

# 1 Building a metadata schema – where to start<sup>1</sup>

## 1.1 Introduction

### Purpose

Metadata has been defined as “*data describing the context, content and structure of records and their management through time*”<sup>2</sup>. It is an inextricable part of managing records in any format. The use of metadata supports methods to identify, authenticate, describe, locate and manage resources in a precise and consistence way that meets business, accountability, and archival requirements.

The key question when implementing a metadata initiative is this: “Is it necessary to create a new metadata schema, or are there already existing metadata schemas which can be adapted for use?” In general, the fewer metadata schemas, the better. We use standards to improve interoperability and to reduce unnecessary variation. It is better and easier to adopt something that already exists, is well modelled, and comprehensively supported. If you build one, then you will also have manage and support it for the life span of the records. This includes updates, backwards and forwards compatibility, metadata about the metadata schema, registry and other infrastructure to support its implementation, etc.

The purposes of this document are to help the reader to decide whether to build or adapt a metadata schema and to provide some advice on implementation.

This document relates directly to:

1. *ISO 23081-1:2006 Information and documentation - Records management processes - Metadata for records - Part 1: Principles*
2. *ISO/TS 23081-2:2007 Information and documentation - Records management processes - Metadata for records - Part 2: Conceptual and implementation issues*

This document also relates indirectly through the above standards to the business needs and records requirements in:

3. *ISO 15489-1:2001 Information and documentation - Records management -- Part 1: General*
4. *ISO/TR 26122-1:2008, Information and documentation - Work process analysis for records*

### Intended audience

The intended audience is the person or group in an organization tasked with creating a formal metadata structure, even though they may have little experience in this type of work.

### Outcomes

Possible outcomes are:

1. Understanding whether to create a new schema or to adapt an existing one
2. Understanding how to get started and key points for compliance

---

<sup>1</sup> A companion paper is under development – *Suggestions for implementing a Metadata Schema or Application Profile, into an Electronic Document and Records Management System (EDRMS)*

<sup>2</sup> ISO 15489-1 s 3.12

## 1.2 Why have you decided to use or create a metadata structure for records?

- “I have been told I need one”
- The organization is implementing software to manage its documents, e.g. an Electronic Document and Records Management System (EDRMS)
- The organization is trying to standardize descriptions across document types, documents created by different groups, databases, websites etc
- The organization wants to improve retrieval of information
- The organization wants to improve the sharing of information
- The organization wants to ensure interoperability across systems
- The organization wants to ensure the preservation of its information over time
- The organization is tasked with/needs to improve its archival descriptions
- The organization wants to demonstrate compliance with standards e.g. for records management
- Some or all of the above

## 1.3 Key concepts

When undertaking a metadata initiative it is important to understand the differences between (and associated benefits of using) ISO 23081 - the international standard on metadata for records, a schema for metadata, an application profile, and an encoding scheme.

- Metadata standard. A high level document which includes principles and implementation issues
- Metadata schema. This document uses “schema” in same way as ISO 23081. “A *schema is a logical plan showing the relationships between metadata elements, normally through establishing rules for the use and management of metadata specifically as regards the semantics, the syntax and the optionality (obligation level) of values.*”<sup>3</sup> Also referred to as an element set.
- Application profile. “An *application profile delineates the use of metadata elements declared in an element set. While an element set establishes concepts, as expressed via metadata elements, and focuses on the semantics or meanings of those elements, an application profile goes further and adds business rules and guidelines on the use of the elements. It identifies element obligations and constraints, and provides comments and examples to assist in the understanding of the elements. Application profiles may include elements integrated from one or more element sets thus allowing a given application to meet its functional requirements.*”<sup>4</sup>
- Encoding scheme<sup>5</sup>. “*Controlled list of all the acceptable values in natural language and/or as a syntax-encoded text string designed for machine processing.*”<sup>6</sup> Includes rules/formats for entering data such as dates, names of people, etc.

