

DIGITISATION:
STANDARDS LANDSCAPE
FOR EUROPEAN MUSEUMS,
ARCHIVES, LIBRARIES

**DIGITISATION:
STANDARDS LANDSCAPE
FOR EUROPEAN MUSEUMS,
ARCHIVES, LIBRARIES**

edited by

ATHENA WP3
“Working Group”
“Identifying standards
and developing
recommendations”

texts by

Gordon McKenna,
Collections Trust (UK)
Chris De Loof,
Royal Museums of Art
and History (Belgium)

The text of this booklet
is included
in deliverable 3.1,
of the ATHENA project
and is based on a survey
of the content that
ATHENA partners
contracted to provide
to Europeana through
the ATHENA project.
<http://www.athenaeurope.org/getFile.php?id=396>

ATHENA

general co-ordinator

Rossella Caffo
ATHENA Project Manager

design

mt milani

Athena logo

Susan Hazan

web version

<http://www.athenaeurope.org/index.php?en/110/promotional-material>

This work by ATHENA
project is licensed under
a Creative Commons
Attribution
Non-Commercial
Share Alike Licence
(CC-BY-NC-SA)
<http://creativecommons.org/licenses/by-nc-sa/3.0/>



table of contents

Introduction

1. Some basic concepts

1.1 Standard

- 1.1.1. Types of standards
- 1.1.2 Open standards

1.2 Digitisation

1.3 Interoperability

1.4 Metadata

- 1.4.1 Definition difficulties
- 1.4.2 Aspects of metadata
- 1.4.3 A 'better' definition?

2. Standards landscape

2.1 Describing standards

2.2 Information schemes (metadata)

- 2.2.1 Museum specific
- 2.2.2 Archive specific
- 2.2.3 Library specific
- 2.2.4 Historic environment specific
- 2.2.5 General heritage
- 2.2.6 Resource discovery
- 2.2.7 Document encoding

2.3 Multimedia formats

- 2.3.1 Text
- 2.3.2 Image
- 2.3.3 Audio
- 2.3.4 Video
- 2.3.5 Virtual reality
- 2.3.6 Vector graphics

2.4 Other technical standards

- 2.4.1 Search and retrieval
- 2.4.2 Transmission
- 2.4.3 Character encoding

3. Conclusions

3.1 Cultural metadata standards

3.2 Technical standards

4. Standards index

introduction

It is the aim of the ATHENA project to support especially museums in providing object data for publication in Europeana.

However, it is known that in the museum landscape in particular there is a great variety of standards used.

It was therefore necessary to gain an overview of the different standards which are in use with the partners in the ATHENA project.

Therefore a survey has been carried out by Working Group 3 of ATHENA and the results were published within one of the deliverables of the project (<http://www.athenaeurope.org/getFile.php?id=396>).

When discussing standards within the community we realised that there is a wider interest in the overview gained and we therefore decided to publish the overview in the form of this special booklet.

Its content results from the information provided from the more than 20 partners from ATHENA coming from museums, libraries and archives, as well as ministries.

Although it is not a complete list of standards currently available and is not covering every sector relevant, we hope it is of practical information for all those interested, as well as instrumental in building more bridges in standards applications for increased interoperability within the cultural heritage sector.

Monika Hagedorn Saupe

WP2 • Awareness and dissemination:
enlarging the network and promoting the service

1. Some basic concepts

1.1 Standard

The British Standards Institution (BSI), the world's oldest standards setting organisation (1901), says:

“Put at its simplest, a standard is an agreed, repeatable way of doing something. It is a published document that contains a technical specification or other precise criteria designed to be **used consistently as a rule, guideline, or definition**. Standards help **to make life simpler** and to increase the reliability and the effectiveness of many goods and services we use. Standards are created by **bringing together the experience and expertise of all interested parties** such as the producers, sellers, buyers, users and regulators of a particular material, product, process or service.”

To the advantages given above **delivering interoperability** can be added.

