Information (CIMI) to test the project's premises and objectives within a real application. A HARMONY participant underscored that this type of crosswalk metadata mapping for interoperability is very difficult to perform. Interoperability between metadata sets requires clarity of the level of metadata being matched across standards

(vocabulary, process, field labels, etc.) Another participant pointed out that development of any metadata standard entails the development of a high level model broken down into entities, which themselves are broken down into elements. It was underscored that recordkeeping professionals need to appreciate these subtleties, process and intellectual challenges.

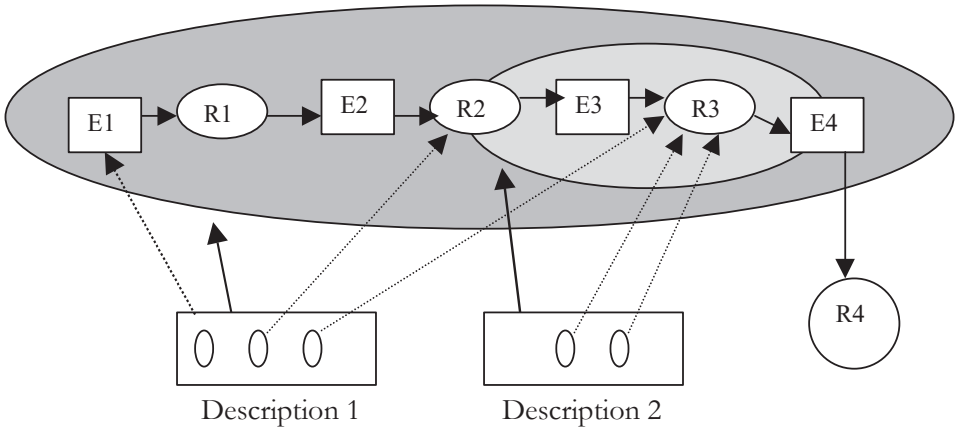

Figure 5. Content, Events, and Descriptions
Key: E = Event; R = Record. (Carl Lagoze)

Participants then re-worked on the Recordkeeping Regime/Function figure examined over the preceding day's discussions to determine where different metadata standards might fit within it (in red font) (See figure 6)

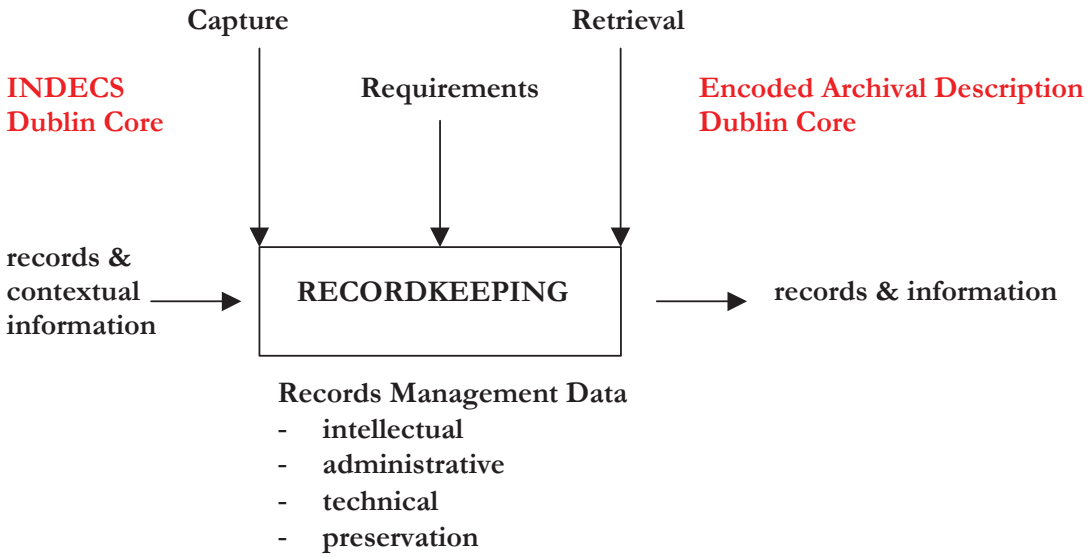

Figure 6. Figure 1 revisited - Fitting in metadata standards

Finally, participants visited the detailed version of the SPIRT recordkeeping metadata model (Figure 7):