<sup>3</sup> ISO 23081.1 s3 Terms and Definitions

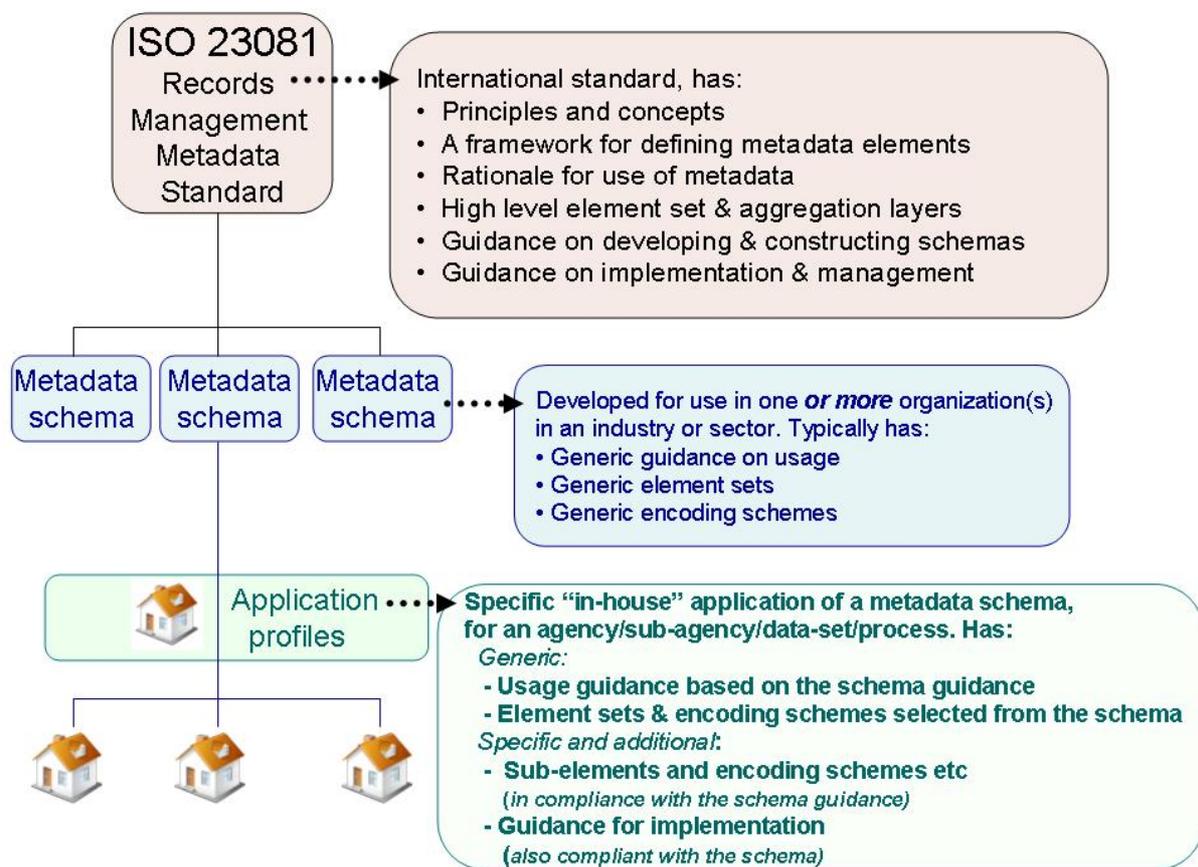
<sup>4</sup> GC RMAP - Government of Canada Records Management Application Profile. S1.3 *What is an Application Profile?*

<sup>5</sup> Note that this definition has a different emphasis to that used in the Dublin Core/resource discovery community. For that community an application profile is the way someone (it could be anyone or any organisation) sets out their **conceptual** view of their use of metadata properties, what vocabularies to use, etc. Although it is about a specific application of metadata for a particular purpose, is still primarily a conceptualisation of metadata use. Once the application profile is done, it is possible to develop a machine-readable schema, which is merely a way of expressing the application profile in a way that is useful at the implementation level.

