

Metadata and the Management of Current Records in Digital Form

Hans Hofman

ICA-committee on electronic and other current records

July 2000

DRAFT

1. Introduction and scope

In the age of digital information the issue of metadata has become more crucial than ever before. Although this is not new as a phenomenon in itself it is new as a term in relation to record keeping and archival management. In the world of paper records archival description (i.e. preparing finding aids at different levels of aggregation) is an important activity. To a certain extent archival description has to do with metadata too. As such it can be characterised as an activity with a retrospective view in the sense that records already exist and are being described when they are managed by an archival institution. The objectives are to improve access, to reflect the arrangement (structure) of the records, and to enable the ongoing understanding of (archival) records.

In a digital world the traditional way of dealing with records may not be sufficient. What has changed? Information and communication technology enables a completely different way of recording and manipulating information. This has several consequences. The most important are:

- Unlike traditional paper records, digital records are no longer fixed physical entities.
- All activities related to carrying out and documenting a business process, including recordkeeping itself, are converging into one virtual world that can be accessed through the desktop (i.e. computer screen).
- This virtual world deals with information of all kinds in order to deliver services or information to users.

These changes caused by IT will also influence record management activities. Some issues in this respect are:

- In order to maintain and manage digital records and to ensure their authenticity and reliability it is necessary to stabilise them as much and as early as possible.
- One way to achieve this is to use metadata that already exist within the

context of the business process. These metadata normally describe the context in which information is created or used and the services that are provided. They contain information about, for instance, the organisation (agent), function or activity, and the transaction. To add these metadata later, for instance when records are captured or even transferred to an archival institution, would be not very cost-effective.

- One aspect needs special attention in a digital environment: that is the linkage between the metadata and the digital objects they describe. They can be easily separated from each other, something which does not happen to the same extent in the paper world. Much of the metadata is stored in different systems (e.g. registry systems, e-mail systems, workflow management systems), while documents are stored in, for instance, document management systems or word-processing systems.

Records managers and archivists are confronted with a changing environment and have to think about the consequences for their management of digital records, especially in relationship with metadata.

Questions in this respect include the following:

- How does metadata compare to archival description?
- To what extent will 'the metadata approach' influence records management or archival activities?
- Which of the existing initiatives in this field are relevant?

This paper will briefly explore these issues by addressing the following: an overview of the different purposes of metadata, an explanation of the relationship between metadata and (digital) records, the identification of initiatives which have as an objective the development of standards and standard approaches to metadata, and a short list of literature and references to information sources.

2. Purpose and Context

The purpose of this short paper is to provide some guidance in understanding what is meant by metadata and in assessing the different activities in this area. It is an introduction and not meant as in-depth study about metadata and records, records management and/or record keeping.

As mentioned previously, the growing use of information and communication technology (ICT) and the subsequent emergence of electronic networked environments change the way we deal with information. Similar to the paper world rules, frameworks and standards are required in order to manage digital information. One of the drivers is the emergence of electronic commerce, including electronic government. Electronic business transactions need a supporting infrastructure, especially for making them reliable. Records created with these transactions document the business done and social activity in general and, as such, they become an important information resource. But the management of such a resource requires an infrastructure based on the application of effective record keeping standards and practices and the use of recordkeeping metadata standards. The

following sections focus on the metadata aspects of this infrastructure.

3. *Records and Metadata*

Metadata are nothing more or less than data about data. Depending on the context in which they are used they provide meaning. For instance information about the author of a letter can tell something about the origin of the communication and the authenticity of the information. It can also be used as an access point for retrieval. As a result, metadata can serve different purposes such as understandability, authenticity, reliability, retrievability, maintenance, management, and appraisal.

Documents are created to play a role within the process of conducting a task (business process) undertaken by an organisation or individual. Documents and their supporting metadata permit the business process to be carried out. These metadata can address factors such as origin, time, place, subject, addressee etc. In a paper world this type of metadata can be found mostly on the document or the record itself. In a digital environment these metadata are managed and stored mostly in other systems, e.g. a workflow management system, a registry system or an email system.

If they are determined to be valuable, documents are captured, stored as records, and kept for as long as necessary. Metadata are gathered during the capture and record keeping processes and allow people to establish the provenance and the authenticity of the records, especially in the case of digital records. Metadata address the content, context, form and structure of a digital record.