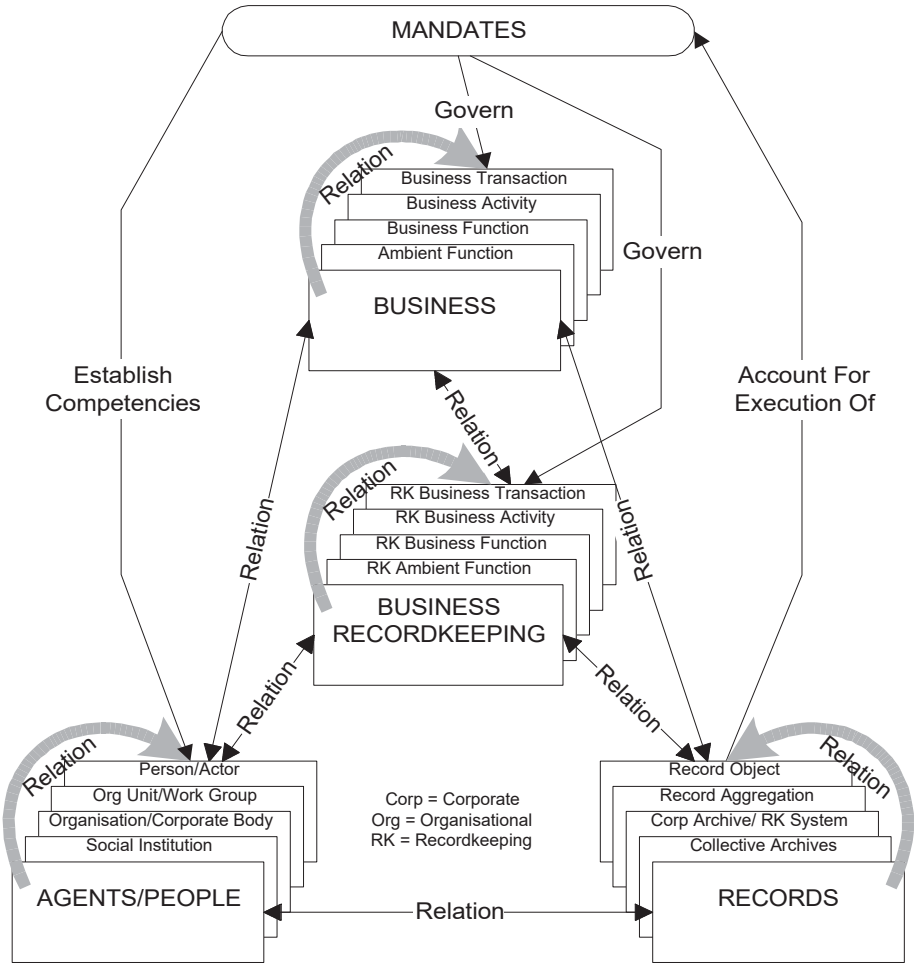

Figure 7. Coverage of Recordkeeping Metadata (S. McKemmish, B. Reed, N. Ward)

It was suggested that in keeping with the focus on events and relationships, the SPIRT model needed more explicitly and concretely articulate what is happening within the relations drawn into the model.

Fifth Working session

Connections between the themes
The themes revisited
Adequacy and applicability
Discovering white spots
Resource person: Sue McKemmish

Resource person Sue McKemmish opened the session by revisiting the some of the more fundamental themes raised over the two preceding days:

- Positioning recordkeeping metadata initiatives within the context and needs of other disciplines (What does the recordkeeping metadata community share with/offer to other communities?

- Applicability of existing recordkeeping metadata standards, such as SPIRT, to other countries

- Understanding metadata requirements within the context of business processes and cultures

- Establishing a common research agenda on recordkeeping metadata

- Developing a common infrastructure and understanding for research on recordkeeping metadata

Related to these are issues associated with: determining which metadata that is exclusive to recordkeeping; the role of archivists as intermediaries in a knowledge transfer process; the potential controllability of particular domains in respect to their recordkeeping practices; what roles can prescribed and forensic metadata play; how to trace events within a metadata model; and, assessing the role played by time and its passage.

The main issue that drove discussion forward from this point was the challenge of managing changes to metadata over time.

The issue of managing change to metadata over time was seen as a very complex problem with few existing models to draw on. The recordkeeping profession is generally more time aware than are other professions. Perhaps it is this focus on time and change over time that provides it with a framework for establishing a distinct role for itself in relation to other communities.