In addition, there are concepts about the relationships between metadata standards:

- Crosswalk. “A specification for mapping one metadata standard to another.”<sup>7</sup> Crosswalks can also occur between schemas and application profiles.
- Harmonization. “The process of enabling consistency across metadata standards. Harmonization of metadata standards is essential to the successful development of crosswalks between metadata standards. Harmonization results in the ability to create and maintain only one set of metadata, and to map the metadata to any number of related metadata standards. The use of harmonization vastly simplifies the development, implementation and deployment of related metadata standards through the use of common terminology, methods and processes”<sup>8</sup>

The diagram below shows the relationship between *ISO 23081* the Records Management Metadata Standard, metadata schemas and application profiles.



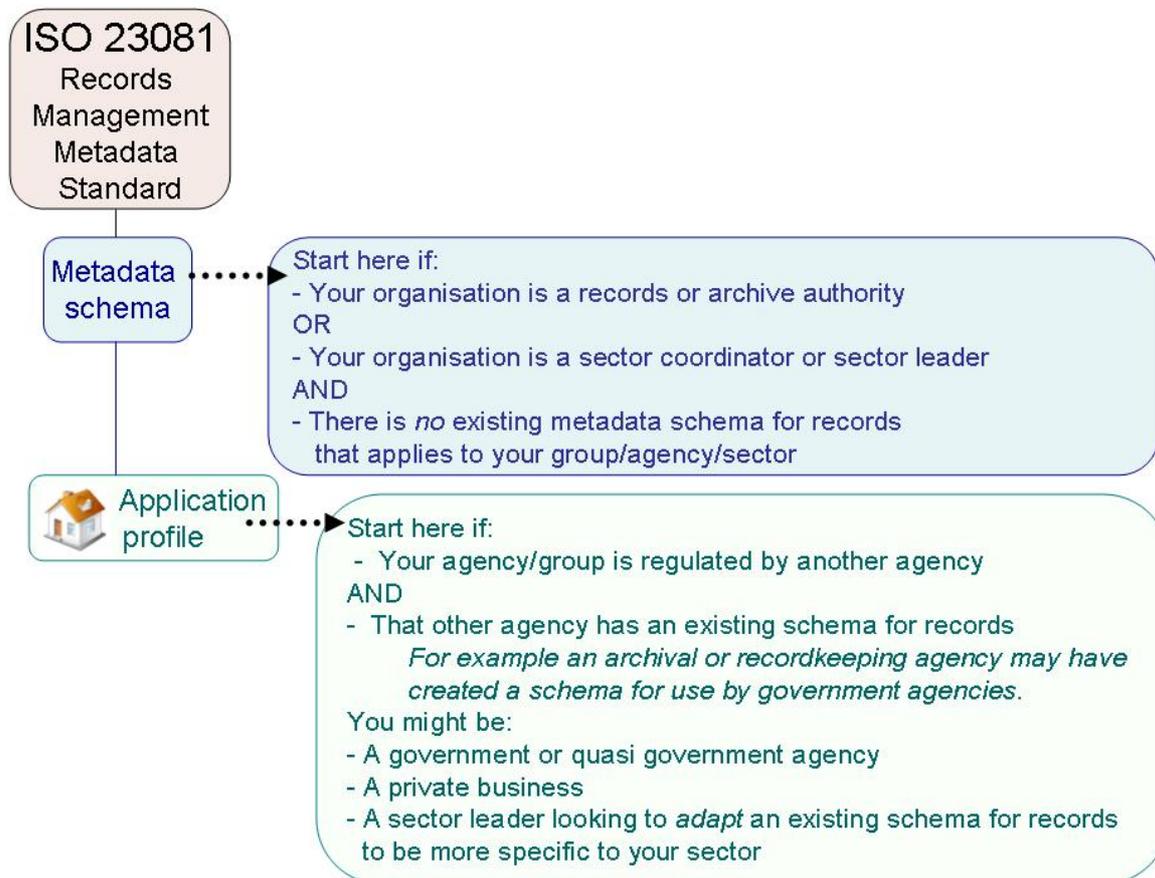
<sup>6</sup> ISO 23081.1 s3 Terms and Definitions