When records and their metadata are captured they need to be maintained and managed. This is a process in itself (the record keeping function) and creates its own metadata, that tell us what has happened during the existence of the records after their capture (e.g. when they were captured, the interrelationships among records (arrangement based on the business process), what has been done in order to maintain them, and in some cases when they were destroyed).

Finally metadata reflecting further explanatory descriptions or discovery elements can be attributed to the records in order to improve access and retrievability.¹ Archival description can also add additional information about the origin of records and the coherence between the records. This becomes necessary when they cross time and/or domain boundaries where tacit or implicit knowledge has to be made explicit.

Very briefly, this is the universe of metadata in connection with records. Two perspectives of metadata are apparent here:

- (1) metadata that position the document or record as part of a business process (both communicated and used) and
- (2) metadata that show the management of the (captured and stored) record.

It is important to make this distinction between the record(s) as created and taking part in the business process and records as object of records

management, because different sets of metadata are required. Records managers deal with the object (i.e. the record), including metadata about its provenance and its management and maintenance as an object.

Secondly it is important to be aware of the difference or distinction between the intellectual or conceptual record and the physical computer file. This is the consequence of the fact that records are no longer fixed physical entities. The intellectual or conceptual record is the record that human beings can read and understand and that can be seen only on a computer screen. The computer file is a physical component in which the record (to a certain extent) is stored in an encoded format. Records could also be stored in more than one computer file. For instance, when a record consists of text and images there will be more than one computer file. As such these computer files have to be read and interpreted by application software in order to recreate the conceptual record. As a result, while both the intellectual or conceptual record and the related computer file or files require their own management, they are or should be closely interrelated. Without computer files (and the accompanying software application) there will be no records.

4. Different purposes

Depending on the perspective taken, different sets of metadata for records are needed. Based on the previous section it is necessary to distinguish between the following three areas: creation and communication, use (all of which focus on the object), and management.

This distinction does not suggest that metadata created in one area could not be used in another area. For instance the metadata created within the business process about the business activity and the organisation responsible for creating the records, could be used in the area of “use” for discovery purposes or in the area of “management”, especially as this relates to intellectual control. It is important to be aware of the fact that metadata created for one purpose can be used for another.

4.1 Creation and communication

Records are created within a business process to enable and to document business activity. As such they are accompanied by metadata to clarify what they are about (subject) and to enable their role. These metadata provide information about the origin or destination, time and place, authentication, and the reliability of the records. As such records are not only by-products of a business process (i.e. serving the purpose of memory), they also provide the triggers that support the conduct of the business process itself. The accompanying metadata are sometimes part of the document itself and sometimes separate (e.g. the e-mail header). This type of metadata, which could be called “business process-metadata”, forms part of the contextual information for the record.

4.2 Use, information resource discovery, and access

Records are created in order to be used, in first instance within the business

process itself (to get things done and as memory), but also for accountability reasons and later in time for research.

If the perspective of using records is taken, the focus will be on the records themselves and the information they contain. Metadata are needed that provide the user (both within the business process itself and for later research purposes) with information on:

- the structure, as one of the elements for establishing the authenticity
- the context, to allow us to understand the provenance and on
- the subject or content, to know what the record is about.

This information helps us to position, to understand, and to interpret the record or set of records. To give an example, if a record is represented on the screen a researcher needs to know in what business process it has been created, by what organisation and by whom, in what case (the subject), and when. It must be possible to check whether the structure (and lay-out) of the record as shown is the original structure (and lay-out). Furthermore it is or can be relevant to know the interrelationship between records, because that is part of the context too and helps the users to understand the record(s) better.

When a record has been retrieved, information is required about the management of the record (or about its custodial history) in order to assess its authenticity. Apart from that information is required about the events that occurred since the creation and capture of the records.

4.3 Management and maintenance

With respect to the management and maintenance of the records information is needed not only about the activities already carried out with the records, but also about the records themselves. In the case of the latter, metadata is required about the state of the record, its location, when it was used, etc.

The ultimate purpose of “management” is to maintain the records in an authentic and accessible (usable) manner.

There are four domains or perspectives under “management”:

- intellectual control,
- administrative control,
- physical or more precisely the technical control of the records, and
- control of the recordkeeping or records management regime itself.

Intellectual control

In the area of intellectual control the role of records managers and archivists is to ensure, maintain, and if necessary, enhance the understandability and accessibility of authentic records through time.