Should archivists develop metadata repositories to track changes to metadata schemes over time? What would be the benefits of doing so? How would versions of different schemes be managed? How would this tracking feed into long term preservation efforts (e.g., emulation, migration, etc.)?

Just as languages evolve over time, so does metadata. How does metadata evolve over time? How and when do you snapshot it? Different metadata schemes will require different approaches. For certain schemes it makes no sense to track every change. For others more granular version control may be required. For example, you cannot (and probably would not need to) track the changes to

the Dublin Core given the way it is being used across the globe today. In other cases, it may be important to link different versions of metadata schemes with their time-associated data. More refined linking might provide for an object's authenticity and increase its understandability to users. Tracking metadata over time will prove to be a complex and difficult task for the recordkeeping profession. Working with other communities to make them more time aware of their metadata schemas is one strategy to manage these long-term challenges.

One participant cautioned against the recordkeeping profession from taking on too much responsibility in this area. There will be knowledgeable and motivated individuals and communities who will labor hard to "translate" aged systems and derive meaning from them when it is important for them to do so. By analogy, classical scholars will labor quite hard to understand ancient texts and material culture. Archivists need to be careful not to over translate each time technology changes. However, modulating when action is required presents a significant challenge.

If archivists manage records over time, how much of the original presentation of the document needs to be kept as evidence of its authenticity? The InterPARES (www.interpares.org) research into permanent authentic electronic records is developing a typology of the types and forms of records created in the electronic environment and specifying the types of contextual metadata that needs to be associated with them over time. It is examining those presentation features of an electronic original that enable the document to generate consequences in the operating environment.

Looping back to the issue of roles and responsibilities, one participant pointed out that if long term access becomes an important enough problem for enough different communities then they will work together to resolve it. If it is of interest only to archivists, then archivists should expect little cooperation or help from others. One computer scientist participant pointed out though, that many other communities have enumerated metadata challenges very similar to the posed by this forum – challenges associate with longevity, authenticity, persistence, fixity, etc…. The struggle facing the recordkeeping metadata community is positioning itself in relation to other communities and making connections with them. Options offered by participants included working on digital signature, public key infrastructure, and e-government initiatives.

BREAK

After reconvening, participants focused their efforts on identifying parallel metadata communities and how recordkeeping metadata can speak to them. One computer scientist participant underscored that many, many communities are developing metadata frameworks that could informed by recordkeeping aspects: transactions, functions and processes, change overtime, etc. What has to happen, though, is for the recordkeeping metadata community to step forward and politic its viewpoints within other communities as they are not likely to seek out the recordkeeping community. The recordkeeping community will need to make its presence known by lobbying and making evident the value of its perpective. The Dublin Core may be one forum, though there is tension even within that community over the need to extend the core beyond its 15 elements. One participant agreed to look into conducting a joint workshop with the Dublin Core community.

There was a sense amongst the participants that the recordkeeping metadata community needed to start picking its best targets for influence. Candidates included the: revision of ISO9000 for quality systems, digital preservation, digital libraries, museums, public key infrastructure, resource discovery,

e-commerce, e-government, e-culture, multimedia, entertainment, rights management, information systems auditing, workflow and business process, IEEE (Institute of Electrical and Electronics Engineers) metadata community, Text Encoding Initiative, Consortium for the Interchange of Museum Information, software companies, etc. It was well recognized that there exist many relevant related communities.

In regards to brainstorming research directions, use of automatic mechanisms emerged as one possibility. Assessing the right mix of human intervention and automatic mechanisms such as data mining and web-archiving is one area that could be explored. It was pointed out that the Internet Archive project (www.alexa.com) is very interested in exploring how automatic data extraction could be used to managing its enormous volume of holdings. In addition the National Library of Australia's PANDORA (Preserving and Accessing Networked Documentary Resources in Australia.) project is creating an archives of selected online publications such as electronic journals, organizational sites, government publications, and ephemera. It is exploring preservation approaches to these materials and is establishing an service to provide persistent links for abstracted and indexed items (See: pandora.nla.gov.au/pandora/)

Other research areas noted included: users; recordkeeping in informal work environments; and persistent linking.

In closing, one participant noted that recordkeeping in society is a complex issue of enormous scale. It was the sense of several participants that it would be unrealistic to take on the entire scope of this challenge. What could be done instead is to first identify a broad range of areas as priority areas and move froward from there.

