Digital Records Pathways: Topics in Digital Preservation

Module 4: An Overview of Metadata

InterPARES / ICA
DRAFT July 2012

Table of Contents

Table of Figures

Digital Records Pathways: Topics in Digital Preservation

1 Preface

Digital Records Pathways: Topics in Digital Preservation is an educational initiative developed jointly by the International Council on Archives (ICA) and the International Research on Permanent Authentic Records in Electronic Systems Project (InterPARES). It offers training to archivists and records professionals in the creation, management and preservation of authentic, reliable and usable digital records. The program assumes that the user has a solid grounding in basic concepts of records management and archival theory, and builds on that knowledge.

Consisting of eight independent modules, *Digital Records Pathways* addresses the theoretical and practical knowledge needed to establish the framework, governance structure and systems required to manage and preserve digital records throughout the records' lifecycle.. Each module addresses a specific topic of relevance to the management and preservation of digital records. The program is provided free of charge on the ICA website at www.ica.org/.

1.1 About the ICA and InterPARES

The ICA and InterPARES are committed to establishing educational materials for the continuing education of archivists and records managers, to build upon foundational knowledge, disseminate new findings, and to equip archivists and records professionals with the necessary specialized knowledge and competencies to manage and preserve digital records.

The International Council on Archives (ICA) (www.ica.org) is dedicated to the effective management of records and the preservation, care and use of the world's archival heritage through its representation of records and archives professionals across the globe. Archives are an immense resource. They are the documentary by-product of human activity and as such an irreplaceable witness to past events, underpinning democracy, the identity of individuals and communities, and human rights. But they are also fragile and vulnerable. The ICA strives to protect and ensure access to archives through advocacy, setting standards, professional development, and enabling dialogue between archivists, policy makers, creators and users of archives.

The ICA is a neutral, non-governmental organization, funded by its membership, which operates through the activities of that diverse membership. For over sixty years ICA has united archival institutions and practitioners across the globe to advocate for good archival management and the physical protection of recorded heritage, to produce reputable standards and best practices, and to encourage dialogue, exchange, and transmission of this knowledge and expertise across national borders. With approximately 1500 members in 195 countries and territories the Council's ethos is to harness the cultural diversity of its membership to deliver effective solutions and a flexible, imaginative profession.

The International Research on Permanent Authentic Records in Electronic Systems (InterPARES) (www.interpares.org) aims to develop the knowledge essential to the long-term preservation of authentic records created and/or maintained in digital form and provide the basis for standards, policies, strategies and plans of action capable of ensuring the longevity of such material and the ability of its users to trust its authenticity. The InterPARES project has developed in three phases:

InterPARES 1 (1999-2001) focused on the development of theory and methods ensuring the preservation of the authenticity of records created and/or maintained in databases and document management systems in the course of administrative activities. Its findings present the perspective of the records preserver.

InterPARES 2 (2002-2007) continued to research issues of authenticity, and examined the issues of reliability and accuracy during the entire lifecycle of records, from creation to permanent preservation. It focused on records produced in dynamic and interactive digital environments in the course of artistic, scientific and governmental activities.

InterPARES 3 (2007-2012) built upon the findings of InterPARES 1 and 2, as well as other digital preservation projects worldwide. It put theory into practice, working with archives and archival / records units within organisations of limited financial and / or human resources to implement sound records management and preservation programs.

1.2 Audience

The audience for this program includes archivists and records and information professionals interested in expanding their competencies in the management of digital records. Taken as a whole, the modules form a suite of resource materials for continuing professional education with particular focus on issues influencing the preservation of reliable, accurate and authentic digital records.

1.3 How to Use the Modules

Each module consists of theoretical and methodological knowledge and its practical application, illustrated through case studies and model scenarios. While the modules have been developed by InterPARES Team Canada, and are therefore illustrated with examples from the Canadian context, each module is customizable for a specific domain or juridical context. For wider applicability, they have been translated into the languages of the ICA partners.

The modules can be studied individually according to need and interest, or as a set, covering the range of competencies required. They can be self-administered by individuals, or offered through professional associations or workplace training. The modules also contain a number of templates that allow universities and professional associations to adapt and to develop specific course curricula, on-site training materials for students and professionals on digital recordkeeping and preservation issues. Universities and professional associations are free to adapt the materials and develop their own context-specific course curricula and training kits.

1.4 Objectives

The modules have the following objectives:

- To provide educational resources based on cutting edge research in digital records issues to professional archival and records management associations for the benefit of their members;
- To provide archivists and records managers with the necessary theoretical knowledge as well as procedural and strategic skills to develop, implement and monitor a digital recordkeeping and/or a preservation program;
- To illuminate theoretical concepts with practical applications through real life examples drawn from case studies, anchored in specific administrative and technological contexts;
- To provide university programs with content and structure for courses on digital records management and preservation.

1.5 Scope

Digital Records Pathways: Topics in Digital Preservation consists of the following modules:

Module 1: Introduction – A Framework for Digital Preservation
Module 2: Developing Policy and Procedures for Digital Preservation
Module 3: Organizational Culture and its Effects on Records Management Selection and Appraisal of Digital Records
Module 4: An Overview of Metadata
Module 5: From *Ad Hoc* to Governed – Appraisal Strategies for Gaining Control of Digital Records in Network Drives
Module 6: E-mail Management and Preservation
Module 7: Management and Preservation of Records in Web Environments
Module 8: Cloud Computing Primer

Each module consists of some or all of the following components as appropriate:

- **Overview** of the topic and scope of the module;
- **Learning objectives** and expected level of knowledge upon completion;
- **Methodology** or the procedures to follow in order to apply the module;
- **Templates (where appropriate)** to facilitate the implementation of the module;
- **Case Study(ies)/Scenarios (where appropriate)** that provide real-world examples of module topic
- **Exercises** covering key learning points;
- **Review questions** to enhance comprehension and understanding of the topic;
- Additional **Resources** for the topic, including **readings, standards** and other **templates** for reference

Overview of the set			
1. A Framework for Digital Preservation 2. Developing Policy and Procedures for Digital Preservation			Foundational
3. Organizational Culture	4. An Overview of Metadata	5. Appraisal Strategies	General purpose
6. E-mail	7. Websites	8. Cloud Computing	Specific purpose
International Terminology Database			Foundational

1.6 International Terminology Database

The terminology used in the modules reflects common usage in archival and records management communities of practice. To ensure common understanding, and minimize potential confusion that may arise from regional or jurisdictional practice, all modules are supported by the International Terminology Database, available at http://www.web-denizen.com/. As well, certain specific terms are included in short glossaries in each module.

Module 4: An Overview of Metadata

2 Introduction

Metadata is often defined as *data about data*. Metadata is information that identifies, describes, facilitates access to, and management of, an information object or objects, regardless of medium. In the digital environment, metadata can more accurately be described as "structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource" (National Information Standards Organization 2004).

All records professionals need to be aware of metadata in the context of the digital records and other digital objects for which they are responsible. The choices you make about metadata will affect your ability to access, retrieve, manage, and preserve your digital records and digital objects, and ensure their enduring authenticity and reliability.

2.1 Aims and Objectives

There are many excellent resources available that give detailed information about current metadata standards for a wide variety of information resources. It is not our intention to duplicate that work here (see Section 5 for an annotated list of resources). The objective of this module is to present a high-level overview of metadata in the context of the management and preservation of digitized or born-digital objects. These objects may be available on the Internet, or they may be digital objects created and managed in networked or stand alone systems belonging to individuals or organizations. They may be born digital, or they may be objects that have resulted from the scanning of paper or other analogue objects. The common factor is that the creator or preserver of the digital objects has a reason to consider using some form of structured metadata to manage his/her digital objects for a specific purpose or purposes.

This module will summarize:

- The development of metadata in the digital environment;
- The different functions of metadata;
- Its attributes and characteristics;
- Some of the main metadata standards relating to the management of digital objects and their preservation; and
- The ways in which metadata is structured and expressed.

2.2 Learning Outcomes

Upon completion of this module you will be able to:

- Understand the primary functions of metadata;
- Understand the different categories of metadata;
- Understand the attributes and characteristics of metadata;

- Understand the main metadata standards used by archival repositories; and
- Know where to locate additional information and resources about metadata.

2.3 Definitions

The definitions presented below relate directly to metadata in the digital environment. For more definitions relevant to records professionals, please refer to the ICA International Terminology Database, at [link].

Digital record: a document in digital form that is managed as a record (A document made or received in the course of a practical activity as an instrument or a by-product of such activity, and set aside for action or reference.) (InterPARES)

Information object (also information resource, digital object): a digital item or group of items, regardless of type or format, that can be addressed or manipulated as a single object by a computer (Gilliland 2008).

Interoperability: the ability of multiple systems with different hardware and software platforms, data structures, and interfaces, to exchange data with minimal loss of content and functionality (National Information Standards Organization 2004).

Metadata: information that characterizes another information resource, especially for the purposes of documenting, describing, preserving, or managing that resource. Metadata defines and describes the structure and meaning of information resources, and the context and systems in which they exist. It is used to support efficient and effective management of these information resources over time (Government of Canada 2010). Data describing context, content, and structure of records and their management through time (ISO 2001).

