Discussion paper on the nature and role of metadata in the creation of reliable and the preservation of authentic records in electronic systems

Anne Gilliland
February 2005

Introduction

This paper addresses questions regarding the conceptualization, definitions and roles of metadata and archival description as they arise from InterPARES 1 and are operationalised in the work of InterPARES 2. Inevitably, however, because of the needs to communicate the work of InterPARES in a meaningful way across not only other disciplines, but also different archival traditions; to interface with, evaluate and inform existing standards, practices and other research projects; and to ensure interoperability across the three focus areas of the project, this discussion cannot address these questions in a vacuum devoid of reference to wider thinking about and developments in recordkeeping[1] and metadata.

Specifically, the paper raises the following questions, either directly or by implication:

- Can "metadata" be applied as an *overarching term* that includes the products of archival description?
- Is metadata a new *concept* that needs to be understood in terms of its implications for archival theory?
- In looking at how to operationalise the Benchmark and Baseline Requirements generated by InterPARES 1, in what ways can metadata ensure the creation of reliable and preservation of authentic records in electronic systems?
- Can a vocabulary be created to assist in the identification of different types and functions of metadata?
- What kind of management regime needs to be put in place to ensure the creation and maintenance of trustworthy metadata?
- Do archival descriptive practices need to be extended to address the needs of electronic records more effectively?
- Can metadata associated with the creation and active use of records ever contribute to archival description, particularly in the capture and elucidation of certain kinds of context and fundamental identification and arrangement information relating to the records?
- Should a metadata specification model generated out of InterPARES 2 support a single or multiple worldviews on the activities, roles, responsibilities, and points of engagement with the record?
- Can metadata-based automated tools support any new kinds of roles and capabilities for the description and use of archival electronic records?

---

1. A note about the use of the terms "recordkeeping" and "archival preservation [or … management, administration and so forth]", in this paper. "Recordkeeping" is often used in standards and the archival literature to signify an archival worldview of the integration and continual interactivity of processes and responsibilities related both to records creation and to archival management of those records. However, this is not a universally accepted premise, with the life cycle model drawing a much clearer demarcation between the management of active records and the preservation of archival records. The issue becomes even more complex if the issue of whether or not records *necessarily* need be physically transferred into archival custody is addressed. In order not to distract from the overall flow of the paper, therefore, I have chosen to use both sets of terms in conjunction with each other, except where a specific distinction is drawn.

These questions have particular relevance for specifying how the Benchmark and Baseline Requirements are implemented within recordkeeping and archival processes and systems design, as well as for the conceptualization and labeling of the models being developed. This discussion, however, arises primarily out of the activities of the Description Cross-Domain as it seeks to address the following research questions posed in the proposal funded by SSHRC:[2]

- What is the role of descriptive schemas and instruments in records creation, control, maintenance, appraisal, preservation, and use in traditional record-keeping systems in the three focus areas?

- What is the role of descriptive schemas and instruments in records creation, control, maintenance, appraisal, preservation, and use in emerging record-keeping systems in digital and web-based environments in the three focus areas? Do new tools need to be developed, and if so, what should they be? If not, should present instruments be broadened, enriched, adapted?

- What is the role of descriptive schemas and instruments in addressing reliability, accuracy and authenticity requirements (including the InterPARES 1 Benchmark and Baseline Authenticity Requirements) concerning the records investigated by InterPARES 2?

- What is the role of descriptive schemas and instruments in archival processes concerned with the long-term preservation of the records in question?

- Do current interoperable frameworks support the interoperability of descriptive schema and instruments across the three focus areas? If not, what kinds of frameworks are needed?

- What are the implications of the answers to the above questions for traditional archival descriptive standards, systems and strategies? Will they need to be modified to enable

---

2. Additional research objectives identified in the funded National Science Foundation and National Historical Publications and Records Commission proposals, per those agencies' funding objectives, include the following:

On the basis of this [theoretical] understanding, the project will then formulate and test technological, metadata, and policy models and use these as a basis with and against which to formulate, analyze and test:

- new and existing methodologies and strategies for ensuring that records created using these systems can be trusted as to their content (that is, are reliable and accurate) and as records (that is, are authentic) while used by the creator;
- new and existing methodologies and strategies for selecting records that have to be kept for legal, administrative, social or cultural reasons after they are no longer needed by the creator;
- new and existing methodologies and strategies for preserving them in authentic form over the long term; and
- advanced technologies for the implementation of these methodologies in different sectors, and disciplinary and socio-cultural contexts …
- Hypothesis of metadata necessary for prototype systems
- Rules for ongoing description of electronic records

And

The Description (Metadata) Research Team, working in collaboration with the ISO group that is currently drafting a new Recordkeeping Metadata standard, is identifying the metadata requirements for life cycle recordkeeping. It is also analyzing existing metadata schema and frameworks within affected sectors in order to identify in what ways they might need to be extended to address records preservation and authenticity concerns, and developing specifications for automated tools for the creation, capture, and long-term preservation of trustworthy metadata.

archival programs to meet new requirements, or will new ones need to be developed? If so, what should they be?

- To what extent do existing descriptive schemas and instruments used in the sectors concerned with the focus areas addressed by this project (for example, the geo-spatial data community) support and inform requirements such as those developed by InterPARES 1? Will they need to be modified to enable these sectors to meet these requirements, or will new ones need to be developed? If so, what should they be?

- What is the relationship between the role of descriptive schemas and instruments needed by the creator and those required by the preserver to support the archival processes of appraisal, preservation and dissemination? What tools are needed to support the export/import/exchange of descriptive data between systems?

- What is the role of descriptive schemas and instruments in rights management and in identifying and tracking records components, versions, expressions, performances, and other manifestations, and derivative works?

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

In addressing these research objectives, the Description Cross-Domain Group has been involved in the following activities:

- Collecting data on the types and sources of metadata being identified through case studies being conducted in other InterPARES' groups.

- Developing a database for analyzing warrant (i.e., the mandate from law, professional best practices, professional literature, and other social sources) requiring the creation and continued maintenance of archival description and other metadata supporting the accuracy, reliability, authenticity and preservation of records. This warrant will be integrated into recommendations made by the Description and other InterPARES2 Groups with regard to evaluating, extending or revising existing descriptive and metadata schemas as well as promoting the Metadata Specification Model (see below).

- Developing and compiling a metadata schema registry that unambiguously describes and analyzes salient features of relevant extant descriptive and other metadata schemas, element sets, standards and application profiles, as well as identify existing cross-walks between them. The registry evaluates each against the Benchmark and Baseline requirements as well as requirements identified from other archival and recordkeeping description and metadata standards and guidelines. The outputs of the evaluation are recommendations for how each registered entity might be extended or otherwise revised to address the reliability, authenticity and preservation needs of records created within the domain, community or sector to which they pertain. The metadata schema registry itself is a product of the research (one that will be integrated into the ISO 23081 Recordkeeping Metadata Standard), in that it provides a standardized framework by which any metadata schema or set could be assessed for its ability to address these needs.

- Developing and testing metadata specifications relating to the activity, entity and data models developed by the Modeling Group which identify the type, source and application of metadata identified in the models, and the existence of relevant metadata schemas. Ultimately, these are to be developed into a metadata specification model that will identify an overall set of metadata requirements that specify what metadata needs to be created, how, and by whom at all points within the InterPARES models. This will also for the basis for a set of specifications for automated tools (not to be confused with descriptive instruments) that can be used to

assist with the creation, capture, management and preservation of essential metadata for active and preserved records.

- Interfacing with other relevant R&D activities such as ISO 23081 development, the Clever Metadata Project and the work of the San Diego Supercomputer Center on the development of metadata tools for the automated creation, harvesting, and end-user manipulation of metadata.

Background

"Metadata" is a concept that is comparatively new to the field of recordkeeping and archives. The term was coined about twenty years ago by the geospatial and data processing communities to refer to "data about data" (essentially salient information--descriptive, provenancial, technical--about scientific data that was not the data itself). The term was appropriated in the 1990s by the bibliographic community and used to refer to value-added information about a resource (i.e., a surrogate for that source) that was created primarily by catalogers. Since then, the term has been used increasingly by information professionals to refer to any kind of descriptive or resource discovery information, whether manually or automatically created.