Wednesday, June 7, 2000

Location: Castle of Bergh

SIXTH WORKING SESSION

Research Agenda
What, how, by whom
Cooperation
Towards a network of excellence
Infrastructure for research.
Resource person: Margaret Hedstrom

The resource person for this session, Margaret Hedstrom initiated discussion by re-visiting the National Historical Publications and Records Commission's (NHPRC) 1990 research agenda for electronic records to see how well they resonate a decade after they were issued.

NHPRC's Research Agenda circa 1990

1. **What functions and data are required to manage electronic records? Do data requirements and functions vary for different types of automated applications?**

2. **What are the technological, conceptual and economic implications of capturing data, descriptive information, and contextual information in electronic form from a variety of applications?**

3. **How can software-dependent data objects be retained for future use?**

4. **How can data dictionaries, information resource directory systems, and other metadata systems be used to support electronic records management and archival requirements?**

5. **What archival requirements have been addressed in major systems development projects and why?**

6. **What policies best advance archival concerns for identification, retention, preservation, and research use of electronic records?**

7. **What are elements should be present in electronic records management programs and how should they be evaluated?**

8. **What incentives can contribute to creator and user support for electronic records management?**

9. **What barriers have prevented archivists from developing and implementing electronic records programs?**

10. **What do archivists need to know about electronic records?**

So what has changed over the past decade? Most significantly for recordkeeping is the explosion of networked computing via the Internet and World Wide Web. A decade ago the focus was on applications. Today there is a more focused emphasis on recordkeeping systems. Another change specific to the recordkeeping profession is the increased number of academics working on electronic records issues. Ten years ago there were comparatively fewer full-time recordkeeping academics, and those that did exist tended not to be interested in/well informed about electronic records issues. The recordkeeping profession has also become more sophisticated in regards to research: how to do it; relationships to between theory and practice; and, collaboration. It is clear that resolving these problems will require sustained research by many projects over many years. One or two research projects cannot be expected to develop the "magic bullet" for others.

It was proposed that there were three big questions facing the participants in regards to research:

- What research?
- By whom?
- How?

To jumpstart discussion, Hedstrom offered a recasting of the NHPRC's 1990 research agenda, edited to reflect current recordkeeping metadata research needs:

Re-worked Research Questions (key: adds underlined, deletions ~~striken through~~)

1. **What functions and metadata are required to ~~electronic~~ manage records? Do metadata requirements and functions vary for different types of ~~automated applications~~ recordkeeping systems/functions/domains?**

2. **What are the technological, conceptual and economic implications of capturing metadata, descriptive information, and contextual information in electronic form from a variety of ~~applications~~ recordkeeping system/domains?**

3. **How can ~~software dependent objects~~ metadata be retained for future use?**

4. **How can data dictionaries, information resource directory systems, and other metadata systems be used to support electronic records management and archival requirements?**

5. **What archival requirements have been addressed in major ~~systems~~ metadata development projects and why?**

6. **What policies best advance archival concerns for identification, retention, preservation, and research use of ~~electronic records~~ metadata?**

7. **What are metadata elements should be present in ~~electronic records management programs~~ recordkeeping regimes and how should they be evaluated?**

8. **What incentives can contribute to creator and user support for ~~electronic records~~ metadata management?**

9. **What barriers have prevented archivists from developing and implementing ~~electronic records programs~~ metadata models, schema, and vocabularies?**

10. **What do archivists need to know about ~~electronic records~~ metadata?**

It was pointed out that there are currently many possible metadata initiatives, standards, and technologies available to the profession for evaluation. However, there is no way to predict which will last and which will fade away. For example, in 1990 it was believed that information resource directory systems would become a major tool for the profession, yet this never materialized.

Hedstrom suggested that the reworked 10 questions should not be viewed as "THE" research agenda. Its purpose was to provide a framework for thinking about what research needed to be carried out, who would do it and how it would get done. Participants were encouraged to use the reworked agenda to develop a more focused set of strategies mindful of possible funding streams. It was underscored that effort was best directed towards developing *fundable* research questions, e.g., questions that reflect the priorities and agendas of funders not accustomed to supporting recordkeeping research.