Metadata schema: in general terms, any organization, coding, outline or plan of concepts. In terms of metadata, a systematic, orderly combination of elements or terms. In terms of DCMI term declarations represented in XML or RDF schema language, (http://dublincore.org/schemas/) schemas are machine-processable specifications which define the structure and syntax of metadata specifications in a formal schema language. In terms of an encoding scheme, is a set of rules for encoding information that supports a specific community of users. (http://dublincore.org/documents/usageguide/glossary.shtml#S)

Encoding scheme: an encoding scheme provides contextual information or parsing rules that aid in the interpretation of a term value. Such contextual information may take the form of controlled vocabularies, formal notations, or parsing rules. If an encoding scheme is not understood by a client or agent, the value may still be useful to a human reader. There are two types of encoding schemes: Vocabulary Encoding Schemes (which indicates that the value of the element is taken from a controlled vocabulary), and Syntax Encoding Schemes (which indicate that the value of a string is formatted in accordance with a formal notation). (http://dublincore.org/documents/usageguide/glossary.shtml#S)

3 Metadata unpacked

Digital communications technologies are ubiquitous in today's networked world. The power of networks is driving knowledge sharing, and the value of networked communication depends on our ability to link, access, manage, preserve, and share information resources. Metadata provides the means to accomplish these tasks over the life cycle of information resources. Metadata is of fundamental importance to information exchange, retrieval, and understandability, and is essential to our ability to manage and preserve the resources it describes. However, it is only as powerful as its capacity to be read and understood across communication boundaries, whether technical or human. Technical boundaries to communication include software and hardware incompatibility, obsolescence or issues of backwards compatibility; human communication barriers include language, vocabulary, and dialect. Regardless of the purpose of metadata for a particular object or set of objects, interoperability is critical. This depends on systematic knowledge representation, structured according to agreed standards.

Any information that is created and kept as a representation of another object or facet of that object may be considered metadata. Examples include the information about archival holdings contained in finding aids, inventories, and archival descriptions, bibliographic information relating to published materials, inventories and registers of business records, and catalogues of artistic works. The described objects may be physical or virtual, and the metadata itself may be human- or machine-readable. The word 'metadata,' however, came into common parlance in the mid-1990s in the communities involved in data management and systems design in general, and the management and sharing of geospacial data in particular. It referred to a suite of industry and domain-specific standards and documentation necessary for identification, representation, interoperability, technical management, performance, and use of data in information systems (Gilliland 2008).

Cultural heritage information professionals – those who work in archives, libraries, and museums – have always used metadata to effect intellectual and physical or technical control over digital objects in their custody or control. Cultural heritage professionals typically manage resources that are no longer used by their creator and have been transferred to their custody and care for continued access, long term management, and preservation. Finding aids, file lists, inventories, catalogue records, thesauri, union lists, are all examples of metadata describing context, content, and structure of digital (and non-digital) resources. Each domain tailors metadata element sets for purposes specific to its requirements.

Library metadata functions primarily as a tool for physical and intellectual access to bibliographic materials. Realized in indexes and cataloguing records, library metadata has been systematically and cooperatively created and shared since the 1960s. *Examples: cataloguing content standards AARC2 and RDA, and transmission standard MARC21.*

Archival and museum metadata facilitates management of and access to holdings by providing structured information that documents contents, tracks rights information, aids resources discovery, establishes relationships and context, and supports preservation. It

traditionally has been created to identify and preserve the evidential value of individual and aggregated resources through detailed description. This metadata documents the contexts and inter-relationships of the resources, facilitating authentication and assisting researchers in interpretation and analysis of the material. Standards for archival description have been developed to facilitate sharing and gain intellectual control over holdings. Standards-compliant descriptions, realized in part through metadata, provide specific, structured information. *Examples of metadata standards for archival description: EAD, RAD, DACS.* Museum metadata standards exist for documentation of humanities collections, art, architecture, and other visual and audio resources, ethnographic and archaeological collections, and natural science collections. *Examples of metadata standards for collections: VRA Core 4.0, DOCAM, Darwin Core.*

Records professionals working with current records also use metadata. Governments, businesses and other records creators rely on metadata as part of their recordkeeping requirements to manage, access, and use their records. Recordkeeping metadata is added at creation and throughout the period of active use of records, enabling implementation of statutory mandates through metadata that addresses all aspects of records management including location and retrieval, access restrictions, retention and disposition. *Example: Minnesota Recordkeeping Metadata Standard, 2003,* at http://www.mnhs.org/preserve/records/docs_pdfs/mnrkms_2003.pdf.

There are also standards for managing specific digital media formats, for managing and protecting intellectual property rights, for digital preservation, for educational resources, for managing and sharing research data sets, among other special purpose metadata standards. These and other standards are described in Section 3.

3.1 Purposes of metadata

The most common purpose for metadata is to help locate and provide access to a resource (resource discovery). However, metadata can serve any purpose for which information about an object is required. Metadata elements can be added, harvested, or otherwise captured that identify an object, provide information on provenance, usage and access rights, trace the history of how the object was created, provide information about how it has been and is to be managed, requirements for its preservation, and how it is related to other resources.

Metadata is classified or categorized according to the purpose it is intended to accomplish (Caplan 2009). The categories are not mutually exclusive, and metadata elements frequently fit into more than one. Common usage identifies the following categories:

- *Descriptive metadata* – data elements or properties that identify a digital resource and are used to find and interpret the resource.
- *Administrative metadata* – data elements or properties used to manage the resource. Administrative metadata may include:
    - *Technical metadata* – data elements or properties that provide information about the technical context of the resource,

    - *Rights metadata* – data elements or properties describing rights and obligations adhering to the digital resource such as ownership, copyright or other intellectual property rights, usage and security restrictions,
    - *Preservation metadata* – data elements or properties describing requirements for preserving the resource over time and across technological change. (Note: Preservation metadata may also be considered a separate category of metadata – e.g., PREMIS preservation metadata – see annotations below.)
- *Structural metadata* – data elements or properties that document the structural relationships between or within digital resources, for example the file structure within which a digital resource resides, or the linkage between pages in a website. Structural metadata supports proper display and use of complex objects.

These categories of metadata derive from the creation, maintenance, and preservation of resources. Another category may be identified based on the use of resources, either from analytics, or user-generated content.

- *Use metadata* - data elements or properties collected about or from the users of the resource (e.g., social tags, access logs, user search logs)

3.2 Key concepts

The focus of this module is metadata that is *structured* information, digitally created, captured, managed and preserved, regardless of the medium of the resource(s) it describes. Structure relies on syntax and vocabularies, expressed through abstract models, and realized in metadata standards, schemas, and application profiles. A metadata syntax is a set of rules governing the form and structure of metadata elements. It allows metadata to be exchanged and understood. Metadata vocabularies are sets of terms chosen to represent metadata elements and values.

Example: The Dublin Core Metadata Element Set is a vocabulary of fifteen properties used for resource description. Two of the **elements** are "creator," defined as "An entity primarily responsible for making the resource," and "date," defined as "A point or period of time associated with an event in the lifecycle of the resource." (www.dublincore.org/documents/dces/). Examples of **values** might be "John Smith" and "January 1, 2012."

The following sections provide an overview of these concepts, and introduce some of the better known metadata standards and resources.

3.3 How is metadata created

Metadata can be added to an information resource in one of two ways: 1) automatically, by the different levels of technology that interact in its creation, and 2) manually, by the creator, custodian, or user of the information resource.

As digital information resources are created, managed, and transmitted across space, time, and technology, they accrue information – metadata – applied by the software, middleware and hardware through which they are created and manipulated. Often not immediately visible to the user, this intrinsic metadata is generally technical in nature, e.g., file path, file format extensions, resolution, bit-depth, frame-rate. It may also identify the resource – for example, e-mail header and routing information.

Metadata may also be created and added to or associated with a resource by human beings. This metadata is the most difficult and expensive to create, but may be the most important for purposes of resource discovery and sharing, management, use, and preservation.