The first usage of the term "metadata" in the archival science literature came in the early 1990s from electronic records archivists who advocated a "metadata systems"-based approach to the management of electronic records systems. They argued that metadata could capture and preserve record context, preserve systems and record structure, generate and retain relevant descriptive information, incorporate appraisal and disposition data, manage records throughout the life, preserve and migrate system functionality, and create inventory/locator systems for organizational information sources. [3] For this potential to be realized, however, they acknowledged that archivists would need to be involved in specifying which metadata to capture and how, which in turn would require that archivists become involved in collaborations with records creators and recordkeeping systems designers.[4]

Included in the metadata systems approach was a recommendation that metadata created by active recordkeeping systems could be used as an input to, or even to replace archival description. This was advocated in part to obviate some of the labour-intensive aspects of manual description conducted by the archivist, especially in the face of large processing backlogs and potentially unprecedented volumes of electronic records. The advocacy of using metadata to support or even replace archival description, however, also emanated out of prevailing sentiments in the electronic records community that contemporary archival descriptive practices failed to address at a detailed level how electronic records should be described,[5] and that

---

3. See for example, David Wallace, "Metadata and the Archival Management of Electronic Records: A Review," *Archivaria* 36 (Autumn 1993): 87-88; Charles M. Dollar. *Archival Theory and Information Technologies: The Impact of Information Technologies on Archival Principles and Methods* (Ancona, 1992); Margaret Hedstrom, "Descriptive Practices for Electronic Records: Deciding What is Essential and Imagining What is Possible," *Archivaria* 36 (Autumn 1993).

4. Wallace, David, "Metadata and the Archival Management of Electronic Records," ibid.

5. Although it was also the case that few attempts had been made to apply full archival hierarchical description to electronic records, since so few had been accessioned to date, and those mainly comprised flat data files. See Jean Dryden, "Archival Description of Electronic Records: An Examination of Current

the *post hoc* nature of archival description meant that sometimes contextual information crucial for understanding the evidentiary nature of the records no longer existed.[6]

This recommendation met with debate in the Canadian archival literature based upon several considerations: that electronic records archivists were mistakenly asserting that archival descriptive practices focused on content rather than context, when in fact this was the foundation of descriptive theory and existing standards; a descriptive approach based on metadata automatically derived from the electronic system and its records would be unable to capture all the documentary and administrative relationships that exist across all types of records, electronic or non-electronic, over time within the archive; while metadata can reveal great detail about the transactionality associated with records, it is unable to provide a broader overview that can elucidate "the geneaological ties that bind it, the personal, familial, professional, and societal influences that shaped it, and the evolution of these factors over time;"[7] it fails to take the needs of secondary users into account and if it did, its integrity could be compromised, thus undermining its ability to provide evidence of the actions and transactions with which it is associated.[8]

In the past ten years, however, archival theoretical thinking about metadata and the technological tools available to exploit metadata have both become considerably more sophisticated. More complex analyses of and recommendations about the specific types and functions of metadata that might support the creation of reliable and preservation of authentic electronic records have resulted in the inception of the term "recordkeeping metadata" and the development in many jurisdictions of recordkeeping metadata standards and guidelines. While there is still not consensus about the scope of recordkeeping metadata,[9] models such as the SPIRT Australian Recordkeeping Metadata Schema (RKMS), which identifies event and process-related metadata, give the construct a purview that is actually quite a bit broader than how metadata is conceptualized in many other communities.[10] Research collaborations,

---

Practices," *Archivaria* 40 (Fall 1995): 99-108. Chapter 9 of the Canadian standard for archival description, *The Rules for Archival Description*, revised August 2003, now is possibly the only descriptive standard that directly addresses the description of records in electronic form, available: http://www.cdncouncilarchives.ca/archdesrules.html

6. Wallace, David, "Managing the Present: Metadata as Archival Description," *Archivaria* 39 (March 1995):10-21.

7. MacNeil, Heather, "Metadata Strategies and Archival Description: Comparing Apples to Oranges." *Archivaria* 39 (March 1995): 22-32.

8. See MacNeil, "Metadata Strategies and Archival Description," ibid; Wendy Duff, "Will **Metadata** Replace **Archival** Description: A Commentary." *Archivaria* 39 (March 1995): 33-38.

9. Archief School, Netherlands Institute for Archival Education and Research, *Proceedings of the Archiving Metadata Forum*, June 5-8, 2000 (draft). Available: http://www.archiefschool.nl/docs/workproc.pdf

10. See McKemmish, Sue, Glenda Acland, Kate Cumming, Barbara Reed, and Nigel Ward, *Australian Recordkeeping Metadata Schema, Version 1.0*, 31 May 2000, available at http://www.sims.monash.edu.au/rcrg/research/spirt/deliver/index.html; McKemmish, Sue, Glenda Acland, Nigel Ward, and Barbara Reed, "Describing Records in Context in the Continuum: the Australian Recordkeeping Metadata Schema," *Archivaria*, 48, (Fall 1999): 3-43.

such as that between the U.S. National Archives and Records Administration and the San Diego Supercomputer Center to develop the Permanent Archives Technology, and the industry/academia partnership of the Clever Recordkeeping Metadata Project are both working to develop automated tools that will exploit metadata for archival purposes as well as to facilitate new kinds of uses.[11] In the same period, use of the term "metadata" by archivists has become more ubiquitous, and it is now, probably due to the influence of the bibliographic community, frequently applied in relation to the standards that apply to archival description and to the products of the archival description process.

The following examples indicate how, in recent years, the term has not only become a part of the parlance of archives and recordkeeping in a range of constituencies, but has become embedded in the language of standards that direct archival activity:

> "The Australian Standard for Records Management provides advice on how to design and implement recordkeeping systems that will capture and manage the content and context of transactions. The standard recommends that records be registered in a recordkeeping system and linked to descriptive information about their context (AS 4390.4, section 6.2.2). Such descriptive information is now referred to by recordkeeping professionals as 'metadata'.
>
> The term 'metadata' originally emerged in the IT community, but the concept has been employed by information professionals for some years to describe information that is used to facilitate intellectual control of, and structured access to, information resources in library collections, file registries and archival holdings. Traditional records management tools such as file registers, file covers, movement cards, thesauri and indexes all provide metadata about records. Such tools help records managers control and manage records, and provide important contextual information about who used records, how and when. Traditionally, archivists provided additional metadata by creating indexes, file lists and other finding aids that helped researchers to locate and understand records once they were transferred from the organisational environment in which they were created to archival custody." **-- *National Archives of Australia Recordkeeping Metadata Standard for Commonwealth Agencies* (1999).**
>
> "Metadata is data describing data. In electronic document and records management, this means data that must be captured along with electronic records to enable them to be understood and to support their management and use.
>
> • Metadata assists with the retrieval of records;
> • Metadata improves the management of records;
> • Metadata is necessary to document the content, context and structure of a record;
> • Metadata is used to document transactions relating to a record; and
> •Contextual and descriptive metadata is essential to the integrity of records and to document a record's commitment as a record." **--*Record Keeping Metadata Requirements for the Government of Canada*. January 2001 (produced by the Records/Document/Information Management (RDIMS) Working Group on Work Processes and Practices (WPPWG) and endorsed by the Information Management Forum sub-group on Metadata).**

11. For further information on these projects, see Robert Chadduck, Joseph JaJa, Fritz McCall, Reagan Moore, and Mike Somrul, "Digital Archiving and Long Term Preservation: An Early Experience with Grid and Digital Library Technologies," available: http://www.archiefschool.nl/docs/workproc.pdf; and Clever Recordkeeping Metadata, http://www.archiefschool.nl/docs/workproc.pdf

"Recordkeeping metadata is information that helps records to be identifiable, accessible and meaningful. It also enables records to be effectively managed by indicating such things as how long they need to be kept and who should have access to them. Furthermore, recordkeeping metadata ensures that there is an audit trail showing what has happened to records over time.

The capture and management of recordkeeping metadata is a standard component of recordkeeping practices. Metadata has traditionally been captured on the covers of files or within registers, indexes, and other means of controlling and managing records. In practice, not all information needed to control and manage records has always been captured as metadata. For example, important information relating to record use or management has frequently been conveyed through the physical nature or arrangement of the records or has been common knowledge among the people using the records. In the modern business environment, however, where records can be created and stored in electronic form and managed in decentralised ways across an organisation, there is a need to make information necessary for record use and management more explicit." **-- Background, *Standard No 5 NSW Recordkeeping Metadata Standard*, Part 1: Introduction to the NSW Recordkeeping Metadata Standard (Section 1 of 1), Issued June 2001.**

"Metadata is often defined as 'data about data.' To elaborate, it is descriptive information that facilitates management of, and access to, other information. A traditional example of metadata would be the bibliographic information found in card catalogs. Recordkeeping metadata facilitates such records management actions as discovery, preservation, and disposition. While optimum metadata for any particular record set may vary, such information often includes items like the name of the record creator, date and time of creation, record identifier, key words, location, and retention information. It can also give reference to applicable policies and laws like the Minnesota Government Data Practices Act and even specific sections within those documents.