Session resource person, Margaret Hedstrom then asked the assembled to discuss their own research and how it is related to metadata

Adrian Cunningham, Director, Recordkeeping and Descriptive Standards, National Archives of Australia. Currently focused on policy and standards setting at the National Archives of Australia. After having rolled out these policies and standards there is now need to evaluate them to see how they are working. There is also the question of building an infrastructure within the distributed network environment to provide users with the rich metadata covered by the policies and standards. Also work with SPIRT.

Gabriel David, Auxiliary Professor, Department of Electrical and Computer Engineering, Faculty of Engineering, Porto University, Portugal. Building a prototype system for multimedia description that captures content and context. Also examining the recordkeeping implications of information systems.

Wendy Duff, Assistant Professor, Faculty of Information Studies, University of Toronto, Canada. Working with users. More specifically, identifying the tasks scholars perform in archives and how they use archival descriptive tools (metadata tools) to do that. How do they judge authenticity in the electronic environment?

Anne Gilliland-Swetland, Assistant Professor, Graduate School of Education and Information Studies, University of California, Los Angeles, USA. Co-directing the US-InterPARES project which is working with computer scientists at the San Diego Supercomputer Center (SDSC). Also working with Sue McKemmish on metadata registries, and involved in a digital library project that is looking at metadata from a multidisciplinary perspective.

Margaret Hedstrom, Associate Professor, School of Information, University of Michigan, USA. Researching how users use EAD (Encoded Archival Description). Also working on a emulation project that, while not initially explicitly oriented towards metadata, is quickly finding that metadata will play a considerable role in the technical (hardware/software) aspects of emulation.

Peter Hirtle, Co-Director, Cornell Institute for Digital Collections, Cornell University, USA. Focusing on creating digital surrogates and how people use them. How do people work with digitized materials. Also working with XML to determine how much hierarchy has to be evident in

collection Also working on Cornell University's PeopleSoft business process reengineering project to factor in archival requirements.

Hans Hofman, Senior Consultant, Ministry of the Interior, The Netherlands. Working on government online from a recordkeeping point of view. Currently developing a data model for describing agencies and their functions. Also involved in digital preservation where metadata plays a critical role.

Peter Horsman, Senior Consultant, Archief School – The Netherlands Institute for Archival Education and Research. Recordkeeping is one of the school's main areas and metadata is seen as a critical aspect of this area.

Carl Lagoze, Digital Library Scientist, University Library and Department of Computer Science, Cornell University, USA. Lagoze's HARMONY project recognizes that mapping between metadata sets involves more than crosswalks and is collaborating working with the Consortium for the Interchange of Museum Information to explore this further. Also involved in the PRISM project, a U.S. National Science Foundation Digital Library II grant. This project is focused is recognition that digital libraries are working with digital content over which they have not direct physical control over the content. The project is exploring access and preservation policies for content that it outside of direct control. Also working on the Open Archives Initiative to develop interoperability between e-print archives.

Heather MacNeil, Assistant Professor, School of Library, Archival and Information Studies, University of British Columbia, Canada Currently involved in InterPARES which is examining the requirements for authentic electronic records. What are the particular requirements for authenticity? Where are they? How are they captured? How are they managed and preserved intact over time?

Sue McKemmish, Associate Professor, School of Information Management and Systems, Monash University, Australia. Now approaching closure on the SPIRT project and exploring meta-modeling tools (such as RDF, object role modeling, and experimental instantiation). How can the relationships within the SPIRT be modeled in a more formal way? Also examining metadata registries to better describe metadata schemas. Also working on the Titanium Project, supported by Australia's Distributed Systems Technology Centre (DSTC). Titanium is working to enable organizational data access and exchange using XML and metadata.

Barbara Reed, Principal Consultant and Director, Recordkeeping Systems Pty. Ltd. Reed works in applied environments and is developing systems that translate SPIRT's work into implemented systems

Christina Ribeiro, Auxiliary Professor, Department of Electrical Engineering and Computing, Faculty of Engineering, Porto University, Portugal. Currently contributing to the MPEG7 standard for the description for multimedia documents. Also working on recordkeeping requirements for information systems.

Meg Sweet, PRO Catalogue Manager and A2A Programme Manager, Public Record Office, United Kingdom. Work at the PRO is currently being driven by the requirement to provide access to e-

records by 2003. To meet this objective the PRO is now focused developing systems of adequate metadata for retrieval and discovery.