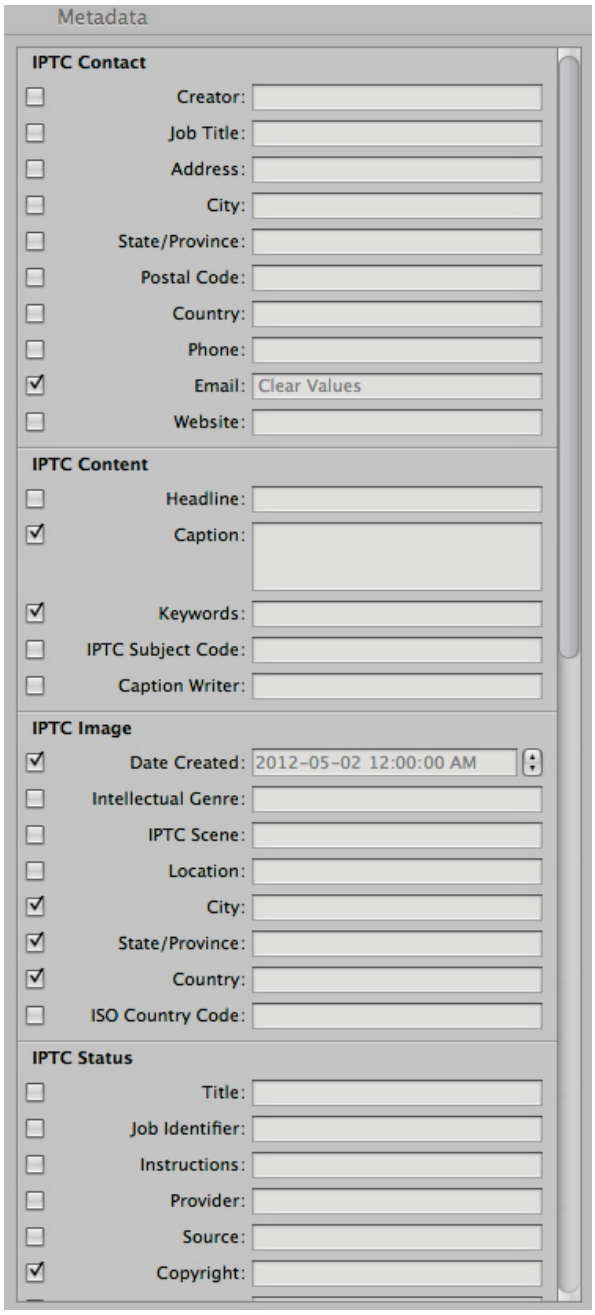


Figure 1: Metadata capture

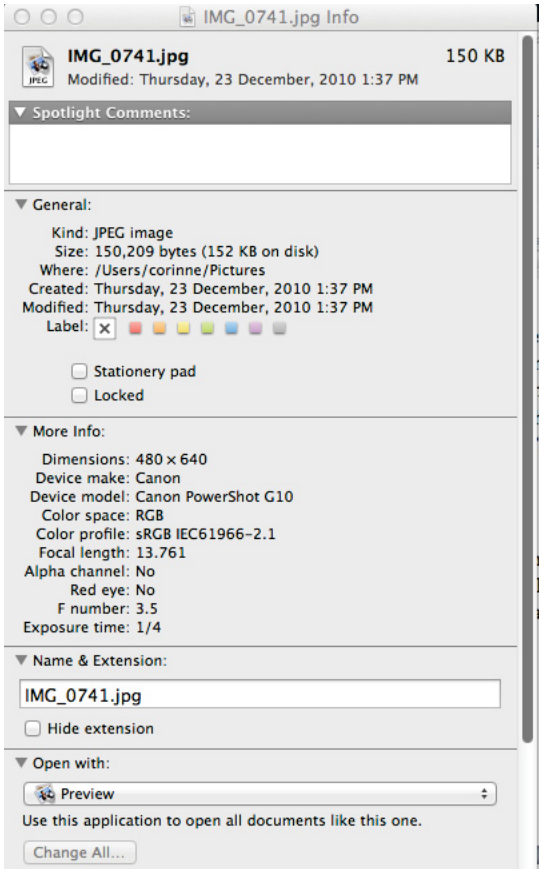
Metadata may be added at any level of aggregation, from the level of individual components of digital objects, to an individual item, to a file, series, or fonds. This may be accomplished in several different ways:

1. Separate metadata records may be created to describe individual 'things' (collection, item, part of an item) and then links made within the metadata record to related files and metadata records – e.g., Dublin Core schema;
2. Complex, multi-level metadata schemas may be created capable of describing different levels within a single metadata record – e.g., SEPIADES schema;

3. Different kinds of metadata may be used to describe the various levels of a complex resource and then linked using special metadata schemas that are intended to structure and coordinate other metadata – e.g., METS schema.

3.4 Where is metadata kept

1. Embedded within a digital file;
2. Within a database (digital asset management system);
3. In a separate XML-encoded file;
4. A combination of these methods.



Metadata embedded within a digital file is usually technical metadata applied by the software or system in which the file is created and managed. The file extension, indication of software version, digital camera information including make and model and exposure information, and activity logs that track access or edits to a resource are all examples of this type of metadata. This implicit metadata is held within the file or embedded within the resource.

A digital asset management system may store information about resources in a database, creating an index of described resources through linked metadata. The database may be queried to search for information about the resources or to locate the resources themselves.

Figure 2: Embedded metadata

3.5 When is metadata added

Metadata may be generated or added to a resource (or component, or collection of resources) at the moment of creation, or at any other time throughout the life cycle of the resource. When and what metadata to add to a resource will depend on the purpose of the metadata and the intention of the person or system adding it. Metadata may be added for resource discovery (e.g., Dublin Core), for preservation (e.g., PREMIS), or by a

community of practice in order to share information across institutions (e.g., Darwin Core, which facilitates sharing of biodiversity datasets).

Archival metadata – that is, metadata for the purpose of archival description – is the source of much current discussion among archivists. Archivists prepare archival descriptions that offer a hierarchical cascade of metadata, moving from the general to the specific, that identifies and describes resources and their contexts to facilitate their access, use, and management. The traditional view of archival description holds that it is an activity undertaken at the time of transfer of resources from the creator, when the purpose for which they were created is finished, to a trusted custodian, who will keep them for evidential, historical, and research purposes.

The Canadian Rules for Archival Description (RAD) 2008 provide a standardized system of rules for arc description grounded in fundamental archival theory and traditional Canadian practice. The guiding principles of RAD are: 1) Archival description should be undertaken with attention to requirements for use, 2) The description of all archival material (e.g., fonds, series, collections and discrete items) should be integrated and proceed from a common set of rules, 3) Respect des fonds *is the basis of archival arrangement and description, 4) Creators of archival material must be described, and 5) Description reflects arrangement (i.e., levels of description are determined by levels of arrangement).*

Traditional archival description moves from the general to the specific in a "top down" approach to records. Starting with the fonds, or the highest level of aggregation, description establishes the context within which records are situated. Traditional description is retrospective, conducted by an archivist entrusted with the records of a creator when that creator no longer needs them. Standards-based description, conducted by a trusted preserver, establishes "grounds for presuming the authenticity of the material by documenting its chain of custody, arrangement, and circumstances of creation and use" (Bureau of Canadian Archivists Planning Committee on Descriptive Standards 2008). This is a key part of establishing grounds for presuming records' authenticity and maintaining their evidential value.

Criticisms of traditional description undertaken "at the terminal stages of the life cycle" with respect to digital records focus on two problems: 1) the enormous backlogs that develop when large aggregations of records are transferred from the creator to archival custody, and 2) the limited ability to capture crucial contextual and structural information throughout the lifecycle that is essential to understand the *fonds.* The solution proposed by these critics is to introduce early in the life cycle of the records a metadata systems strategy for describing and managing digital information (Wallace 1995). These criticisms are countered with the observation that post-hoc description will fail only if the records have been poorly managed in the first place – effective archival description is the consequence of effective records management (MacNeil 1995).

Records managers configure recordkeeping systems to record information about individual records as they participate in transactions, preserving descriptive information

that begins at the item level. Archivists are beginning to exploit the use of item-level metadata for digitized objects so that users can navigate from the "bottom up" (Gilliland 2008). This focuses attention on metadata at the item level. However, this does not replace traditional archival description. Complete metadata is necessary from the moment of records' creation in order to identify, authenticate, and track records; it is a critical resource to facilitate archival description and much of the metadata accumulated throughout the life cycle may become part of the archival description, but is not, nor should it be considered, a replacement for description.

3.6 Structure and standards

For metadata to achieve its potential as a tool in creating, managing, retrieving and using information resources, it must be systematic and structured. Furthermore, for metadata to function in locating and sharing resources across multiple systems with different hardware and software platforms, data structures, and interfaces, its structure must be standardized. Thus an important reason for metadata to be structured and standardized is for interoperability.

Metadata consists of *elements*, also referred to as *fields*, *properties*, or *tags*, depending on the context and the user community. Attributes or characteristics of digital objects are identified by the creator or domain of practice as important for a particular purpose. They are then captured and represented by metadata elements. These elements can then be searched and analyzed in the course of accessing and managing the resource. The collection of all elements established by an organization or user community is referred to as the *element set, metadata scheme or schema,* or *structure standard.*

Structure standards ensure consistency and enable searching and data sharing across a domain of practice. Structure standards exist for many purposes, such as the Australian Recordkeeping Metadata Schema, which allows for standardized information that identifies, authenticates, describes, manages, and makes accessible records that document social and organizational activity and the business contexts in which the records are created, managed, and used, and the VRA Core, used to describe works of visual culture and the images that document them, and MPEG: Moving Picture Experts Group, for coded representation of digital audio and video and related data. Hierarchical standards such as EAD: Encoded Archival Description allow for context as well as content to be described.

Each element is populated by values. For example, an element, "Date", may have the value, "January 1, 2000". Values may be expressed as free text strings, but more often the values are themselves structured and controlled according to *content standards* established by the creator or domain of practice. Values may be structured through rules of syntax, for example, the value "January 1, 2000" may be recorded as 2000/01/01. Another way in which values are structured is through finite lists provided by the creator or domain of practice to represent value choices for a given element, represented in instruments such as controlled vocabularies, thesauri, taxonomies, ontologies, authority lists, and naming conventions. Consistent use of value domains and content rules is another way to support interoperability within and across systems. *Examples include ISO 3166 Country Codes and DCMI Type Vocabulary.*

Metadata mark-up and packaging standards ensure that metadata is machine-readable to enable automated searching, and that objects and their metadata linked and are bound together in standardized ways. *Examples include XML: Extensible Markup Language, METS: Metadata Encoding & Transmission Standard, and OAI: Open Archives Initiative.*

3.7 Interoperability – Dublin Core Metadata Initiative

A key strength of metadata standards depends on their ability to work together. In the digital environment, interoperability is a core functional requirement of metadata standards. Interoperability has been defined as "the ability of different types of computers, networks, operating systems, and applications to work together effectively, without prior communication, in order to exchange information in a useful and meaningful manner" (Woodley 2005). Interoperability requires conventions about semantics, or the meaning of elements, structure, whether human- or machine-readable, and syntax, the rules that govern both. For users of digital resources greater interoperability positively affects resource discovery and metadata re-use. For creators of digital resources, interoperability enhances capacity for distribution of materials.

The Dublin Core Metadata Initiative (DCMI) is an organization "dedicated to fostering the widespread adoption of interoperable metadata standards and promoting the development of specialized metadata vocabularies for describing resources to enable more intelligent resource discovery systems" (Woodley 2005). The Dublin Core Metadata Element Set consists of fifteen core metadata elements for simple and generic resource discovery, and is a standard for cross-domain resource description. It was ratified as an ANSI/NISO standard (Z39.85-2007) and IETF RFC 5013 in 2007, and as an ISO standard (ISO 15836:2009) in 2009.

The worldwide Dublin Core community engages in open consensus building in order to create metadata standards that are neutral with respect to purpose and technology, and have a cross-disciplinary focus. Four levels of interoperability are recognized:

> Level 1: shared natural language (term) definitions, generally limited to a particular application environment, for example a library system or repository federation.
>
> Level 2: formal semantic interoperability, which allows metadata to be freely exchanged across metadata-using applications.
>
> Level 3: description set syntactic interoperability, in which applications share an abstract syntax for validatable metadata records.
>
> Level 4: description set profile interoperability, in which the abstract syntax is further controlled by common constraints and shared vocabularies.