1.1.1. *Types of standards*

There are a number of standards typologies. A common one has:

- *De facto*

Standards not formally recognised by a standards setting body, but widely used and recognised by the sector using them as a standard. These are quite common in the IT industry where the dominance of Microsoft, for good or ill, has led to some of its products becoming *de facto* standards (e.g. *Word for Windows*). They may not be the best solution to a situation but they are often the most economically successful.

- *De jure*

Standards formally recognised by a standards setting body (e.g. ISO). They are developed by the common consent of a group of interested parties, with no one party being dominant. However they take a significant amount to time to develop and establish, sometimes leading to them being over-taken by technological developments.

Standards can also be looked at with regard to the environment they were produced and used:

- *In-house*

Standards developed and used in a particular organisation, for a particular purpose. An example of this is a local place name terminology. This would extend an existing national which only covers geography at a level of granularity too coarse to be useful at the local level.

- *Community*
Standards developed by a set of organisations in the same sector for use within that sector. The UK museum documentation standard *SPECTRUM* was developed with domain experts with the aim to benefit from their experience.
- *National*
Standards developed for use within a single country and recognised at a national level. Nationally recognised terminologies are examples of such standards.
- *International*
Standards recognised and used throughout the world, nearly always approved by an international standards setting body, e.g. ISO 8601 is an international standard for date and time.

For some standards it is possible for them to begin as one type and then, with further work and taking part in an approval process, become another type. For example the *CIDOC Conceptual Reference Model (CRM)*, was originally developed by the CIDOC Documentation Standards Working Group, as a community standard; it is now an ISO standard (ISO 21127:2006).

Another type of standard that is worth examining in detail is the *open standard*.

1.1.2 Open standards

- 1 Kenneth Krechmer, *Open Standards Requirements*, "The International Journal of IT Standards and Standardization Research", 4 (2006), No. 1. See: <http://www.csrstds.com/openstds.pdf>

Kenneth Krechmer gives ten requirements for open standards¹:

- 1. Open Meeting** – All may participate in the standards development process.
- 2. Consensus** – All interests are discussed and agreement found, no domination.
- 3. Due Process** – Balloting and an appeals process may be used to find resolution.
- 4. Open IPR** – How holders of IPR related to the standard make available their IPR.
- 5. One World** – Same standard for the same capability, world-wide.
- 6. Open Change** – All changes are presented and agreed in a forum supporting the five requirements above.
- 7. Open Documents** – Committee drafts and completed standards documents are easily available for implementation and use.
- 8. Open Interface** – Supports proprietary advantage (implementation); each interface is not hidden or controlled (implementation); each interface of the implementation supports migration (use).
- 9. Open Access** – Objective conformance mechanisms for implementation testing and user evaluation.
- 10. On-going Support** – Standards are supported until user interest ceases rather than when implementer interest declines.

Most of the requirements are about the development of a standard. The aim is to make this process transparent and democratic.

It is not clear if any standards conform to all the requirements. However formal national and international standards are more likely to adhere to most of them, at least to some extent.

Open IPR, Open Access and On-going Support are the most important considerations for a potential user of a standard. Therefore in our descriptions of standards we have indicated which we consider 'open' in these three areas.

1.2 Digitisation

Digitisation is the process of transformation of original (analogue) material into digital form. There are three distinct types of digitisation:

- **Reproduction** – Digitisation with the aim to reproduce the original material in digital form as accurately as possible. This category includes images, sound, and video.
- **Retrieval** – Digitisation with the aim to find and retrieve original material. This category includes scanned and indexed documents, for example contracts, letters etc. The purpose is not an accurate reproduction, but to increase usage of the material.
- **Procedural** – Digitisation with the aim of capturing information from analogue (paper) museum catalogue systems with the aim to implement automated collection management.

This booklet will mainly look at the standards associated with reproduction digitisation. However it will also discuss some of the standards associated with the other two.