David A. Wallace, Assistant Professor, School of Information, University of Michigan, USA. Co-directing an NHPRC grant examining recordkeeping in informal team-based environments. How can metadata created in these environments be re-used and re-purpose? Also looking at metadata issues associated with email and the U.S. Electronic Freedom of Information Act (e.g., metadata for discovery, retrieval, and request processing.

Nigel Ward, Senior Research Scientist, Distributed Systems Technology Centre, Australia.
Ward highlighted two aspects of his research: work with SPIRT on data modeling; work on a DSTC project on "information ecology" that is just underway. The "information ecology" perspective states that people, groups, information, and technology work together in a changing ecology. We want to build tools to people work within this ecology. Anthropologists and CSCW experts will work on this project. How can CSCW document repositories, which of the heart of CSCW systems, be made robust with metadata repositories that are capable of describing documents, people, and processes.

BREAK

Participants then moved on to consider if it was fruitful to try to tie these projects/interests together. One participant noted that the recordkeeping profession is so small and that one strategy should be to look outside of the profession for collaboration, especially as recordkeeping metadata questions are questions that others are interested in.

Hedstrom then offered the following framework for research:

- Social, Cultural, and Policy domain
- Use domain
- Aspects related to Time dimensions
- Technical issues (Architecture, Schemas, Mapping, Interoperability, etc…)
- Context, Evidence, Documentation

What is the profession not addressing in this framework? Identifying these "white spots" can help refine research priorities.

One participant asserted that this workshop could have any one of three possible outcomes:

- Move to obtain small funding for a white paper of research issues
- Increase collaboration between participants
- Development of major new research proposals

The hosts indicated their hopes for the meeting. First was to simply to get the group assembled. It was hoped that this workshop would serve as the base for follow-up meetings. Key outputs include production of the proceedings and their posting, along with presentations from the International Seminar (Understanding and Preserving Reliable and Authentic Recorded Information in a Digital World -- Focus on Metadata), to the Archief School's website.

Participants then moved on to a discussion of how to make the recordkeeping endeavor resonate with non-archivists and how to leverage that interest into grants and collaborative research. If it is the values associated with context and the crossing of time and organizational/domain boundaries, then how does the profession raise these and make them relevant to other research domains? What are the convincing arguments to potential funders and collaborators that developing, capturing, and using recordkeeping metadata to move objects across time and domains is an interesting problem? Can the profession attach itself to existing metadata communities and leverage collaboration to build its own infrastructure for research and implementation?

It was argued that the recordkeeping profession needs to tell interesting and compelling stories about current recordkeeping problems and values. Is re-purposing contextual metadata for reasons other than for which it was collected one area of interest to other communities? How should the profession draw on high profile recordkeeping cases (such as the Nazi Gold scandal, the Khmer Rouge archives, the tobacco wars in the U.S., among others) to underscore that similar powerful social lessons from more current cases may not be possible given the fragility of electronic records and the non-capture of critical contextual metadata.

Computer scientist participants suggested that the profession advocate recordkeeping metadata to system designers and that they work to convince computer scientists that recordkeeping is an interesting research problem and one that may challenge some of the strategies they have developed for computer-stored information. One direction would be to begin linking to and collaborating with computer scientists who are interested in crossing boundaries and who see recordkeeping metadata issues as interesting research questions. The San Diego Super Computer Center (SDSC) was offered as an example. It has been collaborating with the InterPARES project and the U.S. National Archives to address issues of longevity and ingesting large volumes of electronic documents into an digital archival repository.

CSCW and digital library research were highlighted as other important communities that the profession should speak to on recordkeeping metadata issues such as authenticity, version control, and longevity. As pointed out by one computer scientist participant, the problems raised by the participants during the workshop appear no different than many of the problems facing digital library researchers, and in fact, digital library researchers often work with archival materials in their projects.

Turning to possible funding streams, the following entities were suggested as possible sources: joint U.S. National Science Foundation (NSF)/European Union digital library initiatives (such as with the U.K. Joint Information Systems Committee (JISC)) International Digital Libraries Initiative; the Long Now and Sloan foundations; and the pharmaceutical and entertainment industries. The more general domains of e-commerce and knowledge management were also seen as a possible entry points.

It was underscored that the recordkeeping metadata community not amputate itself from the larger metadata research community working with computerized information and actively work to develop strategies to enter these communities. It was recognized that there was a need to focus both internally within the profession and externally on other communities and to develop linkages between the two.