These four levels can be represented by the following diagram (http://dublincore.org/metadata-basics/):

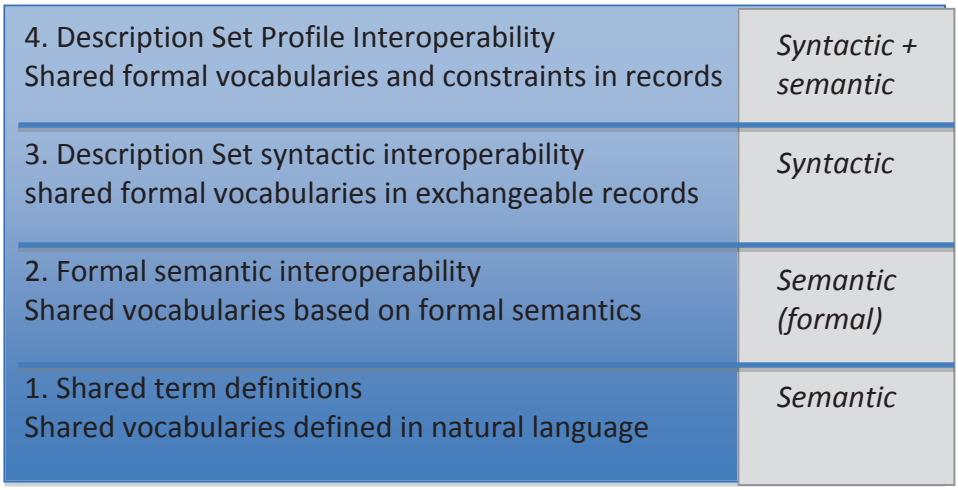


Figure 3: Levels of metadata

Interoperability can also be viewed as synchronic, diachronic, and intentional (Tennis et al, 2010). **Synchronic interoperability** is interoperability at one point in time. This is the ability of one application or system to communicate and work with another. For example, one recordkeeping system may capture distinct names of agents acting on a record: author, writer, originator, while another recordkeeping system does not capture that distinction.

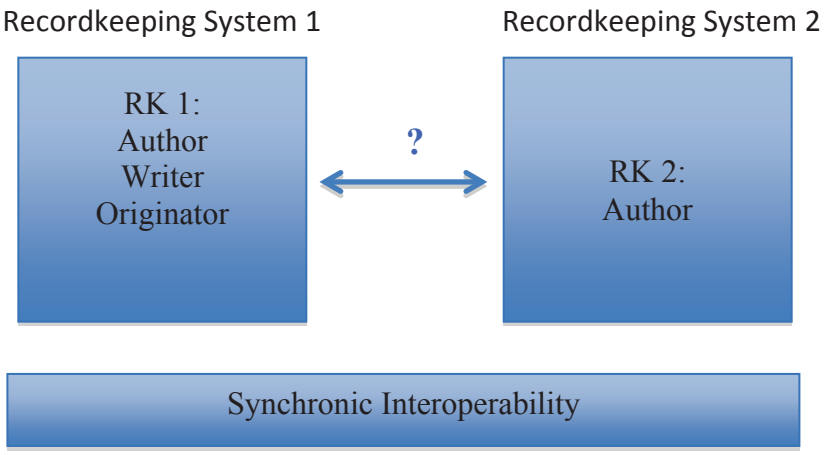


Figure 4: Synchronic interoperability

Applications and systems that are interoperable at one point in time also need to communicate over time regardless of updates to each. This ability to remain interoperable over time is **diachronic interoperability**.

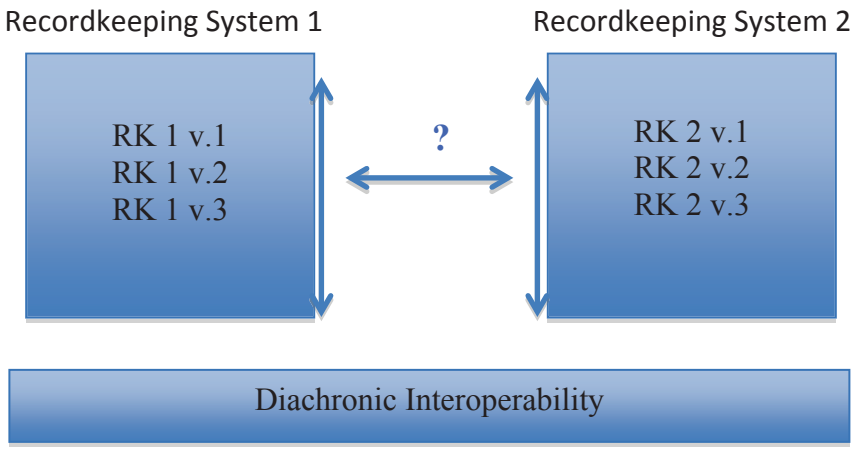


Figure 5: Diachronic interoperability

Because one of the strengths of metadata is its capacity to allow information about resources to be shared, and because metadata is expensive and labour intensive to create and manage, accepted generic metadata standards (or portions of the standards), like Dublin Core, are often adopted, customized, and combined with other accepted standards (or portions) to serve a specific purpose of a user community or domain of practice. **Intentional interoperability** aligns the purpose between pre-existing element sets, resulting in an application profile (see below) that, when published with documentation that clearly states its purpose and functional requirements, may then be adopted by other members of the user community to address common functional requirements.

3.8 Application profiles

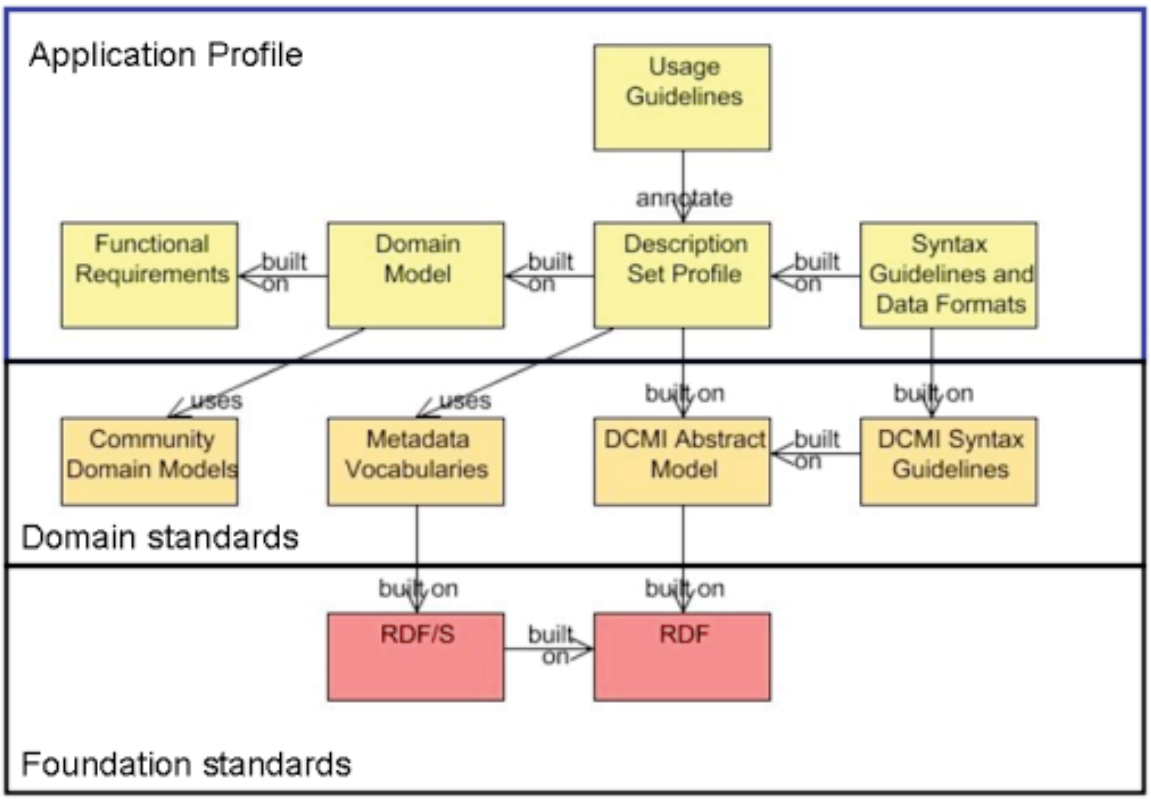
Intentional interoperability is realized through metadata element sets created specifically to address a particular purpose or set of functional requirements. These metadata sets, or **application profiles**, promote the sharing and linking of data within and between user communities.

DCMI has developed a framework for designing Dublin Core Application Profiles (DCAP) which define metadata records that meet specific application needs while providing semantic interoperability with other applications on the basis of globally defined vocabularies and models (http://dublincore.org/documents/profiles-guidelines/). Metadata application profiles describe a set of guidelines, description rules, and constraints used in creating a specific set of metadata records. They provide high level syntactic or structural interoperability. The semantics of the terms used in an application profiles relate to their definitions, and are therefore independent of the application profile. Semantic interoperability therefore works across application profiles.