1.3 Interoperability

Interoperability can be defined as:

“The ability of the systems, procedures and culture of an organisation to be managed in such a way as to maximise opportunities for exchange and re-use of information, whether internally or externally”.²

- 2 Helen Ashby – Gordon McKenna – Matthew Stiff. *SPECTRUM Knowledge*. London: MDA, 2001, p. 63.

This definition was written in the context of maximising the sharing of the collective knowledge of an organisation.

Here we are attempting to maximise the opportunities for European cultural organisations to share their content with IST projects (and therefore benefit from that process).

Paul Miller further divides interoperability into 6 types³:

- 3 Paul Miller, *Interoperability: What is it and Why should I want it?*, “Ariadne”, 2000, No. 4. See: <http://www.ariadne.ac.uk/issue24/interoperability/>

- **Technical interoperability** – Facilitated by using common technical standards (e.g. file types, metadata, etc.).
- **Semantic interoperability** – Facilitated by using common vocabularies for the terminologies used in data (e.g. thesauri).
- **Political/ Human interoperability** – Facilitated by understanding and overcoming the barriers caused by the different experiences and agendas of users and information providers.
- **Inter-community interoperability** – Facilitated by recognising differences between discipline communities and overcoming them by working together (e.g. museums, archives and libraries).

- **Legal interoperability** – Facilitated by following the legal restraints imposed on information providers (e.g. *Freedom of Information* and *Data Protection* legislation).
- **International Interoperability** – Facilitated by recognising and overcoming the barriers caused by cultural and linguistic differences.

This booklet is concerned mainly with the facilitation of technical and semantic interoperability. The other four types may be the subject of other projects.

1.4 Metadata

Metadata has become such a popular term it is worth examining its use in more detail.

1.4.1 Definition difficulties

In its origin it is clear that the metadata has taken as its model words like *metaphysics*, *metalanguage*, where the ‘meta’ element of the word indicates:

“2. (of an academic discipline) concerned with the concepts and results of the discipline”⁴

⁴ *Collins Concise Dictionary*, Third Edition, 1995, p. 835.

The use of ‘data’ element has led to the tendency to restrict its usage to digital ‘objects’. For example an often quoted definition is that its: ‘Data about data’⁵. This implies that metadata is concerned solely with the world of text and multimedia on the Internet or on computers.

⁵ See: <http://www.webopedia.com/TERM/m/metadata.html>

Metadata has emerged in recent years as a new ‘buzz-word’ for information professionals, causing confusion and/or unease in some quarters. For example Tony Gill writes: “... the term ‘metadata’ is now increasingly used in contexts where the term ‘data’ would have sufficed just a few short years ago (for example, descriptions of people, objects and events).”⁶

⁶ Tony Gill, *Metadata and the Web*, in: Baca, M. (ed.) *Introduction to Metadata: Pathways to Digital Information*, 3rd rev., Getty Information Institute, 2008. See: http://www.getty.edu/research/conducting_research/standards/intrometadata/metadata.html

In this wider context metadata can be data usually known in the cultural heritage sector as: collections management data, catalogue records and exhibition texts. In fact any ‘data’ can be thought of as ‘metadata’.

So why use the term 'metadata' at all?
Perhaps looking at some of the key aspects of metadata is the best way to find an answer.

1.4.2 *Aspects of metadata*

A key idea in metadata is that of a *resource*. This is the entity that the metadata is about. A danger here is to restrict the idea of a resource to texts and multimedia 'objects' accessible over the Internet and in particular on the Web. A resource is anything one wishes to describe and give access to in some way. A resource can be:

- Texts (electronic or paper-based)
- Physical objects
- Multimedia (image, sound, and video, etc.)
- Software
- Persons
- Organisations
- Places
- Events
- Concepts
- Collections of all the above

Also some resources are **surrogates** for another resource. A surrogate is a representation of resource in some other form. For example: a digital image or photograph of an artwork, or a virtual reality representation of a place, or a facsimile of an object.