Without adequate metadata, a number of records management problems can arise, particularly with respect to electronic records. To list a few examples, it may be difficult to: locate and evaluate records, pinpoint the official record when multiple copies exist, determine whether a record has been modified since its creation, determine who should have access to a record, and carry out the proper disposition of a record (e.g., archive, destroy) at the end of its retention period. Recordkeeping requirements and associated metadata are best designed into a system as part of its core functionality, not as a tacked-on afterthought." ***--Minnesota Recordkeeping Metadata Standard, IRM Standard 20, Version 1.2*, April 2003.**

"Metadata is not a user-friendly term. However, the underlying concept … is relatively straightforward; metadata is simply meaningful data describing another data object … Indeed, outside information science the term is more confusing than helpful, since it almost suggests that metadata is one specific category of data and that metadata standards should apply to all metadata. Neither of these suggestions is true. If one leaves out the word 'information', metadata becomes just data about any kind of object. What is 'meaningful' depends on the kind of object the data is about, as well as its purpose for use. For archival (meta)data the object is not always one, discrete object, but often a complex of interrelated objects, both physical and abstract, including documents, document aggregations, document creators, business processes, curators, etc., reflected in the three major archival concepts about records: content, structure and context. An archival document is not just one, 'discrete' data object, and the data about it archivists create, collect, update and use, are not just descriptive metadata, but they serve a variety of archival business functions, including retrieval, preservation, storage, appraisal, disposition, etc. They are indispensable for physical, administrative and intellectual control over archival materials." **--Peter Horsman,"**

Metadata and Archival Description" European Union Archive Network (EUAN). Available: http://www.euan.org/euan_meta.html.

"In a records management context, metadata are data describing the context, content and structure of records and their management through time (ISO 15489-1, Clause 3.12). As such metadata are structured or semi-structured information that enables the creation, registration, classification, access, preservation and disposition of records through time and within and across domains. " **--ISO 23081 Recordkeeping Metadata Standard being developed by TC46/SC11-WG1.**

We can place these statements alongside successive professional definitions for archival description:

Description is "the process of establishing intellectual control over holdings through the preparation of finding aids." ***--A Basic Glossary for Archivists, Manuscript Curators, and Records Managers* (Society of American Archivists, 1974).**

"Archival description is the process of capturing, collating, analyzing, and organizing any information that serves to identify, manage, locate, and interpret the holdings of archival institutions and explain the contexts and records systems from which those holdings were selected**."—SAA Working Group on Standards for Archival Description (WGSAD), 1988.**

Archival description is "(1) The process of analyzing, organizing, and recording information that serves to identify, manage, locate, and explain the holdings of archives and manuscript repositories and the contexts and records systems from which those holdings were selected; (2) The written representations or products of the above processes; (3) In records management, a written account of the physical characteristics, informational content, and functional purpose of a record series or system**."—Lynn Bellardo, *Glossary for Archivists, Manuscript Curators, and Records Managers* (SAA, 1992).**

"Archival description is the process of capturing, collating, analyzing, and organizing any information that serves to identify, manage, locate, and interpret the holdings of archival institutions and explain the contexts and records systems from which those holdings were selected**."—Victoria Irons Walch (comp), *Standards for Archival Description: A Handbook*, (Society of American Archivists, 1994). Available: http://www.archivists.org/catalog/stds99/index.html**

"The creation of an accurate representation of a unit of description and its component parts, if any, by capturing, analyzing, organizing, and recording information that serves to identify, manage, locate and explain archival materials and the context and records systems which produced it." **–International Council on Archives, *General International Standard Archival Description* (*ISAD(G)*), Second Edition, 1999.**

"The recording in a standardized form of information about the structure, function and content of records," **--Rules for Archival Description (RAD), rev. August 2003 (Association of Canadian Archivists). Available: http://www.cdncouncilarchives.ca/RAD_Glossary_revised_Aug2003.pdf**

It is also illuminating to look at literary sources that discuss the purposes of, and processes associated with archival description:

"Before one can properly describe a volume or bundle, one must have a clear conception of the dominant idea that presided over its *formation* … It may be said that a definite dominant idea has *always* presided over the formation of a volume, a file, a dossier or a bundle. The archivist must understand this idea if he is to be able to describe well the archival item in question … Generally, this idea is easy to recognize. If, for example, a volume is found composed exclusively of the decisions of a council following in chronological order at fixed intervals of a few days, one may be sure that he has before him the register of resolutions of the council." **–Muller, Feith and**

Fruin, *Manual for the Arrangement & Description of* *Archives*, The H.W. Wilson Co, 1968):101-2.

"The archives should design a system of finding aids that provides essential information about the holdings for users and enables the archivist to retrieve materials. Finding aids should employ first the techniques of group and series description before undertaking item description; a brief description of all records is preferred to a detailed description of some of them. The level of description of records depends on their research value, the anticipated level of demand, and their physical condition. Finding aids may include, as appropriate, guides, inventories or registers, card catalogs, special lists, shelf and box lists, indexes, calendars, and for machine-readable records, software documentation**."-- 1980 SAA Task Force on Institutional Evaluation "Principles for Institutional Evaluation," 1980.**

Archival description involves "the identification of the documents, the assignment to them of an intellectual and physical place in the whole of the authentic documents, that is, their location and description in context, by freezing and perpetuating their interrelationships, ensure that possible tampering will be easy to identify."**—Luciana Duranti, "Archives as a Place," *Archives and Manuscripts* 24 no. 2 (1996): 247.**

"Archival description is the process of capturing, collating, analyzing, controlling, exchanging, and providing access to information about (1) the origin, context, and provenance of different sets of records, (2) their filing structure, (3) their form and content, (4) their relationships with other records, and (5) the ways in which they can be found and used**."—Fredric Miller, *Arranging and Describing Manuscripts* (SAA, 1990).**

"If archival description is defined as the post-transfer process of establishing intellectual control over archival holdings by preparing descriptions of the records, then those descriptions essentially function as cataloguing records, surrogates whose primary purpose is to help researchers find relevant records. In the continuum, archival description is instead envisaged as part of a complex series of recordkeeping processes involving the attribution of authoritative metadata from the time of records creation." … "the concept as being "standardised information about the identity, authenticity, content, structure, context and essential management requirements of records" **--Sue McKemmish and Dagmar Parer, "Towards Frameworks for Standardising Recordkeeping Metadata.' *Archives and Manuscripts*, 26 (1998):24-45. Available: http://www.sims.monash.edu.au/rcrg/publications/recordkeepingmetadata/smckrmp1.html**

"The Bentley research group identified three main purposes of archival description, and the methods by which these purposes are achieved. These purposes are:

- to provide access to archival materials by means of a description that is retrievable, at a minimum, by provenance,
- to promote the understanding of such materials by documenting their context, content, and structure, and
- to establish their authenticity by documenting their chain of custody, their arrangement, and the circumstances of their creation and use**."**

-- Statement of Principles for the CUSTARD Project *(Canadian-U.S. Task Force on Archival Description). Available: http://www.archivists.org/news/custardproject.asp*

"Creating a finding aid is a descriptive process that may begin before a collection ever arrives in your repository. If, for example, you are a manuscript curator or archivist for a collecting repository, you will likely gather crucial data about the origins, provenance, and chain of custody of a collection during the course of identifying the

materials for possible acquisition. You may meet with the creator and/or donor of the records or papers, learn something about the context in which the materials were generated, and glean from background reading, conversation, and examination, bits and pieces of information about the organization, scope, and content of the materials under consideration. A memorandum summarizing your findings may be prepared, filed away, and possibly forgotten about until months or years later when the materials arrive on your doorstep through donation, deposit, purchase, or transfer. At that time, you may pull out this memorandum and compare it to the materials before you. You will likely also review packing lists and legal documents, such as the deed of gift or purchase agreement. Through this discovery process, you will begin to piece together a rudimentary accession record or preliminary catalog record containing some of the basic content of a future EAD finding aid:

- What is the name or title of the collection?
- Who created the material and for what purpose?
- What dates does it cover?
- How much material is there?
- What genres or formats are represented?
- How did it come into the repository's control or possession?
- Who or what was the immediate source of the acquisition?
- Are there restrictions on access or reproduction?
- Has it been assigned a unique identification number for tracking within the repository?
- What storage location will be used for the materials?
- Have any materials been separated for transfer to other units in your repository?