The activities associated with the intellectual control of records include the following:

- o capture and maintain the authenticity (in a paper world through preserving original paper records)
- o guarantee understandability and to ensure provenance (e.g. archival

- description of paper records)
- o enable and carry out (functional) appraisal and disposal appropriately and in time
- o realise, maintain, and improve accessibility to support use (e.g. management, archival description, and preservation).

Administrative control

Related to but also distinct from intellectual control is administrative control. For administrative control, information is required about:

- the state of the records,
- when they are captured or came under custody,
- when are they stored
- what activities are or were carried out with them in order to keep them authentic, usable and understandable, and
- their history of use

Location management (e.g. where are they stored) and logistics, security, back-up, and recovery activities belong to this domain as do issues concerning use rights and copyright.

Technical control

The third domain of management is technical control, which has to do with preservation and maintenance. For preservation reasons we need technical information about digital records such as:

- what is their storage format?
- with what application software and hardware were they originally created?
- with what application software can they be presented again?

Furthermore it is obvious that digital records have to be maintained. New generations of IT will make the technology with which records were once created obsolete and inhibit their access and use through time. The rapid developments in IT make this even more complicated. As a result, there must be a track record of what strategy or strategies have been adopted and within this framework what activities have been carried out to maintain them in a readable, accessible, and authentic manner within this dynamic environment. Depending upon the maintenance strategy adopted, information about the result of the maintenance activities has to be kept (e.g. whether information has been lost during the migration of records).

Record keeping or records management control

The last but not least domain is the recordkeeping or records management regime that controls the recordkeeping function or system. It concerns the rules and procedures that are in place and the changes in this regime that occur during the existence of the recordkeeping system. As a result, the record keeping function or system has to be described as well, because it is part of the context of the record or records in the system. In ideal circumstances, the record keeping system has to describe itself.

Questions that need to be answered include:

- What record keeping regime was valid and during what period?
- What were the recordkeeping requirements during that period?

- What were the rules and procedures for recordkeeping?

In a paper world archivists are used to retrospectively describing the recordkeeping/ - management regime that existed. In a digital world this information is already available, for the most part, in a recordkeeping system, if such a system exists.

5 Existing initiatives

There are many initiatives and activities in the area of metadata, not only in the records management or archival community, but also in other communities such as scientific data libraries, the audio-visual community, and those involved in geographic information systems (GIS) and information resource discovery. They are all trying to identify what metadata are necessary to manage, describe, understand, discover, retrieve, search, and preserve digital information. Furthermore they are all focusing on some kind of standardisation be it on a set of required metadata for a specific purpose (e.g. recordkeeping, discovery, preservation) or on specific aspects such as the data structure, the data content or the data value.

These initiatives can be distinguished between those based on a record oriented, retrospective approach and those based on a business process oriented approach. The primary characteristic of the more traditional record oriented approach is that it takes a retrospective point of view and, in a paper environment, tends to focus on physical objects. It presumes the records already exist and that they proceed through different life-stages. As such they need further processing (e.g. describing, arranging, disposal, etc.). That approach is valid in the physical world of paper records, but may not be appropriate for digital, virtual records, because they require immediate management in order not to be lost. Furthermore in a digital world most of the information that is needed for the above mentioned purposes is already available. What is required is to identify and capture this information.

The business oriented approach takes the business process as a starting point for determining metadata requirements and emphasises the strong interrelationship between records, the business process in which they are created, and the management of the records. The approach tends to see the existence and use of records as a 'continuum', in which the record can serve different purposes at different moments in time and space.

Record oriented, retrospective approach

Examples of initiatives are as follows:

- International Standard on Archival Description, Generic ([ISAD/G](#))
- International Standard on Archival Authority Records for Corporate Bodies, Persons and Families ([ISAAR/CPF](#))

ISAD/G allows a description at different levels of aggregation. This standard is currently under review to adapt it better to the needs of a digital environment. Both standards define structure and data content for archival systems. This (informal) standard is not focusing on records within government organisations, but on archival records within the

custody of archival institutions. As such it is not a recordkeeping metadata standard, but it makes description of these records and their origin possible after they have been appraised and transferred to an archival institution.

Finally in the area of presenting archival information (or description) on the web there is the HTML based initiative, Encoded Archival Description ([EAD, Library of Congress](#)) which is being proposed as an ISO standard. Used mainly in North America, it is an exchange format for placing finding aids on the web.