It is important to distinguish between a resource and its surrogate when creating metadata for the two entities.

Metadata for a surrogate should not describe the original resource. For example, an original artwork might be out of copyright, but a photograph of it might not be.

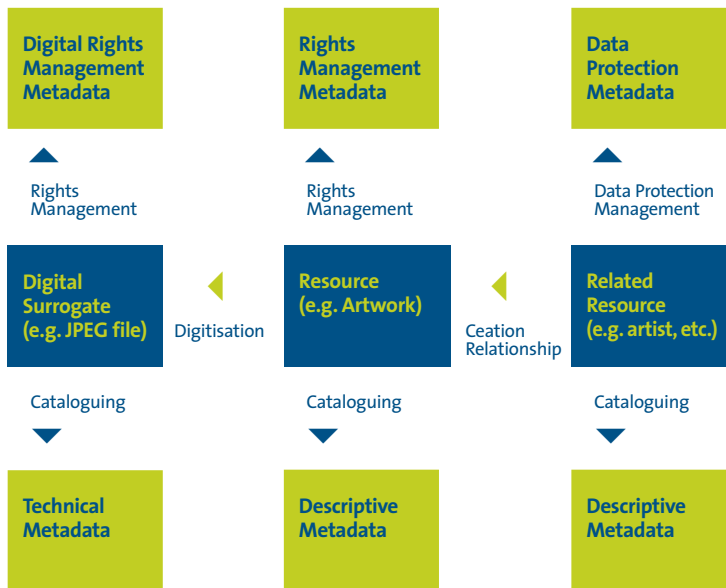
Resources can also be *related* to each in ways other than of original and surrogate. Examples include: creation, making available (e.g. publication), and use.

A number of different typologies for metadata have been proposed. For example Anne J. Gilliland-Swetland⁷ gives:

7 Anne J. Gilliland-Swetland, *Setting the stage*, in: Baca, M. (ed.), *Introduction to Metadata: Pathways to Digital Information*, 3rd rev. Getty Information Institute. 2008. See: http://www.getty.edu/research/conducting_research/standards/intrometadata/setting.html

- **Administrative** – Managing and administering resources (e.g. acquisition, rights, location).
- **Descriptive** – Describing or identifying resources (e.g. catalogue records).
- **Preservation** – Preservation management of resources (condition, and migration data).
- **Technical** – How a system functions or metadata behave (e.g. formats, encryption, passwords).
- **Use** – The level and type of use of resources (user and use tracking information).

The relationships between different resources and their associated metadata can be summarised in the following diagram:



Another type of metadata is that of **resource discovery**. This is metadata aimed at allowing potential user of a resource to find information they need in order to decide whether or not they want to have access to a resource itself. This is a similar situation to putting a term into a Web search engine, viewing the results, and deciding to 'click' on the link. The aim of resource discovery is to give a more accurate and relevant search result for the user.

The most well know resource discovery metadata is *Dublin Core*. This gives information on a resource, gives its **identifier**, and uses the identifier to give *access* to that resource. Access is direct if the identifier is an URL or similar, or indirect in the case of an ISBN (for example).

Finally there is a sense of there being a *metadata movement* taking place with large effort being put into the open development of metadata **schemas**. Schemas are the description of a metadata *element* set, together with a description of how the elements are structured. In turn these schemas are being tested and indeed adopted by organisations and governments. The metadata 'bandwagon' is impossible to ignore!

1.4.3 A 'better' definition?

Metadata is much more than the simple definition given above. Perhaps a more comprehensive one is:

STRUCTURED INFORMATION ABOUT ANY KIND OF RESOURCE, WHICH IS USED TO IDENTIFY, DESCRIBE, MANAGE OR GIVE ACCESS TO THAT RESOURCE.

2. Standards landscape

2.1 Describing standards

Taking the idea of the use of standards to heart, we describe each standard in a Dublin Core (DC) derived format. 9 out of the 15 DC elements are used in the descriptions.