<sup>7</sup> ISO/TC171/SC2 N 471 Document management – Guidelines for the creation of a metadata crosswalk S3 Terms and definitions

<sup>8</sup> *ibid*

## 1.4 Should I start by building a metadata schema or an application profile?

The diagram below shows whether to start - either by creating a metadata schema, or by modifying an existing metadata schema to create an application profile.



1. We strongly urge you to read this document in conjunction with ISO 23081-1 and ISO/TS 23081-2
2. The second step is to discover and then analyse any existing relevant schemas to see if any can be implemented without further change. See *Appendix A - What help is available?* below.
3. It is very probable that any existing schema will need specific changes (and therefore the creation of an application profile) for your organization. Typical changes are:
  - Encoding schemes specific to your organization. Examples: rules on how to enter dates consistently, lists of office locations, activities/services, roles/people
  - Inclusion of refinements (sub-elements) specific to your organization. For example,
    - The element of *Coverage* could have a refinement (sub element) of *Jurisdictional coverage* – a way to provide information on the territorial regions/branches within an organisation
      - *Jurisdictional coverage* for an education organization could have an encoding scheme of *Education District*. This would provide a controlled list of the current territorial school districts

- *Jurisdictional coverage* for a firefighting organization could have an encoding scheme of *Fire District*. This would provide a controlled list of the current territorial districts established to respond to fires
    - A *Language* element could have:
      - A *refinement* of *Dialect*
      - An *encoding scheme* (controlled list) of dialects.
4. If possible, do not introduce any new elements, since this reduces inter-operability between application profiles. See *Application profile - Suggestions for implementing a schema to create an application profile*, below. When creating an application profile from a metadata schema, **most** changes should be in creating:
- Specific refinements (sub-elements)
  - Specific encoding schemes e.g controlled lists of terms, rules for how to enter names, dates etc)

## 2 Metadata schema - suggestions for getting started

### 2.1 Determine the scope of the schema

See *ISO/TS 23081-2 Section 4 Purpose and benefits of metadata*

Identify:

- How the schema relates to other integration/interoperability initiatives in your group/agency/sector
- Which information objects/processes you are going to describe, e.g. text-based documents, images, spatial objects etc.
- Will the schema be used to describe objects in a document management system, website, archival system, business information system/transactional database?
- How the metadata will be used, so that only useful metadata is captured  
See *ISO/TS 23081-2 Section 11.2 Storage and management*
- To whom the schema will apply, e.g. your group, your agency, your sector

### 2.2 Study ISO 23081-1 and ISO/TS 23081-2

- *ISO 23081-1* for:
  - The principles behind schemas
  - The purposes of metadata
- *ISO/TS 23081-2* for:
  - How schemas are constructed and maintained
  - Suggestions for elements and aggregations

### 2.3 Study other existing schemas and contact their creators

See *ISO/TS 23081-2 Section 10, Developing a metadata schema for managing records*

- Look to peer authorities e.g. other recordkeeping/archival agencies. Look to equivalent sector leaders in other regions of your country or in other countries, e.g. if your organization works in education, then look to other leading education agencies within your country, and look at the work done by leading education agencies in other countries.
- Look for useful metadata models
  - In similar sectors/agencies in other countries/jurisdictions
  - Designed to do a similar task of coordinating metadata collection, e.g. on-line learning materials, procedure manuals, consultation records, regulatory processes