The more information that can be captured at this stage the better, especially if the facts are based on oral sources and are unrecorded elsewhere. This initial collection-level description may be viewed by the archivist more as inventory control than an access tool, but gathering and recording the information is an investment in archival description that will reap significant rewards when the data is teased apart and easily mapped to counterpart EAD data categories. From these earliest acquisition and accessioning records, a finding aid author can begin to extract a fundamental description of the collection in its entirety … and start to outline important background information about how the collection was acquired and the conditions under which it is administered by the repository and used by researchers. This latter "administrative information" … will help future finding aid readers know how to approach the collection and make use of the data they find.

At the outset of processing the collection, additional information suitable for inclusion in an EAD finding aid is assembled. In an effort to educate yourself about the materials, you may track down biographies, agency histories, or corporate chronologies about the creator. You may prepare a crib sheet to refer to during processing that identifies the key dates and events in a person's or organization's life. As suggested in section 3.5.1.5, this processing aid can become a public reference tool when included in an EAD finding aid as a biographical note or agency history designed to enhance researchers' understanding of the origins and context of the archival materials. Also easily accommodated in EAD are bibliographies, such as the list of sources you may have prepared during your background research, and other types of "adjunct descriptive data" described in section 3.5.4.

Background reading and consultation of external sources continues throughout processing, but the next stage of organizing and describing a collection involves studying its existing order and structure to identify its majors parts and deduce how those parts have been or could be divided into smaller components. Once the organization has been determined, the focus shifts to issues of arrangement, which relate to how the materials are filed (alphabetical, chronological, etc.) within the higher-level components. During the analysis of the collection, you will likely record

> information about its current organization and arrangement and may incorporate such information into a processing proposal, which outlines how the various parts will be prepared for research use. In delineating both the original and projected structure of the collection, the processing proposal lays the groundwork for building a multilevel EAD finding aid, which, as described in section 3.4, provides a summary description of the whole collection, followed by progressively more detailed descriptions of the parts." ***--EAD Application Guidelines for Version 1.0,*** **Section 3.2. Collecting Data for a Finding Aid. Available: http://www.loc.gov/ead/ag/agcreate.html**

From these statements, one can discern some evidence of a distinction in terms of how archivists are conceptualizing "metadata" (i.e., generic meaningful data describing another data object, which could include types of metadata generated by an electronic system that do not necessarily have relevance to recordkeeping and archives);[12] "recordkeeping metadata" (i.e., documentation of the content, context and structure of a record, documenting the business process in which the record participates, providing an audit trail for the record, supporting the registration, classification, management, retrieval, disposition and preservation of the record, identifying agents and events associated with the record, and, in some cases, archival description), and "archival description" (i.e., the processes associated with, and product of an intellectual analysis and synthesis by the archivist of a range of sources that would help to identify, manage and explain the records, physical and intellectual integration with related records in other formats, and the description of the content, context and structure of records in such a way as to ensure that their continued authenticity can be ensured and demonstrated). Nevertheless, there remains a lack of agreement about exactly how all three terms are used and what are the precise relationships between them. What is frequently missing, also, from these statements, is the fact that metadata can be a record in its own right, for example as a record of a transaction.

Can "metadata" be applied as an *overarching term* to refer to the products of archival description?

From the above discussion, one can see that "metadata" is already being used by some communities and individuals in a way that includes archival description as one, not always distinct, facet. The conceptual nub of this question, if one accepts that it goes beyond disagreements about nomenclature that have roots in different archival traditions and in purposeful appropriation or incidental colonization of the archival field's terms by those in more influential fields, is whether archival description is a distinct intellectual activity that serves a function that is identifiably discrete from metadata. Those who would argue that archival description is not a kind of metadata argue that metadata accrues to the record during the processes of its creation and active life and that archival description is both the manual process and the product

12. One additional set of characteristics of metadata should be noted here, however, since it has direct implications for the preservation of electronic records. There is not always a one-to-one relationship between an individual record or even a record series and a particular category of metadata. While recordkeeping metadata can be associated with individual records and with groups of records, for example, identifying them or notifying of or enforcing policy as in the case of records retention schedules or rights and restrictions metadata, they can also operate at a system level, for example, implementing predictable data structures or documenting administrative and work processes, such as administrative rules. This raises issues about the complexity of relationships and linkages that might exist between records and metadata and how they might be preserved with their referential integrity intact over time.

that occurs when a record is transferred into archival custody. From this viewpoint, archival description not only creates a value-added description that documents the various contexts of the archival record, but is the mechanism through which the record is intellectually incorporated into the archival bond and is, therefore, the primary means by which the continued authenticity of archival records can or should be assured.

To parse the issue of nomenclature for a moment, the archival profession is comprised not only of multiple archival traditions of differing intellectual lineages and juridical contingencies, but also multiple interest areas, all of which develop and push their own perspectives with varying amounts of autonomy or isolation from the rest of the field. This naturally results in differences in terminology. Sir Hilary Jenkinson noted in 1958, "the difficulty—indeed impossibility—of finding literal translations for the technicalities of one countries Archives in the language of another" and the "difficulty of deciding on a *single governing or co-ordinating language* under which information internationally collected could be set out." However, Jenkinson goes on to say "rather regretfully," that there should be an attempt "to secure the greatest possible measure of uniformity in the nomenclature of English Archives and of the methods and instruments we use in dealing with them; simply for our own English convenience."[13] As the previous discussion has indicated, there are already several instances in archival practice where "metadata" has been used to refer to archival description, including in key standards, one of the primary ways whereby nomenclature gets standardized in professional communities. Why has the term been used in this way? Was it consciously and deliberately adopted in these cases to represent a concept for which we did not in the past have an overarching rubric that covered traditional and emerging forms of metadata (albeit that there existed terms to describe types of metadata, such as classification codes, indexes, and locator files), or was it the result of a deliberate attempt to place and explain archival activities, especially those relating to electronic records, alongside those of other information fields and upgrade status? Or was it the result of terminological drift that occurred when larger numbers of archivists were educated in programs together with other kinds of information professionals? In different communities, these questions may have different answers. Differences that are purely differences in nomenclature may never be adequately resolved and may require the use of a glossary and thesaurus to facilitate mutual understanding.

Differences in conceptual understanding of the same term are more problematic, and need to be clarified, either by agreeing upon usage, again a difficult task, or by clarifying the semantics within the context of usage (and it should be noted that application of different nomenclature to the same concept is equally problematic). InterPARES 2 researchers need to arrive at a decision as to whether the project will accept, from a terminological perspective, that the products of archival description are, in a larger sense, forms of metadata or represent a different intellectual activity and concept, and the terminology of the project should reflect this decision.

Is metadata a new *concept* that needs to be understood in terms of its implications for archival theory?

13.Roger Ellis and Peter Walne, eds. *Selected Writings of Sir Hilary Jenkinson* (Chicago, IL: SAA), 2003: 351-352.

As the opening discussion indicated, metadata is a comparatively new piece of terminology for all fields. However, given that archivists have always worked with items such as classification schemes, file lists and locator files, it could be argued that what we are now seeing are merely manifestations of such items in electronic form. In this case, while it is probably convenient to have some uniformity of nomenclature in terms of how we refer to such items, both individually and collectively, one could argue that we are probably not seeing the emergence of a new concept that requires us to re-think archival theory. However, if one looks at some of the research and conceptualization that has been undertaken in recent years, one can see some aspects that are new or at least reconceptualized. For example, electronic records archivists, and not just those who adhere to the records continuum perspective, have exhorted for many years that archivists need to be involved in the design of active recordkeeping systems in order to ensure that they will be capable of generating a reliable record and one that is capable of being set aside and transferred into archival custody. One component of this systems design is the development and building in of metadata structures that will ensure the system documents what is required. There are theoretical implications for this early engagement with the design of recordkeeping processes and systems that have been addressed by records continuum researchers, but these have yet to be comprehensively addressed by those who follow the life cycle approach.

There is also the issue of archivists reconceptualizing the scope of metadata, as has been the case with RKMS, and of requiring creators to generate new and additional types of metadata beyond those they would customarily have generated. To what extent do these activities interfere with the evidential value of the records and their metadata? To what extent is it the mandate of the archivist to help records creators to generate better records?