These elements are:

Title

The name (or names) under which the standard is known. In most cases both the abbreviated and the full name is listed.

Creator

The name of the organisation or individual who originally created the standard.

Publisher

The name of the organisation that makes the standard publicly available.

Date

The date on which the standard was *originally* published.

Identifier

A number or other identifier under which standard is published or a URL which points to the definition of the standard.

Rights

Whether rights restrictions, e.g. patents, apply to the standard.

Description

A textual description explaining the standard and its usage.

Subject

Keywords that identify the nature of the standard.

Relation

Other standards that this standard relates to, and associated websites.

The descriptions are aimed at a general reader. More technical details for the majority of standards discussed can be found in various places on the Web.

The purpose of this section is to allow the reader to have an easy reference to the range of relevant standards in one place.

2.2 Information schemes (metadata)

Most of the standards listed below define descriptive metadata. They focus on recording information about resources: books; archives; museum objects; people, places; and organisations.

In addition SPECTRUM is also a procedural standard. This type of standard defines the activities and the information requirements (a type of administrative metadata) needed to manage a specific area of the cultural sector practise. The standards do not necessarily give guidance about its technical implementation. They can be seen as defining 'good-practice' in a specific field.

The descriptions here have been divided into those created for recording information about material from a particular cultural domain:

- Museum specific
- Archive specific
- Library specific
- Historic environment specific
- General heritage.

or for a particular purpose:

- Resource discovery
- Document encoding.

EAD and TEI are also *tagging* standards for the encoding of documents. They were originally based on the text encoding standard SGML, but more recently have been 'updated' to use XML (both of these are described later).

2.2.1 Museum specific

CDWA

Title	CDWA • Categories for the Description of Works of Art
Creator	Art Information Task Force
Publisher	Getty Research Institute
Date	1990
Identifier	http://www.getty.edu/research/institute/standards/cdwa/index.html
Rights	Getty Research Institute.
Description	Describes the content of art databases by articulating a conceptual framework for describing and accessing information about objects and images. They identify vocabulary resources and descriptive practices that will make information residing in diverse systems both more compatible and more accessible. They also provide a framework to which existing art information systems can be mapped and upon which new systems can be developed.
Subject	documentation (museum)
Relation	http://www.getty.edu/research/conducting_research/standards/cdwa/cdwalite.html (CDWA Lite)

museums

MUSEUMDAT

Title	museumdat
Creator	Fachgruppe Dokumentation im Deutschen Museumsbund Institut für Museumsforschung SMB-PK Zuse-Institut Berlin
Publisher	Fachgruppe Dokumentation im Deutschen Museumsbund / Institut für Museumsforschung SMB-PK / Zuse-Institut Berlin
Date	2006-2007
Identifier	http://museum.zib.de/museumdat/museumdat-v1.o.xsd
Rights	Copyright FG Dokumentation im Deutschen Museumsbund / Institut für Museumsforschung SMB-PK / Zuse-Institut Berlin
Description	A harvesting format (XML schema) optimized for retrieval and publication, meant to deliver automatically core data to museum portals.
Subject	description (cultural object)
Relation	http://museum.zib.de/museumdat/museumdat-v1.o-en.pdf (documentation) http://www.museumdat.org/index.php?ln=de&t=home (German website) http://www.museumdat.org/index.php?ln=en (English website)

museums

OBJECT ID

Title	Object ID
Creator	Thornes, Robin (et al)
Publisher	J. Paul Getty Trust
Date	1999
Identifier	http://www.object-id.com/guide/guide_index.html
Rights	Copyright The J. Paul Getty Trust. All rights reserved
Description	Standard for describing cultural objects. Developed through the collaboration of the museum community, police and customs agencies, the art trade, insurance industry, and valuers of art and antiques.
Subject	description (cultural object)
Relation	http://www.object-id.com/checklist/check_eng.html (checklist)