### 2.4 Determine the structure of the schema

A key decision is whether to have one set of metadata elements (single entity), or whether to establish groups of metadata elements (multiple entities). Multiple entity models have the advantage of grouping elements around what you are trying to describe, e.g. are you describing the Business (organization), Agents (people and roles), Records, Mandate (authority for agents and records), or Relationships.  
See *ISO/TS 23081-2, Section 6, Metadata conceptual model*

Metadata schemas can provide different views of the metadata elements, e.g. when describing a record, you could provide:

- A single entity model (just Record elements), or
- Two entities (Record and Agent) with a separate entity for agents, or
- Three entities (Record, Agent, and Relationship) where relationship is used as an entity to provide the linkages between records and agents.

## **2.5 Register the schema with relevant agencies**

See *ISO/TS 23081-2 Section 10.2, Metadata registries*

## **2.6 Identify useful element sets and encoding schemes**

See *ISO/TS 23081-2 Section 9 Generic metadata elements* and *Section 10.3 Designing metadata schema for managing records*

Identify, from both ISO 23081 and other schemas:

- What **existing** elements, groupings of elements and sub-elements from other schemas can be used in the schema and linked back to their source (cross-walking)
- What, if any, **new** elements or sub-elements are needed
- What existing elements or sub elements are not needed
- Specific encoding schemes to use in this environment  
See *ISO/TS 23081-2 Section 10.3.3 Encoding schemes*
- Useful generic guidance that can be adapted
- What substitutions/changes in use are consistent with /undermine any source schema(s) you have used
- OR  
the reasons for your deviation from the source schema(s)

## 3 Application profile - Suggestions for implementing a schema to create an application profile

### 3.1 Study the existing schema(s)

- The guidance on how to use the existing schema(s)
- Review the structure of the schema – e.g. single entity or multiple entities and whether this structure should be modified to suit your purposes. If changing the structure, e.g. from a multiple entity to a single entity structure, ensure that this is within the schema guidance. See *Section 2.4 Determine the structure of the schema*, above, and see ISO/TS 23081-2, *Section 6, Metadata conceptual model*
- Consult with peers on the selection and use of elements, refinements (sub-elements) and encoding schemes

### 3.2 Study ISO 23081

- ISO 23081-1 to gain an understanding of the principles behind schemas
- ISO/TS 23081-2 to gain an understanding of how schemas are constructed and maintained

### 3.3 Determine the scope of the application profile

- How the application profile relates to other integration/interoperability initiatives in your group/agency
- Which information objects./processes you are going to describe, e.g. text-based documents, images, spatial objects
- Whether/how the metadata will actually be used, so that only useful metadata is captured
- In which systems the metadata will be used, e.g. document management system, website, archival system, a business information system/transactional database
- To whom the application profile will apply, e.g. your group, your agency

### 3.4 Register the application profile with relevant agencies

See *ISO/TS 23081-2, Section 10.2, Metadata registries*

### 3.5 Identify:

- Any changes or additions/selections you need to make
- New refinements (sub-elements) needed, e.g. for an education agency, *Student* might be a sub-element of *Agent*; *School* might be a sub-element of *Jurisdiction*. Adopt existing sub-elements from other schemas that are well maintained.
  - Any elements or sub-elements not needed, e.g. *Audience* might be an optional element in the schema, but not needed in your application profile
- Which elements are mandatory, which are recommended, optional, etc. If an element is mandatory in the parent schema, then it should be retained, and its mandatory status retained. See *ISO/TS 23081-2, Section 10.3.4, Rules for syntax, obligation levels, default values and repeatability*
- Specific encoding schemes to use in this environment, e.g. a list of district offices for a specific organization, a list of activities specific to an organization
- Which substitutions or changes in use are consistent with the source schema and which undermine the source schema. This is very important since inconsistency of usage is the cause of most problems

### 3.6 Minimize the use of “free text” entry

“Free text” is where the user can input text, free of any control over format, content etc. For example, a Description element typically permits “free text” input. The main benefit of a “free text” element is that it provides a place for users to add extra information that does not fit into other elements. The problems with free text are significant, and can include:

- Information is often unusable by automated systems
- Variation in spelling
- Variation in use of abbreviations, formats for dates, etc
- Users may avoid filling out other elements, and instead put unstructured information into free text fields
- Increased operational costs incurred by your organisation to access and retrieve records

### 3.7 Some smart things to do when developing either a metadata schema or an application profile – “Crosswalks” and “harmonization”

- Schemas are usually built for a specific purpose, e.g. discovery, records management preservation, etc. Check for missing elements. There may be several discovery-focused elements (subject, description, title), but are there enough elements for records management, (business function, agent, storage format, ownership, disposal actions and triggers)?
- Include useful elements from other well-maintained schemas, e.g. for geospatial coordinates include elements from geospatial metadata schemas. This is called “cross walking”. See 1.3 Key concepts, above. Remember that if those elements change then your schema must adapt as well
- When combining elements and refinements (sub-elements) from a variety of schemas, make sure they do not overlap. Determine which element set is better at describing formats, which is better at description of content, etc. and select the appropriate elements. Check whether the way you want to use any element or sub-element is consistent both with the source schema and with your purposes. For example a simple *Date* element would not comply with the records management requirement that specific types of dates must be linked to events such as disposal actions
- Link to existing encoding schemes that are well maintained by trusted agencies such as ISO, IEC<sup>9</sup>, ITU<sup>10</sup>, W3C<sup>11</sup>.
- Here are some important ISO encoding schemes. Find them at <http://www.iso.org/iso/store.htm>.
  - *ISO 3166-1:2006, Codes for the representation of names of countries and their subdivisions – Part 1: Country codes.* (Part 2 has codes for subdivisions within countries and Part 3 has codes for formerly used names of countries).
  - *ISO 8601:2004, Data elements and interchange formats – Information interchange – Representation of dates and times.* Also, for date ranges, see *RKMS-ISO8601 Recordkeeping Metadata Schema Extension to ISO 8601*<sup>12</sup>
  - *ISO 19115:2003, Geographic information – Metadata.* This is for spatial descriptors. Look also for more specific/local schemas/application profiles based on this important standard

<sup>9</sup> The International Electrotechnical Commission (IEC) prepares and publishes international standards for all electrical, electronic and related technologies. It cooperates with ISO and ITU to publish joint standards.

<sup>10</sup> International Telecommunication Union (ITU)

<sup>11</sup> The World Wide Web Consortium (W3C) develops interoperable technologies (specifications, guidelines, software, and tools for the Web

<sup>12</sup> <http://www.sims.monash.edu.au/research/rcrg/research/spirt/deliver/schemes.html#rkmsextension>

- Also look for any existing lists relevant to your country or sector, for example:
  - For your country there may be a list of security classifications, e.g. In confidence, Restricted, Secret, Top Secret etc
  - For the education sector there might already be a list of education functions and activities
- Avoid creating new elements. Create refinements (sub-elements) instead. This is fundamental for interoperability. Even if systems don't recognise the refinement, they will recognise the parent element. For example, in an education sector metadata schema, under a “Subject” element there could be a refinement of “Education Curriculum”  
See *ISO 23081- Section 4.2.3, Interoperability*
- Make sure that any use of external encoding schemes is consistent. See
  - *ISO 23081 Section 10.3.3, Encoding schemes*
  - *ISO 23081 Section 10.3.4, Rules for syntax, obligation levels, default values and repeatability*
  - *ISO/TC171/SC2 N 471, Document management – Guidelines for the creation of a metadata crosswalk*

## 4 Summary

Whether creating your own metadata schema or creating an application profile from an existing metadata schema, it is important to:

- Understand ISO 23081-1 and ISO/TS 23081-2
- Research to find existing relevant metadata schemas and application profiles, including:
  - Does the purpose behind each metadata schema/application profile match your needs?
  - What changes you can make that still conform to rules/guidance in the schema(s)