BREAK

Seventh Working Session
Research Agenda continued
Follow-up
Adjournment
Resource person: Peter Horsman

Resource person, Peter Horsman, offerd the following focus for the workshop's last session:

- Solidifying the research agenda (white paper, compelling stories…)
- Determining what pieces of the research agenda have been addressed by the recordkeeping metadata community, what pieces have beed addressed by other metadata communities, and what pieces have yet to be exmained
- Development of a framework for a common understanding of recordkeeping metadata and it value, both within and external to the recordkeeping metadata community
- Identification of the communities to carry out the agenda out, communiteis both internal to and external to the recordkeeping metadata commuity

A key challenge remains the development of a framework that provides for shared concepts and not necessarily terminological harmonization between disciplinary vocabularies. How can different disciplinary vocabularies for similar concepts be clarified and related to each other?

One idea that resonated with participants was for members of the recordkeeping metadata community to branch to other metadata communities to collect information about their metadata initiatives (and also to bring the recordkeeping metadata message to them). Participants would then report back to the recordkeeping metadata community to provide it with a broader and more sophisticated understanding of other metadata work being conducted across the world and across domains. Participants were encouraged to consider how to "infiltrate" other metadata communities.

Two specific avenues advanced as mechanisms for developing these linkages included joining existing funding streams, such as the NSF's Digital Library Initiative, and presenting recordkeeping metadata research to other communiteis professional conferences. The Victoria Electronic Records Strategy's presentation to the ACM digital library community was highlighted as an example of this second mechanism. Initial efforts in this direction will provide a means for making recordkeeping metadata concerns both vital and overt. And efforts directed to other disciplines can be fed back

into and advocated within the recordkeeping community as well. For example, the recordkeeping metadata community can help bridge the divide between the electronic records and the descriptive standards sub-communities within the broader recordkeeping community.

To provide a focus and identity to the workshop's participants, it was decided to name the group the **Archiving Metadata Forum**

As a means for clarifying the scope and mission of the Forum, one participant offered the following reworking of the Society of American Archivists Working Group on Standards for Archival Description matrix as a reference model for understanding standard:

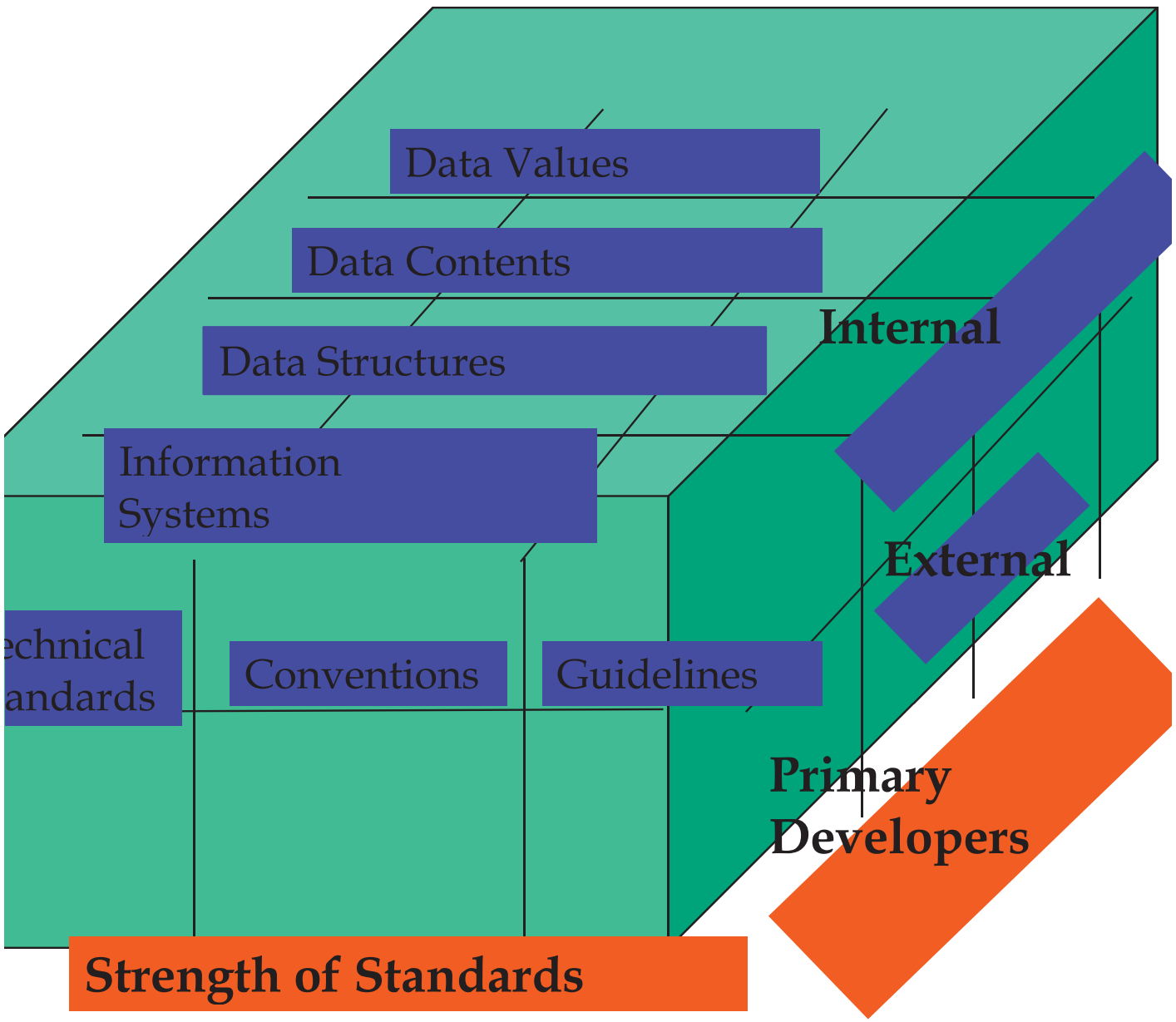
Level of Description
Data Values
Data Contents
Data Structures
Information Systems
echnical andards
Conventions
Guidelines
Internal
External
Primary Developers
Strength of Standards