There are also conceptual implications for *post hoc* archival description. Peter Horsman has written that:

> "although up to now a close comparison between traditional descriptive metadata standards and metadata for electronic records has not been made, it is obvious first, that an integration of both categories should take place, and second, that full archival metadata standards will be broader than ISAD and ISAAR."[14]

Electronic records certainly have technical descriptive needs that are poorly addressed by most existing archival description standards, which would argue for augmenting the standards, but should there be a single end-to-end metadata structure for records? Is there anything about electronic records that might require archivists to approach description in a different way? For example, traditionally archivists would glean information about arrangement, content, salient contexts from documentation created by the creators, as well as what could be gathered from other printed and oral sources. If information about the various ways in which records can be arranged, viewed and reported can be captured automatically and integrated into the archival description, and if the system were able to monitor and analyse transactional and other patterns associated with the record (as has been the subject of experiments at the San Diego Supercomputer Center) and add these into the

14. Peter Horsman, "Metadata and Archival Description." Available: close comparison between traditional descriptive metadata standards and metadata

archival description, would this represent a profound conceptual challenge to the role of archival description?

In looking at how to operationalise the Benchmark and Baseline Requirements generated by InterPARES 1, in what ways can metadata ensure the creation of reliable and preservation of authentic records in electronic systems?

The primary set of conditions against which metadata schemas registered in the InterPARES2 metadata schema registry are measured are the Benchmark and Baseline Requirements that were generated out of the InterPARES 1 Project (see APPENDIX A). The Benchmark requirements are based on the notion of a trusted record-keeping system. They include requirements that support the presumption of the authenticity of electronic records before they are transferred to the preserver's custody. The Baseline Requirements are based on the notion of the preserver as trusted custodian, and support the production of authentic copies of electronic records after they have been transferred to the preserver's custody. These are the only extant sets of requirements that specifically address how creators and archivists can assess the authenticity of records. As noted in the InterpARES 1 Authenticity Task Force Report,

> The benchmark requirements identify the record attributes (metadata) that need to be 'explicitly expressed and inextricably linked' to a record in order for its identity and integrity to be asserted. The benchmark requirements also identify 'the kinds of procedural controls over the record's creation, handling and maintenance that support a presumption of its integrity'.[i] The role of the benchmark requirements is to act as a tool for preservers to use in assessing the authenticity of electronic records. The higher the number, and the greater the degree to which a system meets these requirements, then the stronger the presumption of the authenticity of the electronic records held within it. [p. 3]
>
> In contrast, the baseline requirements specify the requirements that must be met in order to produce authentic copies of electronic records from a preservation system. This includes archival descriptive metadata documenting 'the records juridical-administrative, provenancial, procedural and documentary contexts', and controls over the records transfer and reproduction processes to ensure the maintenance of the records' identity and integrity.15

The Benchmark and Baseline Requirements, however, were only expressed conceptually, and in narrative form, by InterPARES 1, and have not yet been operationalised for any kind of technological implementation, for example, as a set of logical propositions or production rules. Nor have the requirements been deconstructed in a way that would specify how other processes and metadata might help to meet them. For example, how might the different types of context identified in InterPARES1 be manifested or documented through metadata? One way of addressing this problem is to decompose archival and recordkeeping notions of "context" into types that can then be associated with specific processes and attributes. InterPARES 1 identified five different types of contexts as being relevant to the maintenance of authentic records over time: juridical-administrative,

15. Report of the Authenticity Task Force of the InterPARES Project. Available: http://www.interpares.org.

provenancial, procedural, documentary, and technological.[16] Some of these types need to be further decomposed in order to identify their constituent metadata manifestations.[17] Operationalising these requirements, therefore, in terms of the extent to which they might be met through the development of a rigorous and thorough metadata regime, has been a major aspect of developing the analytical framework underlying the metadata schema registry.[18]

Many of the Benchmark and Baseline Requirements could potentially be implemented through metadata and archival description, particularly such aspects as identity, linkages, documentation of documentary forms, juridical requirements, business rules and technical procedures, access privileges, establishment of the authoritative record when multiple copies exist, and transfer of relevant documentation (Benchmark Requirements); and almost every aspect of the Baseline Requirements. Accordingly, the work of the Description Cross-Domain Group on the Metadata Schema Registry seeks to operationalise these narrative requirements in terms of how they might be satisfied both through the metadata associated with the active record and recordkeeping system and archival description, and then to analyse existing scheams, standards, guidelines, best practices, implementation profiles, and so forth to assess the extent to which they meet the requirements, given their stated scope. Where they fall short, researchers will be in a position to recommend augmentations or modifications to ensure that the schemas, etc. meet those requirements that fall within the stated scope of each schema. Researchers will also be able to recommend companion metadata schemas that can be used to address those parts of the requirements that are unaddressed because they are out of scope. In this way, multiple models for managing records can be supported – both those that seek to apply an end-to-end recordkeeping metadata schema, and those where different parties have responsibility for different aspects of recordkeeping and archival preservation.

---

16. InterPARES 1 Project. *The Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project*. Available: http://www.interpares.org/book/index.cfm

17. For example, perhaps the juridical-administrative type could be decomposed to address specific types of juridical-administrative requirements that manifest themselves directly in emerging metadata initiatives, such as those relating to rights management for records. Digital rights management (DRM) metadata are increasingly being integrated into systems by creators, publishers, and information providers, for example, as mechanisms for expressing and automatically enforcing rights and licensing requirements relating to information resources. In an age where records are more and more often the product of private activity, or collaboration or of outsourcing relationships between government and the private sector, or academic research and industry, such developments not only reflect these changes in records creation but can have significant implications for both researchers and the types of preservation regimes to which the records may be subject.

18. In order to draw on as many perspectives as possible and to try to identify where there might be consensus or divergences about relevant recordkeeping requirements (especially where there might appear to be differing view points emerging from the life cycles and records continuum perspectives), requirements were also derived from an examination of ISO 15489 Information and documentation -- Records Management (2001), the U.S. Department of Defense's Design Criteria Standard for Electronic Records Management Software Applications(DoD 5015.2-STD, 2002), and the European Union's Model Requirements for the Management of Electronic Records (MoReq) specifying requirements for Electronic Records Management Systems (ERMS). The primary requirements, however, are drawn from InterPARES1.

Can a vocabulary be created to assist in the identification of different types and functions of metadata?

Arguably, metadata is a term that expresses a generic concept for which a universal term did not previously exist, but, by equal measure, the term is vague unless accompanied by a qualifier that indicates such things as the type, function, or source of the metadata. In order to be able to develop metadata specifications and to have these match up to the labeling used on the InterPARES2 models, there is a need to develop an unambiguous terminology that is sufficiently granular as to allow specific types of metadata and products of archival description and their associated processes, to be identified. Since no such granular vocabulary, currently exists, this could be an influential product of InterPARES2.

What kind of management regime needs to be put in place to ensure the creation and maintenance of trustworthy metadata?

As Chris Hurley has noted:

> "Contextual metadata documents circumstances relevant to the making of the record, who, when, how, why … Efforts now being made to regularize the process whereby knowledge of context is captured as metadata for electronic recordkeeping should not blind us to a fundamental truth. Because records themselves are timebound, metadata must be verified within a context which is both current **and** historical. Records cannot remain current unless the metadata is externally validated."[19]

If a more comprehensive and rigorously delineated metadata and archival description infrastructure is integral to creating reliable records and maintaining and demonstrating the authenticity of archival records, then there is a concomitant need for overt integrity control and transparency of that metadata and archival description. This can only be the case if the metadata themselves are trustworthy and comprehensively managed for as long as they are required. In other words, reliability and authenticity are concerns for recordkeeping metadata *as well as* for the records and recordkeeping processes to which they relate. Metadata generated and managed by records creators and archival description generated by archivists, must be sufficient, appropriate, understandable, and of high quality. Archivists must be cognizant that the accession records, finding aids, and use records they typically create today are not only part of the archival description for the records to which they relate, but they are also records in their own rights. The scrutiny, therefore, that archivists give to the records and recordkeeping metadata of others in order to assess and validate their management and reliability, they must also give to their own (hopefully, in the future, with the assistance of the analytical framework and metadata schema registry being developed by InterPARES2). Both the provenance and the version of metadata must be clearly identifiable. Moreover, metadata, which often capture much of the context of records, potentially also offer researchers a rich source of intelligence about records creators and record creation processes.