## Appendix A - What help is available?

### Generic

ID/Name	Context	Link
ISO 23081-1 ISO/TS 23081-2	Generic guidance and are an essential starting point	<a href="http://www.iso.org/iso/store.htm">http://www.iso.org/iso/store.htm</a>
ISO/TC171/SC2 N 471 <i>Document management – Guidelines for the creation of a metadata crosswalk</i>	Metadata crosswalks	<a href="http://www.iso.org/iso/standards_development/technical_committees/list_of_iso_technical_committees/iso_technical_committee.htm?commid=53650">http://www.iso.org/iso/standards_development/technical_committees/list_of_iso_technical_committees/iso_technical_committee.htm?commid=53650</a>
ICA <sup>13</sup> “ <i>Metadata and the Management of Current Records in Digital Form</i> ”-	Overview/guidance	<a href="http://www.ica.org/en/node/30038">http://www.ica.org/en/node/30038</a>

### Metadata schemas and application profiles

Individual metadata schemas and any accompanying application profiles have been developed by metadata and subject experts who have experience in describing the information objects created in their sectors and jurisdictions, e.g. documents, images, web pages, voice recordings, archives, databases etc.

Metadata schemas have been created by some records or archives authorities. Here are a few suggestions:

Name	Context	Link	Source
MADRAS project <sup>14</sup>	This is “a register of existing metadata schemas, an analysis of how they relate to several standards, and recommendations for proper metadata sets in preservation.” It was developed through the InterPARES2 project <sup>15</sup> -	<a href="http://www.gseis.ucla.edu/us-interpares/madras/guidelines.php">http://www.gseis.ucla.edu/us-interpares/madras/guidelines.php</a> <a href="http://www.interpares.org/">http://www.interpares.org/</a>	International
PREMIS <sup>16</sup>	Preservation metadata. “The PREMIS data dictionary and XML schema are maintained by the PREMIS maintenance activity, hosted by the Library of Congress”	<a href="http://www.oclc.org/research/projects/pmw/g/">http://www.oclc.org/research/projects/pmw/g/</a>	International

<sup>13</sup> International Council on Archives (ICA)

<sup>14</sup> Metadata and Archival Description Registry and Analysis System (MADRAS)

<sup>15</sup> International Research on Permanent Authentic Records in Electronic Systems 2 (InterPARES2)

<sup>16</sup> Preservation Metadata Implementation Strategies (PREMIS)

Name	Context	Link	Source
GC RMMS	Government of Canada - Records Management Metadata Standard. It “defines a records management metadata element set recommended for use in the Government of Canada” Alphabetical listing of [50] elements in order to promote. It is a "flat" set in that it does not invoke parent-child (i.e. hierarchical) relationships or nesting (i.e. sub-division) of elements.	<a href="http://www.collectionscanada.gc.ca/government/002/007002-5000-e.html">http://www.collectionscanada.gc.ca/government/002/007002-5000-e.html</a>	Canada
GC RMAP	Government of Canada - Records Management Application Profile. “Defines the business rules delineating the use of records management metadata elements declared in the Government of Canada Records Management Metadata Standard”		
(GC) ECMAP	Executive Correspondence Metadata Application Profile		
SPIRT RKMS <sup>17</sup>	Recordkeeping metadata schema. Multiple-entity approach: Business, Agent, and Record	<a href="http://www.sims.monash.edu.au/research/rcrg/research/spirt/">http://www.sims.monash.edu.au/research/rcrg/research/spirt/</a>	Australia
AGRMS	Australian Government Recordkeeping Metadata Standard (currently an Exposure Draft). “Multiple-entity approach: Record, Agent, Business, Mandate, Relationship The standard allows for both multiple-entity and single-entity implementation”.	<a href="http://www.naa.gov.au/records-management/publications/AGRMS.aspx">http://www.naa.gov.au/records-management/publications/AGRMS.aspx</a>	Australia
Queensland Recordkeeping Metadata Standard and Guideline 2008	Multiple-entity approach: Record, Agent, Function. Mapped to SPIRT and to a predecessor to AGRMS.	<a href="http://www.archives.qld.gov.au/metadata.asp">http://www.archives.qld.gov.au/metadata.asp</a>	Australia
Archives New Zealand Electronic Recordkeeping Metadata Standard Technical Specifications 2008	Multiple-entity approach, linked to AGRMS. (Currently an Exposure Draft).	<a href="http://www.archives.govt.nz/continuum/rkp_ublications.php">http://www.archives.govt.nz/continuum/rkp_ublications.php</a> <sup>18</sup>	New Zealand
KS X ISO 23081-1 Records Metadata - Principles Issued in November 2007.	Records metadata standard for current and semi-current records. The standard is a "public agency standard", not a national standard. It applies to all public agencies which manage public records		Korea