Application profiles developed in compliance with this framework (called the Singapore Framework from the location of the meeting in 2007 at which the Framework was

adopted) are packets of documentation consisting of three mandatory and two optional components:

1. Functional requirements (mandatory): describe the purpose the metadata will serve the community ;
2. Domain model (mandatory): represents the resources being described and their relationships;
3. Description Set Profile (DSP) (mandatory): lists the metadata elements that will be used;
4. Usage guidelines (optional): outlines the rules for applying and using the metadata elements; and
5. Encoding syntax guidelines (optional): defines how the metadata will be encoded



(from http://dublincore.org/documents/singapore-framework/)

Figure 6: Application profile – Singapore Framework

Functional requirements describe the purpose the metadata is to fulfill. Clearly articulating the functional requirements allows you to identify what information must be captured. Once the metadata elements have been determined, they can be validated against the functional requirements for internal consistency and completeness.

Determining the functional requirements of your metadata may involve input from community stakeholders, records professionals, and system designers.

Statement of the functional requirements should address the following questions (this is not an exhaustive list – questions to ask will be context-specific):

- What is the primary purpose of this metadata?
- What will this application profile provide that cannot be found in other metadata standards?
- What other metadata standards will contribute to this profile?
- How will this metadata be used?
- Who will use this metadata?

InterPARES 3 General Study – Developing an Application Profile for Record Authenticity

The International Research on Permanent Authentic Records in Electronic Systems (InterPARES) 3 approached the problem of implementing the findings of InterPARES 1 and 2 in small and medium-sized institutions to facilitate the creation, maintenance and preservation of digital records that could be trusted as authentic and reliable over time. Digital records are the lasting traces of actions and transactions, and their evidentiary capacity depends on their authenticity, reliability and accuracy, established at creation, and maintained and preserved over time and across technological change. Foundational findings of InterPARES 1 identified the elements of a record's identity and integrity that were required for the presumption of authenticity. While much research has focused on functional and design requirements for metadata to facilitate access and preservation, little work has explicitly connected the theory of digital records and recordkeeping with functional and design requirements for metadata attesting to the presumption of authenticity of those records. InterPARES 3 recognized that extant metadata schemas (or as they are increasingly being called, metadata vocabularies) do not explicitly account for all the elements of identity and integrity required to be captured for the presumption of authenticity over a record's life cycle. To address this gap, InterPARES 3 established a general study to develop a metadata application profile for authenticity. The following examples are drawn from that general study (Tennis and Rogers 2012).

Example: IP3 Application Profile – Functional Requirements

Researchers identified six functional requirements the authenticity metadata application profile that can be expressed in the following statement: These metadata should be necessary and sufficient to support the presumption of authenticity of records, interoperate between systems and across time, be adequate for archival description, and be useful for both retrieval and meaningful display of records.

1. *Presumption of authenticity: The InterPARES Benchmark and Baseline requirements provide the elements necessary for the presumption of authenticity.*
2. *Interoperability should be supported between systems (synchronic) and across time (diachronic).*

3. *Parsimony: metadata elements are both necessary and sufficient to support the presumption of authenticity.*
4. *Adequacy for archival description: All of the metadata generated will be considered temporary unless needed for the other functional requirements, and will be adequate for a thorough archival description.*
5. *Retrieval*
    a. *Semantic: metadata will be added to aid retrieval that carries proof of authenticity, as needed, for example, in e-discovery,*
    b. *Technical: metadata that will guarantee the technical components of the record can be reassembled to create the record with fixed form and stable content.*
6. *Meaningful display: metadata elements will be clearly linked to the functions, persons, and contexts that participate in the lifecycle of the record.*

The **domain model** defines the entities described by the application profile and their relationships. The DCAP (Singapore Framework) does not require or specify any particular modeling technique. The model may be visual or descriptive.

Example: IP3 Application Profile – Domain models

The domain model, in the context of metadata application profiles, describes in both words and in entity-relationship diagrams, what is to be described and how the resources to be described relate to one another. The researchers constructed three levels of the domain. At the highest level is the aggregation of records – the series, or the fonds depending on the context. The next layer is the record, and the third layer is the record attributes. Below are the entity relationship diagrams for the second and third layer- the record and the record attributes.

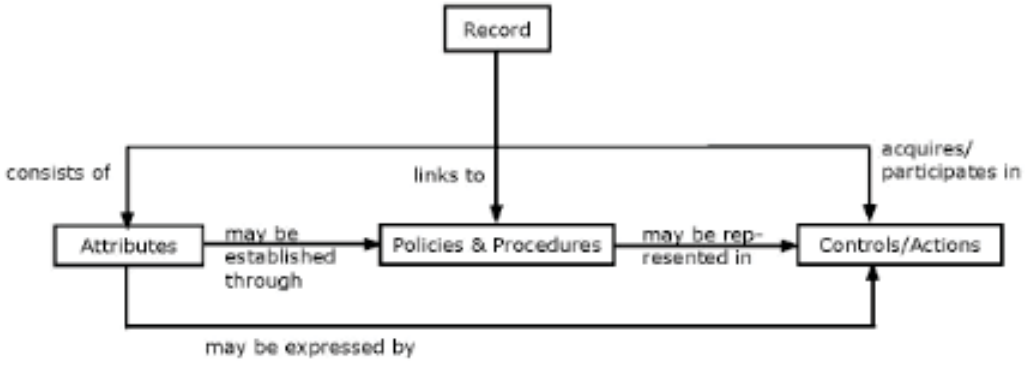


Figure 7: The record in the domain model

The record consists of attributes, links to, or is linked to policies and procedures, and acquires controls as it moves from document to individual record to part of an aggregation of records (see Figure 1). Attributes comprise both intellectual and digital components. These attributes express the identity and integrity of the record in a digital system, statements about which are required to assert, on behalf of the preserver, whether the record can be presumed to be authentic. Thus, any metadata that takes as its purpose the presumption of authenticity must document the identity and integrity of the record by recording the intellectual and digital components. The intellectual and digital components document the record's identity, whereas the policies, procedures and controls primarily (though not exclusively) document the integrity of the record (see Figure 2).

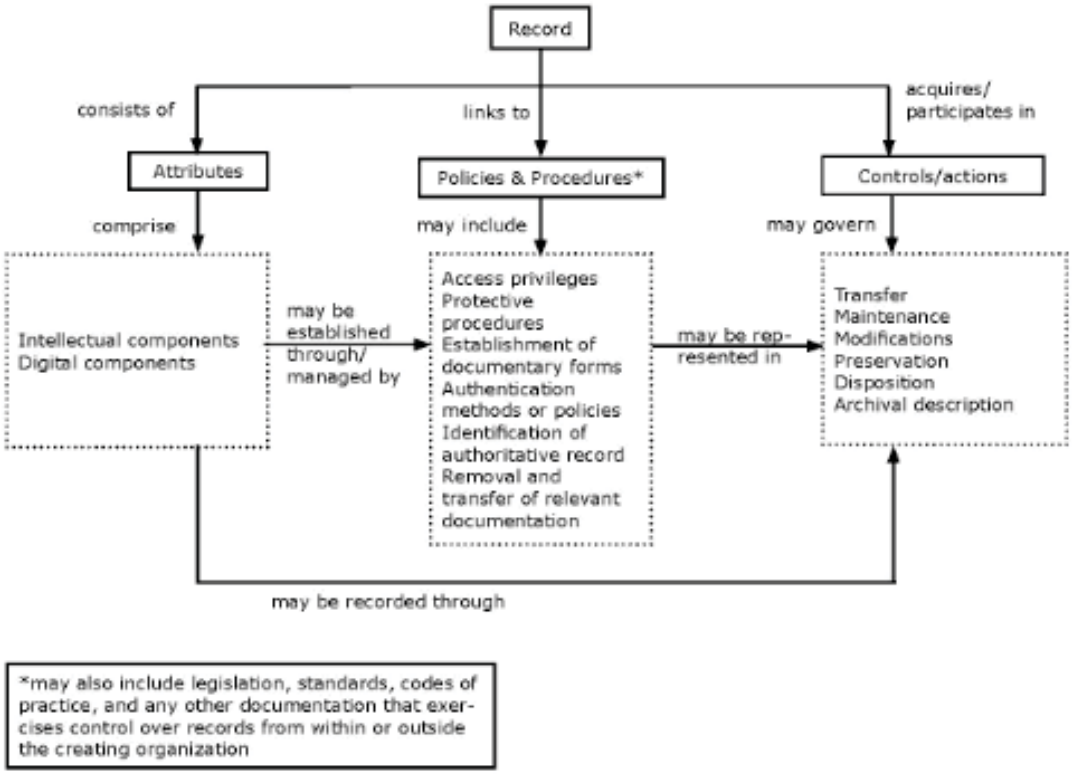


Figure 8: The attributes, policies and procedures, and controls of a record

By mapping to the InterPARES Chain of Preservation model, the "Benchmark Requirements Supporting the Presumption of Authenticity of Electronic Records", and

the "Baseline Requirements Supporting the Production of Authentic Copies of Electronic Records" (InterPARES 2008), the researchers identified all the elements required for identification and capture to satisfy the first functional requirement (presumption of authenticity). Figure 3 shows these elements.