SPECTRUM

Title	SPECTRUM • SPECTRUM: The UK Museum Documentation Standard, 3rd Edition (Version 3.1)
Creator	McKenna, Gordon (ed) Patsatzi, Efthymia (ed)
Publisher	Collections Trust [MDA]
Date	2007
Identifier	1 900642 14 X (ISBN)
Rights	Copyright Collections Trust. All rights reserved. [Open Standard]
Description	Standard for the collections management documentation. Built around 21 procedures that commonly occur in museums. Supported by definitions of 'units of information' – the data needed to support the procedures. There are Dutch language versions for Flanders and the Netherlands. An XML schema is available.
Subject	documentation (museum)
Relation	http://www.collectionstrust.org.uk/spectrum (UK version) http://www.collectionstrust.org.uk/spectrum-nl (Netherlands version) http://www.collectionstrust.org.uk/spectrum-nl-be (Flanders version) http://www.collectionstrust.org.uk/schema (XML schema)

museums

2.2.2 Archive specific

EAD

Title	EAD • Encoded Archival Description
Creator	Encoded Archival Description Working Group of the Society of American Archivists Network Development and MARC Standards Office of the Library of Congress
Publisher	Society of American Archivists
Date	2002
Identifier	ftp://ftp.loc.gov/pub/ead/ead.dtd (DTD) http://www.loc.gov/ead/ead.xsd (W3C schema)
Rights	Copyright Society of American Archivists. [Open Standard]
Description	DTD and schema for the encoding archival finding aids. Also used to describe collections (collection description).
Subject	archive description collection description document encoding
Relation	http://www.loc.gov/ead (EAD website)

ISAD(G)

Title	ISAD(G) • ISAD(G): General International Standard Archival Description, Second Edition
Creator	Committee on Descriptive Standards (ICA/CDS) (adopter)
Publisher	International Congress on Archives (ICA)
Date	2000
Identifier	0-9696035-5-X (ISBN) http://www.ica.org/sites/default/files/isad_g_2e.pdf
Rights	Copyright International Congress on Archives
Description	General rules for archival description that may be applied irrespective of the form or medium of the archival material. The rules accomplish these purposes by identifying and defining twenty-six (26) elements that may be combined to constitute the description of an archival entity.
Subject	archival description
Relation	http://www.ica.org (ica website)

archives

ISAAR (CPF)

Title	ISAAR (CPF) • ISAAR (CPF): International standard archival authority record for corporate bodies, persons and families, Second Edition
Creator	ICA Committee on Descriptive Standard (preparer)
Publisher	International Congress on Archives (ICA)
Date	2004
Identifier	2-9521932-2-3 (ISBN) http://www.ica.org/sites/default/files/ISAAR2EN.pdf
Rights	Copyright International Congress on Archives
Description	General rules for the establishment of archival authority records that describe the corporate bodies, persons, and families that may be named as creators in descriptions of archival documents.
Subject	archival description
Relation	http://www.ica.org (ica website)

2.2.3 Library specific

FRBR

Title	FRBR • Functional Requirements for Bibliographic Records
Creator	IFLA Study Group on the Functional Requirements for Bibliographic Records International
Publisher	International Federation of Library Associations and Institutions (IFLA)
Date	1998
Identifier	ISBN 3-598-11382-X
Rights	Copyright International Federation of Library Associations and Institutions
Description	A conceptual entity-relationship model that relates user tasks of retrieval and access in online library catalogues and bibliographic databases from a user's perspective.
Subject	bibliographic description
Relation	http://www.ifla.org/VII/s13/frbr/frbr.pdf (Final report)

libraries

Title	MAB2 • Maschinelles Austauschformat für Bibliotheken (Automated Library Exchange Format)
Creator	Deutsche Nationalbibliothek
Publisher	Deutsche Nationalbibliothek
Date	2001
Identifizier	http://www.d-nb.de/standardisierung/txt/titelmab.txt [bibliographic data (MAB-TITEL)]
Rights	Deutsche Nationalbibliothek
Description	A German format for machine-readable exchange of all data generated within a library environment (bibliographic, authority and local data).
Subject	bibliographic description
Relation	http://www.d-nb.de/eng/standardisierung/formate/mab.htm (German National Library MAB web page)