<sup>17</sup> Strategic Partnerships with Industry, Research & Training - Recordkeeping Metadata Schema (SPIRT)

<sup>18</sup> For the Exposure draft, see <http://continuum.archives.govt.nz/electronic-recordkeeping-metadata-standard.html#exposure>

## **Guidance on rich media document types**

Through the United States Library of Congress and the National Information Standards Organisation

<b>Name</b>	<b>Context</b>	<b>Link</b>	<b>Source</b>
METS <sup>19</sup>	“The METS schema is a standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library, expressed using the XML schema language of the World Wide Web Consortium. The standard is maintained in the Network Development and MARC Standards Office of the Library of Congress, and is being developed as an initiative of the Digital Library Federation”	<a href="http://www.loc.gov/standards/mets/">http://www.loc.gov/standards/mets/</a>	United States
NISO Z39.87 Data Dictionary – Technical Metadata for Still Images	“This standard defines a set of metadata elements for raster digital images to enable users to develop, exchange, and interpret digital image files. The dictionary has been designed to facilitate interoperability between systems, services, and software as well as to support the long-term management of and continuing access to digital image collections.”	<a href="http://www.niso.org/kst/reports/standards/">http://www.niso.org/kst/reports/standards/</a>	United States
Library of Congress Audio-Visual Prototyping Project. Video (Source) Data Dictionary	The prototyping projects are developing approaches for the digital reformatting of moving image and recorded sound collections as well as studying issues related to "born-digital" audio-visual content. The projects include explorations of the scanning of motion picture film and the reformatting of video recordings from tape to digital files. [ ] The first phase (1999-2004) made a preliminary assessment of transfer technology (audio workstations) together with a thorough examination of digital-object packaging and METS metadata. The second phase is elaborating on the development of transfer technology and extending the division's use of the MAVIS collection management software into the realm of recorded sound.	<a href="http://www.loc.gov/rr/mopic/avprot/DD_VS_MD.html">http://www.loc.gov/rr/mopic/avprot/DD_VS_MD.html</a>	United States

<sup>19</sup> Metadata Encoding and Transmission Standard (METS)

Name	Context	Link	Source
NISO Metadata for Images in XML (NISO MIX) <sup>20</sup> Technical Metadata for Digital Still Images Standard	“The Library of Congress' Network Development and MARC Standards Office, in partnership with the NISO Technical Metadata for Digital Still Images Standards Committee and other interested experts, is developing an XML schema for a set of technical data elements required to manage digital image collections. The schema provides a format for interchange and/or storage of the data specified in the Data Dictionary - Technical Metadata for Digital Still Images (ANSI/NISO Z39.87:2006). [ ] MIX is expressed using the XML schema language of the World Wide Web Consortium.	<a href="http://www.loc.gov/standards/mix/">http://www.loc.gov/standards/mix/</a>	United States

---

<sup>20</sup> National Information Standards Organisation (NISO) Metadata for Images in XML Schema (MIX)