Figure 8. Reference Model for Assessing Metadata Standards (M. Hedstrom)

The sense of the participants was that this reference model can be made useful to the **Archiving Metadata Forum**, but that its terminology and concepts need to be refined and clarified.

In regards to next steps to push the **Archiving Metadata Forum's** agenda forward, the following were suggested:

- Establish a web presence for the Forum. The Archief School expressed a willingness to maintaining and updating the Forum's web presence. Initial content for the web presence include the workshop proceedings and the presentations from the June 8th International Seminar on Understanding and Preserving Reliable and Authentic Recorded Information in a Digital World. Additional content identified for the web presence include one page descriptions of metadata research being conducted by Forum participants.

- Explore using CSCW tools (such as egroups.com) and establishing a listserv to share information on research and to enable collaboratively authored documents pertientent to the Forum's mission.

- Drafting a companion piece to the proceedings to announce the formation, mission, and objective of the Forum. *D-Lib* Magazine was indicated as a useful outlet for this effort.

- Become involved in the drafting of the ISO Records Management standard

- Meet again. It was suggested that the Public Record Office, U.K, would be very interested in hosting the next meeting.

At this point, the hosts thanked the assembled for providing three days of stimulating discussion and for establishing the **Archiving Metadata Forum**. Participants closed the workshop by asserting their commitment to continue the working within the Forum to promote reccrdkeeping metadata concerns.

ENDNOTES

[1] SPIRT Research Team, "Australian Recordkeeping Metadata Schema, Version 1.0," (29 May 2000). [Authored by: Sue McKemmish, Glenda Acland, Kate Cumming, Barbara Reed, Nigel Ward.]

[2] Sue McKemmish, Glenda Acland, Nigel Ward and Barbara Reed, "Describing Records in Context in the Continuum: the Australian Recordkeeping Metadata Schema," *Archivaria* (forthcoming).

[3] Standards Australia, *Australian Standard – Records Management* (AS4390 – 1996).

[4] See: Geoffrey C. Bowker and Susan Leigh Star, *Sorting Things Out: Classification and its Consequences* (Cambridge, Massachusetts: The MIT Press, 1999).

[5] See: www.sdsc.edu/NARA/

[6] See: www.naa.gov.au/recordkeeping/control/rkms/contents.html

[7] See: http://www.indecs.org/. See also: Godfrey Rust, "Metadata: The Right Approach -- An Integrated Model for Descriptive and Rights Metadata in E-commerce," *D-Lib Magazine* (July/August 1998). <http://www.dlib.org/dlib/july98/rust/07rust.html>

[8] See: www.ilrt.bris.ac.uk/discovery/harmony/