Title	MARC • MACHine-Readable Cataloguing MARC 21 Concise Format for Bibliographic Data
Creator	Network Development and MARC Standards Office of the Library of Congress
Publisher	Library of Congress
Date	2002 (updated)
Identifier	http://www.loc.gov/marc/bibliographic/ ecbdhome.html
Rights	[Open Standard]
Description	Standard for the representation and communication of bibliographic information in machine-readable form.
Subject	bibliographic description
Relation	http://www.loc.gov/marc/authority/ ecadhome.html (authority information) http://www.loc.gov/marc/holdings/ echdhome.html (holdings data) http://www.loc.gov/marc/classification/ eccdhome.html (classification data) http://www.loc.gov/marc/community/ eccihome.html (community data)

Title	Information and documentation • Format for Information Exchange
Creator	International Organization for Standardization (ISO)
Publisher	International Organization for Standardization (ISO)
Date	1996
Identifier	ISO 2709:1996
Rights	[Open Standard]
Description	Specifies the requirements for a generalized exchange format suitable for bibliographic descriptions. Describes a framework for communications between data processing systems. Replaces the second edition.
Subject	bibliographic description
Relation	MARC

METS

Title	METS • Metadata Encoding and Transmission Standard
Creator	McDonough, Jerome (et al)
Publisher	Digital Library Federation
Date	2007 (Version 1.7)
Identifier	http://www.loc.gov/standards/mets/mets.xsd
Rights	Copyright Digital Library Federation
Description	XML schema is a standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library.
Subject	digital library
Relation	http://www.loc.gov/standards/mets/mets-schemadocs.html (METS website)

libraries

Title	MODS • Metadata Object Description Schema
Creator	Library of Congress' Network Development MARC Standards Office
Publisher	Library of Congress
Date	2008 (version 3.3)
Identifier	http://www.loc.gov/standards/mods/v3/mods-3-3.xsd
Rights	[?]
Description	XML schema for a bibliographic element set that may be used for a variety of purposes, and particularly for library applications
Subject	bibliographic description
Relation	http://www.loc.gov/standards/mods/ (website)

Title	MIDAS • MIDAS Heritage
Creator	English Heritage Forum on Information Standards in Heritage (FISH)
Publisher	English Heritage
Date	2008
Identifier	http://www.english-heritage.org.uk/upload/pdf/MIDAS_Heritage_Part_One.pdf http://www.english-heritage.org.uk/upload/pdf/MIDAS_Heritage_Part_Two.pdf http://www.english-heritage.org.uk/upload/pdf/MIDAS_Heritage_Part_Three.pdf
Rights	Copyright Forum on Information Standards in Heritage (FISH)
Description	The UK data standard for information about the historic environment. It states what information should be recorded to support effective sharing of the knowledge of the historic environment, and the long-term preservation of those records. It covers: the individual assets that form the historic environment (buildings, archaeological sites, shipwrecks, areas of interest and artefacts); the work that is undertaken to understand, protect and manage change to those assets.
Subject	documentation (historic environment)
Relation	http://www.fish-forum.info (FISH website)
Relation	http://www.heritage-standards.org.uk (FISH Interoperability Toolkit)

2.2.5 *General heritage*

These standards are not 'tied' to any of the domains but can be used by any of them.