What this argues for is an end-to-end metadata management regime that addresses which metadata need to be carried forward in time, for what purposes, by whom,

---

19. Chris Hurley, "Abandoned Children to Zoos," *Archivaria* 40, Fall 1995. Available: http://www.sims.monash.edu.au/research/rcrg/publications/ambientf.htm

and how they are to be preserved and validated. Bound up with this, however, are sticky issues associated with how much metadata one ends up accumulating and managing over time (including metadata associated with the preservation, reproduction and dissemination aspects of the archival function), and whether certain metadata can be efficiently segregated and eliminated after validation, certification and summarization by a preserver. Potentially, one aspect of the metadata specification model to be developed by the Description Cross-Domain Group in concert with the Modeling Group would be the identification of these aspects for each type of metadata specified.

Can metadata associated with the creation and active use of records ever contribute to archival description, particularly in the capture and elucidation of certain kinds of context and fundamental identification and arrangement information relating to the records?

As already discussed, one aspect of metadata creation and management that makes some in the archival community nervous is the notion, raised anew by projects such as the Archivists' Workbench,[20] that certain types of metadata created while the records are active could be captured or analysed automatically and used to partially automate, or even to replace archival description. As identified by InterPARES1, records have many types of interacting contexts that need to be documented. Often with electronic records, because of their virtual nature and also their complexity, it can be more difficult to identify these contexts than it might be with traditional records. However, often it is the case that the system within which the record has been created or maintained has in place metadata mechanisms, or could be designed to have them, that document some of the context in which archivists are interested (albeit that these are generally created contemporaneous with the record and lack the hindsight and birds-eye view of the archivist).

Indeed, what is distinctive about recordkeeping metadata is the range of ways in which they can automatically capture salient contexts of records as they move through time, space, systems, and types of use and user. For example, metadata can provide detailed descriptions of business processes and logs or audit trails of any changes made to records and associated dates. It can also describe the functionality of the original technical environment and enable users to distinguish the authoritative record from drafts and derivative versions. Metadata can also link separately stored data or record content to the appropriate documentary form to facilitate creating an imitative authentic copy of the original (an approach akin to that being used with the Persistent Archives Technology).

In the future, time and cost concerns as well as new technological capabilities are likely to necessitate that even archival description may be created, at least partially, by automated means, likely including harvesting and re-purposing metadata created by others prior to the records coming into archival custody. For this to be acceptable as an assistance or augmentation to archival description, however, a) the metadata harvested should supplement manual description or should capture some aspect that it is difficult or impossible to do manually; and b) archivists should assess what they do manually in traditional description and identify at the point of recordkeeping systems design what could be captured automatically out of the system. Neither of

20. See San Diego Supercomputer Center, Archivists' Workbench Project Summary. Available: http://www.sdsc.edu/NARA/Publications/nhprc_summary.pdf

these activities, however, necessarily usurps the archivist's prerogative to supplement and synthesize the metadata gathered automatically in the process of creating a descriptive instrument. Moreover, because the metadata thus gathered is likely to be in digital form, the archivist now has the option of retaining it both in its original form, as evidence of the records and recordkeeping to which it relates, and to transform it into a form that is more useful for secondary use.

Should a metadata specification model generated out of InterPARES 2 support a single or multiple worldviews on the activities, roles, responsibilities, and points of engagement with the record?

This question is surfaced by the Description Cross-Domain Group's work with regard not only to the scope of the metadata schemas that are being registered and analysed by the metadata schema registry (as already discussed), but also to the extent to which both the life cycle and continuum views on archival roles can be respected or even reconciled through the analytical approach embedded in the metadata schema registry. One of the great contributions, and benefits, of the InterPARES research over the past several years has been that it has brought together archival researchers not only from academe and practice, but also from very different archival traditions. This, however, has also led to moments of confusion and even contention as the divergent underlying perspectives and practices emerge and must be disambiguated and addressed. The Description Group researchers have found themselves faced with two alternatives—one being the development of research products that tolerate and support more than one approach, the other being to attempt to reconcile approaches that appear at first, and maybe even at second glance, to be irreconcilable.

The Description Group is at present attempting to straddle both of these alternatives, although, in the interests of full disclosure, it needs to be said that the analytical framework developed by the Group has turned out to be situated closer to the conceptualizations of recordkeeping metadata as expressed in particular by RKMS (the most delineated of existing recordkeeping metadata schemas) and supported by continuum thinking. This has been an inevitable consequence of taking the position that many different types of metadata and archival description are needed to satisfy even the InterPARES Benchmark and Baseline Requirements. Moreover, having made a conscious decision to assess the metadata implications of both of the dominant existing models, the relative extensiveness of the Continuum model, with the dimensionality afforded by its four axes of identity, evidentiality, transactionality and recordkeeping entity,[21] necessitated that the Description Cross-DomainGroup take a more complex view of metadata and archival description than might have been needed if it had looked only at supporting a Life Cycle model.

The activity models developed in InterPARES 1 were based on a life cycle view and presumed a custodial approach to the preservation of archival records. The Benchmark and Baseline Requirements identified responsibilities and capabilities for both the *creator* and the *preserver*, but were still predicated upon the physical

21 Upward, Frank, "Structuring the Records Continuum Part One: Post-custodial Principles and Properties," *Archives and Manuscripts* 24 no. 2 (Nov 1996): 268-85; Upward, Frank, "Structuring the Records Continuum Part Two: Structuration Theory and Recordkeeping,"*Archives and Manuscripts* 25 no. 1 (May 1997):10-35.

transfer of records into an archival repository. However, the Description Cross-Domain Group has also had to address the fact that while these two theoretical models currently exist (and it is, of course, quite possible, that further models might emerge in the future), many different kinds of implementations also exist. Some of these implementations adhere to the traditional life cycle view, but increasingly continuum thinking is influencing practices not only in Australia, but also in Northern Europe and the United States. What is more, archivists and other recordkeepers who are grappling with the challenges of electronic records, are developing their own hybrids of both approaches. In this context, it should be noted that although historically they have been linked closely together, conceptually it is not required that custodialism and non-custodialism be tied to adherence to the life cycle and continuum worldviews, respectively. It is also important to bear in mind that the world outside of archival science does not use these models, at least not conceived of in these terms, but communities other than archival communities are also targeted user groups for the metadata schema registry and analytical framework and their needs to be addressed.[22]

The work of the Description Cross-Domain Group attempts to resolve the tensions between the two approaches on a number of levels. In the Analysis section of the metadata registry, for example, it analyzes schema against criteria of both models. In doing so, the similarities between the two models will surface, and the overlapping and correspondence of criteria will be uncovered. The differences will also be made clear and will enable a detailed comparison of the requirements, their strengths and weaknesses, which will enable to determine whether the two are complete, mutually exclusive or even complementary. Working together with the InterPARES2 Modeling Group on the activity and entity models should also help to identify the points when metadata are or should be created or applied. Finally, the Description Cross-Domain Group intends to look at both sets of outcomes side-by-side with the intent of seeing whether or not they really are mutually exclusive, and whether or not it is viable to fold them into a unified metadata specification model that could be used within any existing or future recordkeeping approach.

Can metadata-based automated tools support any new kinds of capabilities for the description and use of archival electronic records?

Recordkeeping metadata are created in a variety of ways and by a variety of agents—they may be created manually (as is the case with most archival description) or automatically (as, for example, would be the case with an inverted index of terms culled from a text document). They may also be automatically inferred, derived or harvested from the records and recordkeeping systems themselves, an approach that looks increasingly attractive as systems developers and information professionals of all types become more aware of the burgeoning overhead of metadata creation and management necessary to support the online provision of trustworthy information. They may even be exploited and re-used for purposes for which they were never intended, such as for corporate knowledge mining, developing new institutional market segments, or developing learning objects. In the archival community, research and development activities such as

22 The Open Archival Information System (OAIS) Reference Model is a good example of a high-level model that at first glance seems to be a re-expression of a life cycle model, but upon further scrutiny could equally well support a continuum approach.

the Archivists' Workbench and PERM Projects of the San Diego Supercomputer Center have begun to explore the development of automated tools for metadata creation and management, as well as for the manipulation of records by end users, and the Clever Recordkeeping Metadata Project described in another article in this issue, which is looking at innovative ways of multi-purposing harvested recordkeeping metadata.

Approaches such as these potentially not only offer archivists a faster and less labor-intensive way to gain a measure of intellectual control over large volumes of electronic records, but also offer secondary users a much richer set of tools through which to access, manipulate and interpret archival records. They can also potentially support validation mechanisms for recordkeeping metadata and monitor the continued integrity of critical linkages that exist between records and their metadata. Perhaps the most important potential use of automated metadata tools, however, might be to support a metadata management regime, something which, if not automated, would be practically unimplementable for archivists.