- University of British Columbia (UBC) project (now succeeded by the [Inter Pares project](#)).
Based on the discipline of diplomatics the UBC project has defined the requirements and metadata necessary to ensure the authenticity and reliability of records. The set, structured into a template consisting of 26 elements, describes the components of a record. These components are the medium, the content, the intellectual and physical form of a record, the persons (roles) involved and the acts (activities/transactions).

Business process oriented approach

The other category is the business process oriented approach which embraces the following informal standards in the area of recordkeeping and archival management:

- the standardised set of Recordkeeping Metadata of the SPIRT-project, the [Records Continuum Research Group, Monash University](#);
Based on the continuum model, the SPIRT project is an effort to formulate a standard for recordkeeping metadata. It provides a high level set of metadata describing Agents, Business (activities, including recordkeeping) and Records and their interrelationships throughout the continuum.
The project has compared this model with other existing sets of metadata, such as BAC, AGLS, Dublin Core, and UBC.
- the [recordkeeping metadata standard of the National Archives of Australia](#)
The National Archives of Australia has issued this standard to encourage government organisations to apply a consistent approach to records identification and description. It has a strong link to the SPIRT model of recordkeeping metadata. This Metadata Standard is part of a coherent set of recordkeeping guidelines, policies, standards, manuals etc. supporting electronic government.
- metadata for [Business Acceptable Communications \(BAC\)](#), Pittsburg project on metadata for reliable and authentic (archival) records.
The Pittsburgh project has defined a 6-layer model of metadata that enables compliance to the requirements of business acceptable communications (transactions), and also to satisfy the requirements of evidence and effective and continuing management. The six layers are: handle layer, terms & conditions layer, structural layer, contextual layer, content layer, and use history layer. The model proposes that records be encapsulated within the context of these layers.

These initiatives attempt to integrate all aspects of recordkeeping including the archival aspects (although the UBC project is mainly focusing on archival management).

Information Resource Discovery

Another perspective is the field of information discovery on the web. The purpose is to enable searching and retrieval of information resources on the world wide web. The most important initiatives here are:

- [Dublin Core](#), and
- Government Locator Systems: such as the Australian GLS ([AGLS](#)), together with the Australian Government's Interactive Functions Thesaurus ([AGIFT](#)), that enables end users to search in natural language for government information by linking this to terms established in the Australian government. It is an elaboration on the 'function' term in AGLS.

Libraries are especially active in this field. Interesting projects in this area are:

- the PANDORA Logical Data Model of the [National Library of Australia](#)
- the activities of the [US Research Libraries Group \(RLG\)](#)
- the activities of the [Council on Library and Information Resources](#).

More generic standards for metadata (structures) are being proposed. One that should be of interest to the record keeping and archival communities is the reference model of the [Open Archival Information System \(OAIS\)](#) that is now in the process of becoming an ISO standard. This reference model identifies, describes and structures functions, processes, and sets of metadata that are needed for (long term) preservation. It is not aimed at preserving records in particular, but at preserving digital information in general. Nevertheless it provides a very useful framework for (archival) records.²

In order to fulfil their role properly in a world of digital information metadata about recordkeeping cannot exist on their own. The characteristics of digital information, especially on the world wide web, are that such information be transparent, omnipresent, virtual, timeless, and without boundaries. As a result, recordkeeping metadata needs to be interoperable with other sets of metadata that exist in other communities, especially if communication and discovery are to be made possible (e.g. metadata to allow information resource discovery (e.g. Dublin Core, AGLS), metadata for managing cultural heritage, multimedia, etc.)

One of the main challenges is to identify the boundaries between the domain of recordkeeping and the related domains described earlier in this paper. This is important for two reasons:

- to learn what metadata already exist and can be derived from recordkeeping purposes
- to connect to initiatives in the field of information resource discovery or other relevant communities in order to integrate information about records as much as possible in search engines in order to make the records visible and retrievable.

Based on that knowledge it should be possible to identify what needs to be done inside the recordkeeping domain and to what extent it is possible to establish interfaces with other communities or initiatives, all with the intent to achieve interoperability.

Appendix A

In the following two charts an attempt is made to position the different main, proposed, and informal standards that exist. The first chart takes the domains of control as a perspective, the other the elements of a record.