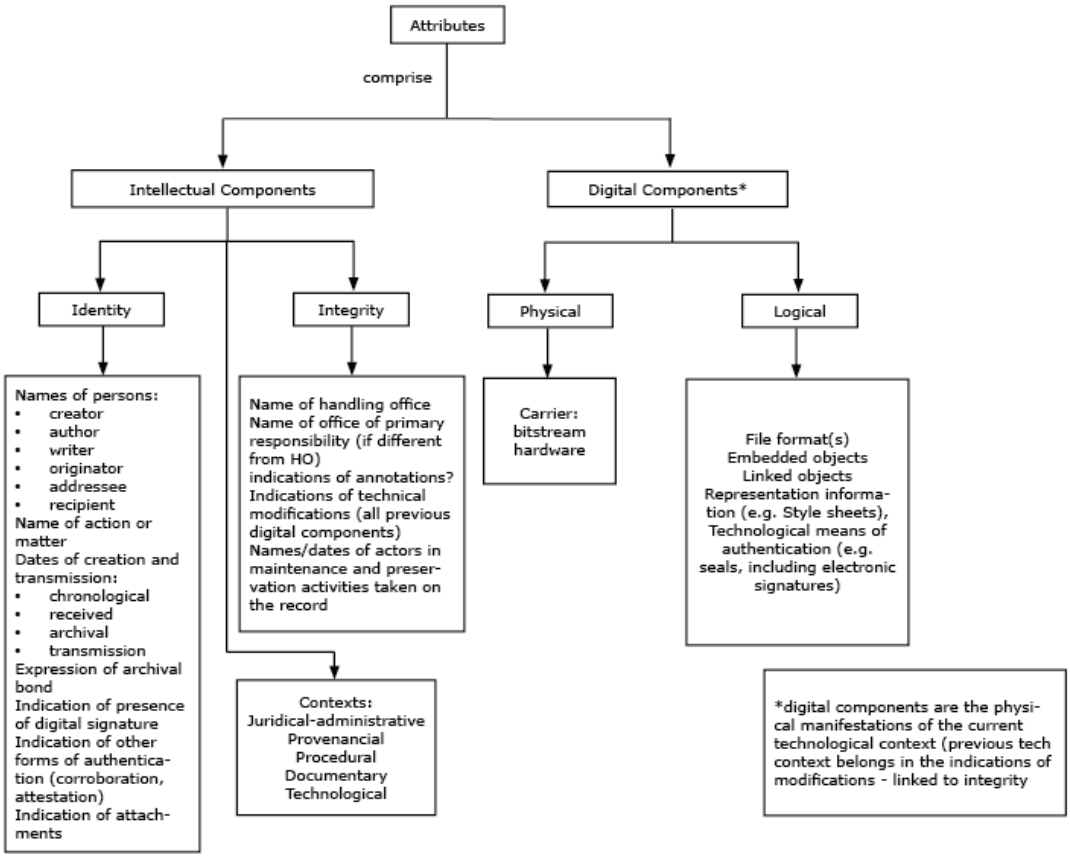


Figure 9: Attributes: the intellectual and digital components of a record

The **Description Set Profile** (mandatory) defines the metadata elements or properties that that will describe the entities in the model.

Example: IP3 Application Profile –Description set profile
The researchers identified the following areas of description (metadata elements):

- *Persons (the persons collaborating in the creation of a record)*
- *Dates*
- *Subject (of action or matter in which the record participates – for example, document title)*
- *Bond (the contexts of the record, expressed in terms of relationships through classification codes and similar identifiers)*
- *Attachments*

- *Technical information*
- *Elements of form*
- *Indication of signs and seals*
- *Indication of means of authentication*
- *Rights and access information*
- *Links to external documentation and system metadata*

Usage guidelines (optional) describe the application profile, its implementation, and define the properties and terms.

Encoding syntax guidelines (optional) describe any encoding syntax that may be used (see **Representation or Encoding** below).

3.9 Metadata crosswalks

A metadata crosswalk is a mapping of elements in one metadata schema to equivalent elements in another metadata schema. A crosswalk acts as a blueprint when transferring metadata from one schema to another, or when developing application profiles. Crosswalks support interoperability between and sharing of metadata records.

Example: an excerpt of a crosswalk between Dublin Core, DACS, and EAD (from http://www.getty.edu/research/publications/electronic_publications/intrometadata/crosswalks.html)

Dublin Core	DACS (Describing Archives: a Content Standard)	EAD (Electronic Archival Description)
Type	3.1 Scope and Content	<controlaccess> <genreform> (in archdesc>)
Creator	2.6 Name of creator(s)	<author> (in <eadheader>) <name> <origination> <persname> <corpname> <famname> (in <archdesc>)

Crosswalks can be difficult to create without an in-depth knowledge of the standards being mapped. A complete, or fully-specified crosswalk requires both a semantic mapping and a metadata conversion specification that will transform the metadata record content compliant with the source standard to metadata record content that is also compliant with the target standard.

Issues to be resolved in element to element mapping include (but are not limited to):

- One-to-many or many-to-one
- Elements existing in only one of the sources or targeted standards
- Equivalent elements qualified as mandatory in one and optional in the other
- How to handle hierarchical elements
- Single vs. multiple object standards (St. Pierre and LaPlant, Jr. 1998; Woodley 2008)

Example: IP3 Application Profile – Crosswalks
The researchers conducted an element to element crosswalk between several metadata schemas. The following is a small snapshot:

Purpose	MoReq	Dublin Core	Inter-PARES CoP Model	InterPARES Terminology	PREMIS	Inter-PARES Codes	Description
Acquisition	M14.4.30	Identifier	A43.2.1.2	acquisitionCode	2.4 (or no match)	B12	Transfer registration number assigned by the transferring agent
Acquisition	M14.4.29		A43.2.1.3	acquisitionDate	2.3	D17	Date and time the transfer was received
Acquisition			A43.2.1.4	registrationPerson	3.2, 2.6, 2.6.1, 2.6.2, 2.6.3	P23	Name of the person registering the transfer
Acquisition		Identifier	A4.3.2.1.5	registrationCode	2.1, 2.1.1, 2.1.2	B13	Transfer registration number assigned by the person registering the transfer
Acquisition			A4.3.2.1.6	acquisitionInfo	2.4 (or no match)	DO16	Indication of the reason/authorization for the transfer (e.g., reference to the relevant terms and conditions of transfer)
Acquisition			A4.3.2.1.7	acquisitionInventory	2.5, 2.5.1	DO17	Indication of records and other transfer documentation received
Acquisition			A43.2.1.8	notificationOfReceiptRecipient	3.2, 2.6, 2.6.1, 2.6.2, 2.6.3	P24	Name of person(s) to whom a notification of recipt of transfer was issued
			A4.3.2.1.9	notificationOfReceiptPerson	3.2, 2.6, 2.6.1, 2.6.2, 2.6.3	P25	Name of the person who issued the notification

3.10 Representation or encoding

Metadata can be encoded in many ways. Three common encoding languages used to format metadata are HTML, XML, and RDF.

3.10.1 HTML: Hypertext Markup Language

HTML tags format Web resources for display, and their elements are prescribed.

3.10.2 XML: Extensible Markup Language

XML encodes metadata elements for meaning and exchange (http://www.w3.org/XML/). It was created to structure, store, and transport information. It provides a software- and hardware-independent way to store data, facilitating sharing and transport between incompatible systems. Because XML has no predefined tags, metadata schema elements can be turned into tags and "wrapped around" specific values. XML is the foundation of

many metadata standards, such as METS: Metadata Encoding and Transmission Standard.

3.10.3 RDF: Resource Description Framework

The Resource Description Framework is a language for representing information about resources in the World Wide Web (http://www.w3schools.com/rdf/default.asp). Like XML, RDF encodes metadata elements for meaning and exchange. It provides the foundation for processing and exchanging machine-readable metadata, and promotes automated processing of Web resources. RDF retains the capability to exchange metadata between application communities, while allowing each community to define and use the metadata that best serves it. (DC Glossary).

For more information see http://www.w3.org/TR/rdf-primer/ and http://www.w3.org/RDF/.

Example: IP3 Application Profile –RDF

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
        xmlns:dc="http://purl.org/dc/elements/1.1/"
        xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
        xmlns:foaf="http://xmlns.com/foaf/0.1/"
        xmlns:owl="http://www.w3.org/2002/07/owl#"
        xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
        xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
        xmlns:dcterms="http://purl.org/dc/terms/">

<!-- version 0.1 Corinne Rogers, and Joseph T. Tennis authors for the InterPARES 3
Research Project interpares.org -->

<rdf:Description rdf:about="http://ipam.info#A2.2.2.0">
        <rdfs:isDefinedBy rdf:resource="http://ipam.info#COPModel"/>
        <rdfs:label rdf:resource="http://ipam.info#A2.2.2.0"/>
        <rdfs:label xml:lang="en">chronDate00</rdfs:label>
        <dcterms:description xml:lang="en">the date of document
creation</dcterms:description>
        <rdfs:domain rdf:resource="http://ipam.info#Record"/>
        <dcterms:identifier rdf:resource="http://ipam.info#D00"/>
</rdf:Description>