APPENDIX A.

Benchmark Requirements Supporting the Presumption of Authenticity of Electronic Records[23]

Preamble

The benchmark requirements are the conditions that serve as a basis for the preserver's assessment of the authenticity of the creator's electronic records. Satisfaction of these benchmark requirements will enable the preserver to infer a record's authenticity on the basis of the manner in which the records have been created, handled, and maintained by the creator.

Within the benchmark requirements, Requirement A.1 identifies the core information about an electronic record—the immediate context of its creation and the manner in which it has been handled and maintained—that establishes the record's identity and lays a foundation for demonstrating its integrity. Requirements A.2–A.8 identify the kinds of procedural controls over the record's creation, handling, and maintenance that support a presumption of its integrity.

Benchmark Requirements (Requirement Set A)

To support a presumption of authenticity the preserver must obtain evidence that:

REQUIREMENT A.1: Expression of Record Attributes and Linkage to Record	the value of the following attributes are explicitly expressed and inextricably linked to every record. These attributes can be distinguished into categories, the first concerning the identity of records, and the second concerning the integrity of records.
	A.1.a Identity of the record:
	A.1.a.i Names of the persons concurring in the formation of the record, that is: • name of author[24] • name of writer 25(if different from the author) • name of originator26 (if different from name of author or writer) • name of addressee27
	A.1.a.ii Name of action or matter
	A.1.a.iii Date(s) of creation and transmission, that is:

23 Available: http://www.interpares.org

24 The name of the physical or juridical person having the authority and capacity to issue the record or in whose name or by whose command the record has been issued.

25 The name of the physical or juridical person having the authority and capacity to articulate the content of the record.

26 The name of the physical or juridical person assigned the electronic address in which the record has been generated and/or sent.

27 The name of the physical or juridical person(s) to whom the record is directed or for whom the record is intended.

- chronological date28
- received date29
- archival date30
- transmission date(s)31

A.1.a.iv	Expression of archival bond32 (e.g., classification code, file identifier)
A.1.a.v	Indication of attachments

A.1.b Integrity of the record:

A.1.b.i	Name of handling office33
A.1.b.ii	Name of office of primary responsibility34 (if different from handling office)
A.1.b.iii	Indication of types of annotations added to the record35
A.1.b.iv	Indication of technical modifications;36

922

REQUIREMENT A.2: Access Privileges	the creator has defined and effectively implemented access privileges concerning the creation, modification, annotation, relocation, and destruction of records;

923

REQUIREMENT A.3: Protective Procedures: Loss and Corruption of Records	the creator has established and effectively implemented procedures to prevent, discover, and correct loss or corruption of records;

924

28 The date, and possibly the time, of compilation of a record included in the record by the author or the electronic system on the author's behalf.
29 The date, and possibly the time, when a record is received by the addressee.
30 The date, and possibly the time, when a record is officially incorporated into the creator's records.
31 The date and time when a record leaves the space in which it was generated.
32 The archival bond is the relationship that links each record, incrementally, to the previous and subsequent ones and to all those participate in the same activity. It is originary (i.e., it comes into existence when a record is made or received and set aside), necessary (i.e., it exists for every record), and determined (i.e., it is characterized by the purpose of the record).
33 The office (or officer) formally competent for carrying out the action to which the record relates or for the matter to which the record pertains.
34 The office (or officer) given the formal competence for maintaining the authoritative record, that is, the record considered by the creator to be its official record.
35 Annotations are additions made to a record after it has been completed. Therefore, they are not considered elements of the record's documentary form.
36 Technical modifications are any changes in the digital components of the record as defined by the Preservation Task Force. Such modifications would include any changes in the way any elements of the record are digitally encoded and changes in the methods (software) applied to reproduce the record from the stored digital components; that is, any changes that might raise questions as to whether the reproduced record is the same as it would have been before the technical modification. The indication of modifications might refer to additional documentation external to the record that explains in more detail the nature of those modifications.

REQUIREMENT A.4: Protective Procedures: Media and Technology	the creator has established and effectively implemented procedures to guarantee the continuing identity and integrity of records against media deterioration and across technological change;

925

REQUIREMENT A.5: Establishment of Documentary Forms	the creator has established the documentary forms of records associated with each procedure either according to the requirements of the juridical system or those of the creator;

926

REQUIREMENT A.6: Authentication of Records	if authentication is required by the juridical system or the needs of the organization, the creator has established specific rules regarding which records must be authenticated, by whom, and the means of authentication;

927

REQUIREMENT A.7: Identification of Authoritative Record	if multiple copies of the same record exist, the creator has established procedures that identify which record is authoritative;

928

REQUIREMENT A.8: Removal and Transfer of Relevant Documentation	if there is a transition of records from active status to semi-active and inactive status, which involves the removal of records from the electronic system, the creator has established and effectively implemented procedures determining what documentation has to be removed and transferred to the preserver along with the records.

929

930

931 **Appendix 2.**

932

933 **Baseline Requirements Supporting the Production of Authentic Copies of**
934 **Electronic Records37**

935 ***Preamble***

936 The baseline requirements outline the minimum conditions necessary to enable the
937 preserver to attest to the authenticity of copies of inactive electronic records.

938 ***Baseline Requirements (Requirement Set B)***

939

	The preserver should be able to demonstrate that:	
REQUIREMENT B.1: Controls over Records Transfer, Maintenance, and Reproduction	the procedures and system(s) used to transfer records to the archival institution or program; maintain them; and reproduce them embody adequate and effective controls to guarantee the records' identity and integrity, and specifically that	
	B.1.a	Unbroken custody of the records is maintained;
	B.1.b	Security and control procedures are implemented and monitored; and
	B.1.c	The content of the record and any required annotations and elements of documentary form remain unchanged after reproduction.
REQUIREMENT B.2: Documentation of Reproduction Process and its Effects	the activity of reproduction has been documented, and this documentation includes	
	B.2.a	The date of the records' reproduction and the name of the responsible person;
B.2.b		The relationship between the records acquired from the creator and the copies produced by the preserver;
	B.2.c	The impact of the reproduction process on their form, content, accessibility and use; and
	B.2.d	In those cases where a copy of a record is known not to fully and faithfully reproduce the elements expressing its identity and integrity, such information has been documented by the preserver, and this documentation is readily accessible to the user;
REQUIREMENT B.3: Archival Description	the archival description of the fonds containing the electronic records includes—in addition to information about the records' juridical-administrative, provenancial, procedural, and documentary contexts—information about changes the electronic records of the	

37 Available: http://www.interpares.org

creator have undergone since they were first created.

Commentary on the Benchmark Requirements Supporting the Presumption of Authenticity of Electronic Records

The assessment of the authenticity of the creator's records takes place as part of the appraisal process. That process and the role of the benchmark requirements within it are described in more detail in the "Appraisal Task Force Report." This assessment should be verified when the records are transferred to the preserver's custody.

A.1: Expression of Record Attributes and Linkage to Record

The presumption of a record's authenticity is strengthened by knowledge of certain basic facts about it. The attributes identified in this requirement embody those facts. The requirement that the attributes be expressed explicitly and linked inextricably[38] to the record during its life, and carried forward with it over time and space, reflects the task force's belief that such expression and linkage provide a strong foundation on which to establish a record's identity and demonstrate its integrity. The case studies undertaken as part of the work of the task force revealed very little consistency in the way the attributes that specifically establish the identity of a record are captured and expressed from one electronic system to another. In certain systems, some attributes were explicitly mentioned on the face of the record; in others they could be found in a wide range of metadata linked to the record or they were simply implicit in one or more of the record's contexts. In many cases, certain attributes (e.g., the expression of the archival bond) were not captured at all. The task force's concern is that, in the absence of a precise and explicit statement of the basic facts concerning a record's identity and integrity, it will be necessary for the preserver to acquire enormous, and otherwise unnecessary, quantities of data and documentation simply to establish those facts.

The link between the record and the attributes listed in Requirement A.1 is viewed by the task force as a *conceptual* rather than a *physical* one, and the requirement could be satisfied in different ways, depending on the nature of the electronic system in which the record resides. For example, in electronic records management systems, this requirement is usually met through the creation of a record profile.[39] In other types of systems, the requirement could be fulfilled through a topic map. A topic map expresses the characteristics (i.e., *topics*) of subjects (e.g., records or record attributes) and the relationships between and among them.