Table 1 Main proposed standards in relation with the different domains of control

	Intellectual control	Technical control	Administrative control	Access / Information discovery
SPIRT	Provides standard elements for describing agents (organisations) and business processes (functions, activities); recognises the necessity to describe the logical state of records	Identifies the technical origin of records	Provides standard-elements for describing recordkeeping activities and recordkeeping rules/ regime	Yes
Dublin Core	Identifies 15 elements to describe information resources on the web; as such it enables to a certain extent intellectual control	--	--	Yes
GLS AGLS	Provides a structure for describing government information sources	--	--	Yes
BAC	Included in contextual and	Included in structural	Included in terms & conditions and	Included in handle and

	content layers	layer	use history layers	terms & conditions layers
UBC	To a certain extent in Intellectual form, content, persons, acts	Technical form, medium	Partly in Annotations (in intellectual form)?	Possible through intellectual form (contextual information), content, persons, and acts
ISAD/G	Provides elements in Identity Statement Area and Context Area	Very partially in Conditions of Access and Use Area (not yet for digital records)	Contains elements in Content and Structure Area and Conditions of Access and Use Area that allow administrative control	Context Area

Table 2 Proposed/informal standards in relation with the elements of a record

	Structure	Content	Presentation/Form	Contextual information
EAD	--	--	Offers a layout and exchange-format for presenting finding aids on the web	--
ISAD/G	--	Provides elements for describing content	--	Provides elements to describe 'Context area'
ISAAR/CPF	--	--	--	Provides elements for describing context, i.e. authority files concerning persons,

				families and organisations
XML/XMS	Provides a syntax to describe the structure of documents and other information sources. For each type of record/document, including databases, a DTD or Schema can be made for that purpose.	--	In a style sheet (XMS) the look-and-feel can be described	--
Dublin Core, GLS/AGLS	--	Provides an indication of the content through 15 or more (AGLS) elements	--	--
AGIFT	--	A thesaurus of government functions (specification of 'function'-element of AGLS)	--	Offers a controlled list of terms for describing government activities/functions
UBC	Partly in Intellectual form	Content (element)	Extrinsic elements of Intellectual form	Persons, acts

Appendix B

Further Reading

MacNeil, H., *Metadata Strategies and Archival Description: comparing Apples to Oranges*, in: *Archivaria*, Vol 39, 1995, p.22-32

Duff, W., *Will Metadata replace Archival Description: a Commentary*, in: *Archivaria*, Vol 39, 1995, p.33-38.

Bearman, D., *Towards a Reference Model for Business Acceptable Communications*, 1994

Bearman, D., Miller, E., Rust, G., Trant, J., & Weibel, S. (1999), [A common Model to Support Interoperable Metadata: Progress Report on reconciling metadata requirements from the Dublin Core and INDECS/DOI Communities](#), D-Lib Magazine

Bearman D., and Sochats, K., [Metadata Requirements for Evidence](#), paper published in 1999 (?) on internet.

Cunningham, A., *Dynamic Descriptions: Australian Strategies for the Intellectual Control of Records and Recordkeeping Systems*, in: P.J. Horsman, F.C.J. Ketelaar en T.P.H.M. Thomassen (red.), *Naar een nieuw Paradigma in de Archivistiek*, Jaarboek, The Hague 1999, p.133-144.

McKemmish, S., Acland, G., Reed, B., *Towards a Framework for Standardising Recordkeeping Metadata: The Australian Recordkeeping Metadata Schema*, in: *Records Management Journal* 9, no. 3 (December 1999), p.177- 202.

McKemmish, S., [Accessing Essential Evidence on the Web: Towards an Australian Recordkeeping Metadata Standard](#), paper published in 1999 on internet.

McKemmish, S., Acland, G., Ward, N., Reed, B., *Describing Records in Context in the Continuum: the Australian Recordkeeping Metadata Schema*, (to be published in 'Archivaria' no. 48 Fall 1999)

Reed, B., [Metadata: Core Record or Core Business?](#), paper published in 1999 on internet.

Wallace, W., *Managing the Present: Metadata as Archival Description*, in: *Archivaria*, Vol. 39, 1995, p.11-21

Footnotes:

1. The 15 elements of Dublin Core can be considered as such. It is questionable whether keywords should be considered as metadata. It is also doubtful whether adding keywords is part of records management or record keeping.

2. Another initiative is the Metadata Coalition Open Information Model (MDCOIM), that offers a generic framework for structuring metadata and is not focused on records or libraries.