<rdf:Description rdf:about="http://ipam.info#A2.2.2.1">
        <rdfs:isDefinedBy rdf:resource="http://ipam.info#COPModel"/>
        <rdfs:label rdf:resource="http://ipam.info#A2.2.2.1"/>
        <rdfs:label xml:lang="en">chronDate01</rdfs:label>
        <dcterms:description xml:lang="en">chronological date (and possibly time) of
compilation and capture</dcterms:description>
        <rdfs:domain rdf:resource="http://ipam.info#Record"/>
```

<dcterms:identifier rdf:resource="http://ipam.info#D01"/>
</rdf:Description>

3.11 Metadata Harvesting

Metadata may be automatically harvested from different distributed, independent sources who wish to share their metadata, and aggregated to facilitate broader resource discovery. The current standard protocol for metadata harvesting is the OAI-PMH: Open Archives Initiative Protocol for Metadata Harvesting, an "application-independent interoperability framework" (http://www.openarchives.org/OAI/openarchivesprotocol.html) based on HTML and XML. OAI-PMH facilitates resource discovery through automated and efficient dissemination of resource metadata.

3.12 What metadata should be created and maintained

Metadata can be expensive and time-consuming to create and manage. Businesses and cultural heritage organizations must decide what metadata they require for their business purposes. This decision will balance their rights and obligations with the cost of adding, harvesting, and maintaining metadata. If you are considering developing a metadata application profile or adopting an existing metadata standard you should consider the following questions (not all may be relevant, and there may be other questions specific to your organization) (adapted from www.dcc.ac.uk):

- What are your requirements for metadata? Why do you need metadata - to serve what purposes?
- What do you want to accomplish with the metadata you want to create or capture?
- What are the functional requirements of the metadata you need to create?
- How and with whom will the digital objects be shared?
- What is the budget for metadata creation? For metadata maintenance?
- What is the format of the digital objects being described by the metadata?
- How will the metadata be captured?
- How will the metadata and digital objects be stored, accessed, and delivered?
- When will the metadata be added/captured in the life cycle of the digital objects?
- Do metadata structure standards already exist as de facto standards that are applicable?
- Will you need to build a specific application profile to address your functional requirements?
- What content standard(s) will you use?
- What encoding standard will you use?
- What transmission standard will you use?

4 Metadata Standards

This section provides a brief annotated guide to some of the better known standards.

4.1 Dublin Core

http://dublincore.org/

Dublin Core is a simple and generic metadata schema for resource description. Intended to be capable of describing any type of resource, it is widely used and adapted. Developed from the mid-1990s through a process of international collaboration, it is maintained by the Dublin Core Metadata Initiative (DCMI). In its simple 15-element form, Dublin Core has been disseminated as part of the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) and has achieved IETF RFC 5013, ANSI/NISO Standard Z39.85-2007, and ISO Standard 15836:2009 standardization. DCMI has developed a larger set of elements and sub-elements (DCMI Metadata Terms) and a framework for the development of application profiles (Dublin Core elements combined with specialized vocabularies developed for particular purposes). Dublin Core can be encoded with a variety of syntaxes, including text, HTML, XML, and RDF.

4.2 Recordkeeping Metadata

4.2.1 Archives New Zealand: Electronic Recordkeeping Metadata Standard (June 2008)

http://archives.govt.nz/advice/continuum-resource-kit/continuum-publications-html/s8-electronic-recordkeeping-metadata-stand

Electronic Recordkeeping Metadata Standard establishes principles and minimum requirements for creating and managing recordkeeping metadata in electronic environments. The standard is mandatory for public offices and local authorities and supports the Public Records Act 2005. It addresses point-of-capture metadata and recordkeeping process metadata in order to identify and describe the content, context, and structure of records, conditions of their use and security, relationships with other records, people, and business being transacted, and to identify past and future events which document recordkeeping actions affecting authenticity and integrity.

4.2.2 Australian Government Recordkeeping Metadata Standard (Version 2.0, July 2008)

http://aa.gov.au/Images/AGRkMS_Final%20Edit_16%2007%2008_Revised_tcm2-12630.pdf

This standard is a revision of the *Recordkeeping Metadata Standard for Commonwealth Agencies Version 1.0*, published by the National Archives of Australia in 1999 to guide Australian Government agencies and vendors in the development of electronic records management systems. It is based on a multiple-entity model, allowing for description of five separate entities: Record, Agent, Business, Mandate, and Relationship. It describes the minimum metadata necessary to ensure that records remain accessible and usable over time, and some of the metadata necessary to manage the preservation of digital records for ongoing agency business needs, or when those records are held in a digital archive.

4.2.3 Australian Recordkeeping Metadata Schema (RKMS)

http://www.infotech.monash.edu.au/research/groups/rcrg/projects/spirt/deliverables/austrkms.html

The Australian Recordkeeping Metadata Schema (RKMS) is a main deliverable of the SPIRT Recordkeeping Metadata Project led by Monash University, and provides a standardized set of structured recordkeeping metadata elements, a framework for developing and specifying recordkeeping metadata standards, and a framework for reading or mapping metadata sets to allow semantic interoperability. The RKMS metadata elements provide standardized information that identifies, authenticates, describes, manages, and makes accessible records that document social and organizational activity and the business contexts in which the records are created, managed, and used. It was developed to promote compatibility between related resource management tools. In particular, the RKMS is harmonized with Dublin Core and Australian Government Locator Service metadata initiatives. The *Recordkeeping Metadata Standard for Commonwealth Agencies Version 1.0* is considered a subset of the RKMS.

4.2.4 ISO 23081-1:2006 "Information and Documentation – Records Management Processes – Metadata for Records – Part 1

http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=40832

Part 1 covers the principles governing records and their metadata, the processes that affect them, the systems in which they are created and maintained, and organizations responsible for their management.

4.2.5 ISO 23081-2:2009 "Information and Documentation – Records Management Processes – Metadata for Records – Part 2:Conceptual and Implementation Issues"

http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=50863

Part 2 establishes the framework for defining metadata elements according to the principles articulated in Part 1, in order to enable standardized description of records, and to support interoperability of records and metadata over time, across space, and across applications. It identifies issues arising in the implementation of metadata for managing records and the options for addressing these issues.

4.2.6 ISO/TR 23081-3:2011 "Information and Documentation – Records Management Processes – Metadata for Records – Part 3: Self-assessment method"

http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=57121

Part 3 provides guidance on conducting a self-assessment in order to identify the current state of metadata capture and management, identify priorities and key requirements,

evaluate progress in the development of a metadata framework, and evaluate system and project readiness for including metadata functionality in a system.

4.2.7 Treasury Board of Canada – Standard on Metadata

http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?evttoo=X&id=18909&section=text

The Treasury Board of Canada Standard on Metadata sets out guidelines for applying recordkeeping metadata to information resources of business value to the Government of Canada using the ISO 23081 generic set of metadata elements, elements specifically defined for use in Government of Canada recordkeeping repositories, and properties from the Dublin Core Metadata Initiative (DCMI) Metadata Terms. The standard also applies ISO 639-2/T Codes for the Representation of Names of Languages, and the World Wide Consortium Date and Time Formats (W#CDTF) syntax encoding schema.

4.2.8 United Nations Standard on Recordkeeping Metadata – Archives and Records Management Section (ARMS)

http://archives.un.org/unarms/doc/arms_standard_on_recordkeeping_metadata.pdf

This Standard describes the metadata that the United Nations Archives and Records Management Section (ARMS) recommends should be captured in recordkeeping systems in all UN offices. It is intended to be used in conjunction with the ARMS *Functional Requirements for Electronic Recordkeeping Systems*, the ARMS *Manual for the Design and Implementation of Recordkeeping Systems*, and other documents about recordkeeping at the UN. The document details the importance of standardized recordkeeping metadata to ensure the recording of adequate contextual information about transactions, assist in retrieval of records, control access, facilitate transfer, reduce fraudulent use and unauthorized access, promote efficiency and economy, and provide a benchmark for measuring quality and supporting auditing.

4.3 Archival Metadata

4.3.1 DACS: Describing Archives: a Content Standard

http://www.archivists.org/governance/standards/dacs.asp

Adopted by the Society of American Archivists as an official SAA standard, DACS is widely adopted in the US as the standard for archival description. It supersedes APPM: Archives, Personal Papers, and Manuscripts, and is the US implementation of ISAD(G) and ISAAR(CPF). DACS is a multi-level standard – an "output-neutral set of rules" applicable to all media.

4.3.2 EAD: Encoded Archival Description

http://www.loc.gov/ead/

The EAD metadata schema provides an XML encoding for archival descriptions. It adopts a multi-level approach to description, providing information about a collection as a whole and then breaking it down into groups, series and (if significant) individual

items. EAD grew out of work done at UC Berkeley in the mid 1990s and was influenced by TEI: Text Encoding Initiative and ISAD(G) (see below). Version 1.0 was released in 1998 with a major revision in 2002 (Version 2002). EAD is maintained by the US Library of Congress and Society of American Archivists, but is used internationally, including the UK. The DACS content standard (see above) provides guidelines for US archivists on how to enter data into EAD.

4.3.3 ISAD(G): General International Standard Archival Description

http://www.icacds.org.uk/eng/ISAD%28G%29.pdf

ISAD(G) outlines metadata elements that should be used in the description of archival collections. It adopts a multi-level approach to description, providing information about a collection as a whole and then breaking it down into groups, series and (if significant) individual items. ISAD(G) has influenced national archival standards and the development of the international archival encoding schema: EAD (see above) and the European SEPIADES schema (see below). ISAD(G) is in its 2nd edition, published in 1999.

4.3.4 METS: Metadata Encoding and Transmission Standard

http://www.loc.gov/standards/mets/METSPrimerRevised.pdf

The Metadata Encoding and Transmission Standard (METS) is a data encoding and transmission specification, expressed in XML, that provides the means to convey the metadata necessary for both the management of digital objects within a repository and the exchange of such objects between repositories. The METS XML schema was developed in 2001 under the sponsorship of the Digital Library Federation, is supported by the Library of Congress, and governed by the METS Editorial Board. It received NISO registration in 2004, renewed in 2006. A key function of the METS standard is to structure or package other metadata or data for exchange or delivery. METS can embed or link to other XML-based metadata (e.g., MODS or PREMIS). Any number or type of digital files can be described and linked together by a METS record, enabling it to represent very complex digital resources (e.g., a whole digitized book, with bibliographic data, images and transcribed text).