When a record is exported from the live system, migrated in a system update, or transferred to the preserver, the attributes should be linked to the record and available to the user. When pulling together the data prior to export, the creator should also ensure that the data captured are the right data. For example, in the case of distribution lists, the creator must ensure that if the recipients specified on "List A" were changed at some point in the active life of records, the accurate "List A: Version 1" is exported with the records associated with the first version, and that the second version is sent forward with those records sent to recipients on "List A: Version 2."

A.2 Access Privileges

Defining access privileges means assigning responsibility for the creation, modification, annotation, relocation, and destruction of records on the basis of

---

38 For the purposes of this requirement, inextricable means incapable of being disentangled or untied, and link means a connecting structure.

39 If the attribute values contained in the profile are also expressed independently as entries in a register of all records made or received by the creator, then, in addition to establishing the identity and supporting the inference of the integrity of the record, they would corroborate such identity and strengthen the inference of integrity.

competence, which is the authority and capacity to carry out an administrative action. Implementing access privileges means conferring exclusive capability to exercise such responsibility. In electronic systems, access privileges are usually articulated in tables of user profiles. Effective implementation of access privileges involves the monitoring of access through an audit trail that records every interaction that an officer has with each record (with the possible exception of viewing the record). If the access privileges are not embedded within the electronic system but are based on an external security system (such as the exclusive assignment of keys to a location), the effective implementation of access privileges will involve monitoring the security system.

A.3 Protective Procedures: Loss and Corruption of Records

Procedures to protect records against loss or corruption include: prescribing regular back-up copies of records and their attributes; maintaining a system back-up that includes system programs, operating system files, etc.; maintaining an audit trail of additions and changes to records since the last periodic back-up; ensuring that, following any system failure, the back-up and recovery procedures will automatically guarantee that all complete updates (records and any control information such as indexes required to access the records) contained in the audit trail are reflected in the rebuilt files and also guarantee that any incomplete operation is backed up. The capability should be provided to rebuild forward from any back-up copy, using the back-up copy and all subsequent audit trails.

A.4 Protective Procedures: Media and Technology

Procedures to counteract media fragility and technological obsolescence include: planning upgrades to the organization's technology base; ensuring the ability to retrieve, access, and use stored records when components of the electronic system are changed; refreshing the records by regularly moving them from one storage medium to another; and migrating records from an obsolescent technology to a new technology.

A.5 Establishment of Documentary Forms

The documentary form of a record may be determined in connection to a specific administrative procedure, or in connection to a specific phase(s) within a procedure. The documentary form may be prescribed by business process and work-flow control technology, where each step in an administrative procedure is identified by specific record forms. If a creator customizes a specific application, such as an electronic mail application, to carry certain fields, the customized form becomes, by default, the required documentary form. It is understood that the creator, acting either on the basis of its own needs or the requirements of the juridical system, not an individual officer, establishes the required documentary form(s) of records.

When the creator establishes the documentary form in connection to a procedure, or to specific phases of a procedure, it is understood that this includes the determination of the intrinsic and extrinsic elements of form40 that will allow for the maintenance of the authenticity of the record. Because, generally speaking, that determination will vary from one form of a record to another, and from one creator to another, it is not possible to predetermine or generalize the relevance of specific intrinsic and extrinsic elements of documentary form in relation to authenticity.

A.6 Authentication of Records

In common usage, to authenticate means to prove or serve to prove the authenticity of something. More specifically, the term implies establishing genuineness by adducing legal or official documents or expert opinion. For the purposes of the benchmark requirements, authentication is understood to be a declaration of a

---

40 The extrinsic and intrinsic elements of form are defined and explained in the Authenticity Task Force's *Template for Analysis,* Appendix 1 <j app01>.

record's authenticity at a specific point in time by a juridical person entrusted with the authority to make such declaration. It takes the form of an authoritative statement (which may be in the form of words or symbols) that is added to or inserted in the record attesting that the record is authentic.**41** The requirement may be met by linking the authentication of specific types of records to business procedures and assigning responsibility to a specific office or officer for authentication.

The authentication of copies differs from the validation of the process of reproduction of the digital components of the records. The latter process occurs every time the records of the creator are moved from one medium to another or migrated from one technology to another.

A.7 Identification of Authoritative Record

An authoritative record is a record that is considered by the creator to be its official record and is usually subject to procedural controls that are not required for other copies. The identification of authoritative records corresponds to the designation of an office of primary responsibility as one of the components of a record retention schedule. The Office of Primary Responsibility is the office given the formal competence for maintaining the authoritative (that is, official) records belonging to a given class within an integrated classification scheme and retention schedule. The purpose of designating an Office of Primary Responsibility for each class of record is to reduce duplication and to designate accountability for records.

It is understood that in certain circumstances there may be multiple authoritative copies of records, depending on the purpose for which the record is created.

A.8 Removal and Transfer of Relevant Documentation

This requirement implies that the creator needs to carry forward with the removed records all the information that is necessary to establish the identity and demonstrate the integrity of those records, as well as the information necessary to place the records in their relevant contexts.

Commentary on the Baseline Requirements Supporting the Production of Authentic Copies of Electronic Records

The establishment and implementation of the baseline requirements take place as part of the function of managing preservation. The preservation function and the role of the baseline requirements within it are described in more detail in the "Preservation Task Force Report."

B.1 Controls over Records Transfer, Maintenance, and Reproduction

The controls over the transfer of electronic records to archival custody include establishing, implementing, and monitoring procedures for registering the records' transfer; verifying the authority for transfer; examining the records to determine whether they correspond to the records that are designated in the terms and conditions governing their transfer; and accessioning the records.

As part of the transfer process, the assessment of the authenticity of the creator's records, which has taken place as part of the appraisal process, should be verified. This includes verifying that the attributes relating to the records' identity and integrity have been carried forward with them (Requirement A.1), along with any relevant documentation (Requirement A.8).

The controls over the maintenance of electronic records once they have been transferred to archival custody are similar to several of the ones enumerated in the benchmark requirements. For example, the preserver should establish access

41 The meaning of authentication as it is used by the Authenticity Task Force in this report is broader than its meaning in public key infrastructure (PKI) applications. In such applications, authentication is restricted to proving identity and public key ownership over a communication network.

privileges concerning the access, use, and reproduction of records (Requirement A.2); establish procedures to prevent, discover, and correct loss or corruption of records (Requirement A.3), as well as procedures to guarantee the continuing identity and integrity of records against media deterioration and across technological change (Requirement A.4). Once established, the privileges and procedures should be effectively implemented and regularly monitored. If authentication of the records is required, the preserver should establish specific rules regarding who is authorized to authenticate them and the means of authentication that will be used (Requirement A.6).

The controls over the reproduction of records include establishing, implementing, and monitoring reproduction procedures that are capable of ensuring that the content of the record is not changed in the course of reproduction.

B.2 Documentation of Reproduction Process and its Effects

Documenting the reproduction process and its effects is an essential means of demonstrating that the reproduction process is transparent (i.e., free from pretence or deceit). Such transparency is necessary to the effective fulfillment of the preserver's role as a trusted custodian of the records. Documenting the reproduction process and its effects is also important for the users of records since the history of reproduction is an essential part of the history of the record itself. Documentation of the process and its effects provides users of the records with a critical tool for assessing and interpreting the records.

B.3 Archival Description

Traditionally it has been a function of archival description to authenticate the records and perpetuate their administrative and documentary relationships. With electronic records, this function becomes critical. Once the records no longer exist except as authentic copies, the archival description is the primary source of information about the history of the record, that is, its various reproductions and the changes to the record that have resulted from them. While it is true that the documentation of each reproduction of the record copies**42** may be preserved, the archival description summarizes the history of all the reproductions, thereby obviating the need to preserve all the documentation for each and every reproduction. In this respect, the description constitutes a collective attestation of the authenticity of the records and their relationships in the context of the fonds to which the records belong. This is different from a certificate of authenticity, which attests to the authenticity of individual records. The importance of this collective attestation is that it authenticates and perpetuates the relationships between and among records within the same fonds.

[i] Authenticity Task Force, 'Appendix 2: Requirements for Assessing and Maintaining the Authenticity of Electronic Records', in *The Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project*, InterPARES, September 2002, http://www.interpares.org/book/index.htm.

42 Although, technically, every reproduction of a record that follows its acquisition by the preserver is an authentic copy, it is the only record that exists and, therefore, should normally be referred to as "the record" rather than as "the copy."