4.4 Library Metadata (Cataloguing Standards)

4.4.1 AARC2: Anglo-American Cataloguing Rules

http://www.aacr2.org/

See RDA: Resource Description and Access.

4.4.2 MARC21

http://www.loc.gov/marc/

The MARC (Machine-Readable Cataloguing) standards are metadata transmission standards used by libraries for the representation and communication of bibliographic and

related information in machine-readable form. MARC21 resulted from the combination and revision of the US and Canadian MARC formats to make it more accessible internationally.

4.4.3 RDA: Resource Description and Access

http://www.rda-jsc.org/rda.html
http://www.rdatoolkit.org/

RDA is a new cataloguing standard for resource description and access developed by the North American, British, and Australian library communities and built on (and intended to succeed) AACR2. RDA provides guidelines and instructions on resource description and access for all types of content and media. RDA provides guidelines on cataloguing digital resources, and supports the clustering of bibliographic records in order to show relationships between works and their creators. RDA Toolkit is an integrated, browser-based, online subscription product that includes RDA instructions, workflows, mappings of RDA to difference schemas, and other related resources.

4.4.4 MODS: Metadata Object Description Schema

http://www.loc.gov/standards/mods/

MODS is a schema for a bibliographic element set developed by the Library of Congress' Network Development and MARC Standards Office, particularly for library purposes. As an XML schema, it can be used to carry selected data from existing MARC21 records as well as to create original resource description records. It can be used to expose metadata for harvesting, represent original resource description in XML syntax, and offers an element set that is richer than Dublin Core, compatible with library data, and simpler than the full MARC format (MODS: Uses and Features, http://www.loc.gov/standards/mods/mods-overview.html).

4.5 OAI-PMH: Open Archives Initiative Protocol for Metadata Harvesting

http://www.openarchives.org/OAI/openarchivesprotocol.html

The Open Archives Initiative is an important initiative to facilitate the interoperability of metadata records. The OAI Protocol for Metadata Harvesting (OAI-PMH) provides an automated means of requesting metadata records from OAI-compliant repositories, and aggregating the metadata so that it is searchable from one place. Data providers make their metadata sets or a selection of their metadata sets available for harvesting using simple Dublin Core in a standard XML format. Service providers harvest the metadata, making it broadly available as is or with value added after harvesting, for use.

4.6 PREMIS

http://www.loc.gov/standards/premis/

PREMIS provides a Data Dictionary of core metadata elements intended to support digital preservation. Specifically, the Data Dictionary defines preservation metadata that

"supports the viability, renderability, understandability, authenticity, and identity of digital objects in a preservation context; represents the information most preservation repositories need to know to preserve digital materials over the long-term; emphasizes 'implementable metadata': rigorously defined, supported by guidelines for creation, management, and use, and oriented toward automated workflows; and embodies technical neutrality: no assumptions made about preservation technologies, strategies, metadata storage and management etc." The official website provides an XML-encoding for PREMIS, which is intended to facilitate its use with other XML-based metadata such as METS.

4.7 SEPIADES: SEPIA Data Element Set

http://www.ica.org/7363/paag-resources/sepiades-recommendations-for-cataloguing-photographic-collections.html

SEPIADES is a multilevel data element set to catalogue photographic collections, recommended by the European Commission on Preservation and Access (Amsterdam 2003). Hierarchical description is determined by the user, who may create as many levels and sublevels as required, from the level of the institute or repository down to the single item.

5 Review

This module has presented a high level overview of the primary functions, categories, attributes and characteristics of metadata, focusing on the digital environment. It has distinguished between:

- Metadata standard
- Metadata schema
- Metadata application profile
- Encoding scheme

When creating metadata, keep the following points in mind:

- Create or capture metadata as close to the point of resource creation as possible;
- Automate the creation or capture of metadata whenever possible;
- Re-use existing standards whenever possible;
- Always design metadata application profiles to fit functional requirements;
- Adhere to principles that will enhance interoperability; and
- Balance cost, completeness, and functional requirements.

5.1 Review Questions

1. What are some of the different purposes of metadata? Give examples to illustrate your answers.

2. What is the difference between a metadata standard, and a metadata schema, as outlined in this module?
3. What is an application profile?
4. What are the two types of encoding schemes?
5. Why is interoperability important?
6. What is the difference between interoperability and harmonization?
7. In general, what categories of metadata exist and what are their purposes?

6 Additional Resources

Canadian Heritage Information Network: CHIN

http://www.pro.rcip-chin.gc.ca/normes-standards/guide_normes_musees-museum_standards_guide/metadonnees-metadata-eng.jsp

DCC Digital Curation Manual – Installment on Archival Metadata

http://www.dcc.ac.uk/sites/default/files/documents/resource/curation-manual/chapters/archival-metadata/archival-metadata.pdf

Dublin Core Metadata Initiative website (DCMI)

http://dublincore.org/
The Dublin Core Metadata Initiative website is the primary source of information about DCMI. You can find definitions, background, DCMI specifications and training resources, links to other standards, and information about user communities and events.

Joint Information Systems Committee (JISC)

JISC, Cross-media: Managing your digital resources

This series of advice documents is aimed at those who provide support for managing digital collections, or to small collection owners so they can manage their digital assets. The series covers common issues in organizing and managing collections of moving and still images and audio. Topics covered include but are not limited to: an introduction to digital preservation, an introduction to metadata, asset management, rights management, controlled vocabularies, file formats and file naming conventions, accessibility, e-learning, and online content delivery.

JISC (2010) An Introduction to Metadata

http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/an-introduction-to-metadata
This is the first in a series about metadata, and is aimed at those developing managed and sharable digital collection. This document defines metadata and introduces basic concepts.

JISC (2010) Metadata and Digital Images, Metadata and Audio Resources, Metadata and Digital Video

http://www.jiscdigitalmedia.ac.uk/stillimages/advice/metadata-and-digital-images/
http://www.jiscdigitalmedia.ac.uk/audio/advice/metadata-and-audio-resources/
http://www.jiscdigitalmedia.ac.uk/movingimages/advice/metadata-and-digital-video/
These documents introduce concepts and issues to consider when creating metadata for digital image collections or for audio collections. Starting with advice on identifying required metadata, the documents introduces metadata standards and interoperability, crosswalking between metadata schemas, management systems and metadata, and vocabularies.

JISC (2010) Metadata Standards and Interoperability

http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/metadata-standards-and-interoperability/
This document offers a comprehensive overview of metadata standards and the principles behind using them. As with the other JISC advice papers on metadata, the intended audience is someone developing and managing collections of digital image, video, and audio assets.

JISC (2010) Putting Things in Order: a Directory of Metadata Schemas and Related Standards

http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/putting-things-in-order-links-to-metadata-schemas-and-related-standards/
This useful annotated directory introduces the reader to formal metadata schemas and related standards, and provides links to further information. It is intended to be read in conjunction with the other metadata advice papers.

JISC (2010) Controlling Your Language: a Directory of Metadata Vocabularies

http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/controlling-your-language-links-to-metadata-vocabularies/
After introducing the concept of controlled vocabularies, this document lists more than 70 vocabulary sources currently available. These are organized according to their type: thesauri, classifications, and authority lists.

Michael Day, UKOLN. Metadata: Mapping between metadata formats

http://www.ukoln.ac.uk/metadata/interoperability/
This document is listed as a resource in the JISC advice papers on metadata (see above). It was last updated 22 May, 2002, and although many of the links are broken, many other link to crosswalks that have been updated since that time, making it a valuable resource none the less.

Library of Congress – Metadata for Digital Content

http://www.loc.gov/standards/mdc/

National Information Standards Organization (2004) Understanding Metadata

http://www.niso.org/publications/press/UnderstandingMetadata.pdf
This booklet is available free on the NISO website. This version is a revision and expansion of *Metadata Made Simple: A guide for libraries*, published in 2001. *Understanding Metadata* is a short guide that defines metadata, the main types and functions of metadata, and offers a practical approach to structuring and sharing metadata schemas.

The Open Data Foundation

http://www.opendatafoundation.org/

W3C – Linked Data

http://www.w3.org/standards/semanticweb/data

Linked data lies at the heart of the Semantic Web. It is the large scale integration of, and reasoning on, data on the Web. Linked data depends on data and metadata expressed in standard formats that are reachable and manageable by Semantic Web tools. Linked data and Semantic Web technologies are developed by the World Wide Web Consortium (W3C), an international community of member organizations, staff, and the public, working together to develop Web standards. HTML5, XML, and RDF (among others) are cornerstone standards of the Web developed and recommended by W3C.

7 References

Bureau of Canadian Archivists Planning Committee on Descriptive Standards. 2008. *Rules for Archival Description*. Ottawa: Bureau of Canadian Archivists. http://www.cdncouncilarchives.ca/RAD/RADComplete July2008.pdf.

Caplan, Priscilla. 2009. *Understanding PREMIS*. Library of Congress. http://www.loc.gov/standards/premis/understanding-premis.pdf.

Gilliland, Anne. 2008. "Setting the Stage." In *Introduction to Metadata*, ed. Murtha Baca. Version 3.0. Los Angeles, CA: Gregory M. Britton. http://www.getty.edu/research/publications/electronic_publications/intrometadata/.

Government of Canada, Treasury Board of Canada. 2010. "Standard on Metadata". Standard. http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=18909&section=text.

ISO. 2001. *ISO-15489 (2001) Information and Documentation-Records Management*. http://www.iso.org/iso/iso_catalogue.htm.

MacNeil, Heather. 1995. "Metadata Strategies and Archival Description: Comparing Apples to Oranges." *Archivaria* 39 (Spring): 22–31.

National Information Standards Organization. 2004. *Understanding Metadata*. Bethesda, MD: NISO. www.niso.org.

St. Pierre, Margaret, and William P. LaPlant, Jr. 1998. *Issues in Crosswalking Content Metadata Standards - National Information Standards Organization*. White Paper. NISO. http://www.niso.org/publications/white_papers/crosswalk/.

Tennis, Joseph T. 2010. "Metadata Developments: Interoperability Across Systems, Time, and Conceptions" presented at the InterPARES 3, 4th International Symposium, Oslo, Norway.

Tennis, Joseph T., and Corinne Rogers. 2012. *General Study 15: Metadata Application Profiles for Authenticity*. University of British Columbia.

Wallace, David. 1995. "Managing the Present: Metadata as Archival Description." *Archivaria* 39 (Spring): 11–21.

Woodley, Mary S. 2005. "DCMI Glossary." http://dublincore.org/documents/usageguide/glossary.shtml.

———. 2008. "Crosswalks, Metadata Harvesting, Federated Searching, Metasearching: Using Metadata to Connect Users and Information." In *Introduction to Metadata*, ed. Murtha Baca. Los Angeles, CA: Gregory M. Britton.