Title	CIDOC-CRM • CIDOC Conceptual Reference Model
Creator	CIDOC Documentation Standards Working Group
Publisher	International Organization for Standardization (ISO)
Date	2006
Identifier	ISO 21127:2006.
Rights	[Open Standard]
Description	A conceptual object-oriented model that provides the extensible ontology for concepts and information in cultural heritage and museum documentation.
Subject	documentation (museum)
Relation	http://www.iso.org (ISO website) http://cidoc.ics.forth.gr (CRM website)

general heritage

Title	VRA • VRA Core (Version 4.o)
Creator	Visual Resources Association's Data Standards Committee
Publisher	Visual Resources Association
Date	2007
Identifier	http://www.vraweb.org/projects/vracore4/VRA_Core4_Element_Description.pdf
Rights	Copyright Visual Resources Association [?]
Description	A metadata element set (units of information such as title, location, date, etc.), as well as an initial blueprint for how those elements can be hierarchically structured. The element set provides a categorical organization for the description of works of visual culture as well as the images that document them.
Subject	visual culture
Relation	http://www.vraweb.org/projects/vracore4/index.html (website) http://www.vraweb.org/projects/vracore4/vra-4.o.xsd (XML schema)

general heritage

DUBLIN CORE

Title	Dublin Core • The Dublin Core Element Set Version 1.1
Creator	Dublin Core Metadata Initiative
Publisher	Dublin Core Metadata Initiative
Date	1999
Identifier	http://dublincore.org/documents/1999/07/02/dces/
Rights	[Open Standard]
Description	The Dublin Core is a simple metadata element set intended to facilitate discovery of electronic resources. Elements can be grouped into those having data on: Content – Coverage, Description, Type, Relation, Source, Subject, Title; Intellectual Property – Contributor, Creator, Publisher, Rights; Instantiation – Date, Format, Identifier, Language. Its use has been mandated by several governments in Europe (e.g. UK) and throughout the world (e.g. Australia).
Subject	resource discovery

resource discovery

Title	TEI • TEI P5: Guidelines for Electronic Text Encoding and Interchange
Creator	Burnard, Lou (ed) Bauman, Syd (ed)
Publisher	Text Encoding Initiative Consortium
Date	2007
Identifier	http://www.tei-c.org/P5X/
Rights	Copyright TEI Consortium 2007 Licensed under the GPL. Copying and redistribution is permitted and encouraged. [Open Standard]
Description	Defines a set of tags (markers) for inserting into the electronic form of a document (a text) in order to identify the structure and other features of that document. The aim of these tags is to allow the processing of the text by computer. They are aimed at use with texts in any 'natural' language, of any date and of any genre. This version is XML-compatible.
Subject	document encoding
Relation	XML SGML http://www.tei-c.org (Text Encoding Initiative Consortium website)

document encoding

2.3 Multimedia formats

2.3.1 Text

Text encoding standards deal with how texts are stored in computer systems.

They vary from industry standards (*de facto*) standards, such as Microsoft Word, to World Wide Web Consortium (W3C) endorsed standards, such as XML.

Title	DjVu
Creator	AT&T Labs
Publisher	LizardTech Incorporated
Date	1996 onwards
Identifier	http://djvu.org/docs/DjVu3Spec.djvu [version 3]
Rights	[Open Standard]
Description	Designed mainly to store scanned images, especially those containing text and line drawings.
Subject	document rendering document structure
Relation	http://djvu.org [DjVu website]

Title	HTML • HyperText Markup Language
Creator	Berners-Lee, T
Publisher	World Wide Web Consortium (W3C)
Date	1989 onwards
Identifier	http://www.w3.org/TR/html4/ [version 4.01]
Rights	[Open Standard]
Description	The Hypertext Markup Language was designed by Tim Berners Lee to create a semantic network of documents. The markup language was based on SGML (an SGML application). The emphasis of HTML is on rendering (display) of documents rather than on representing document structure. The HTML standard is maintained by the World Wide Web consortium. The Cascading Style Sheet standard is an addition to HTML to facilitate the lay-out and design of web pages. Version 4 of HTML uses Unicode as its standard character set.
Subject	document rendering
Relation	SGML XML XHTML Unicode

Title	PDF • Portable Document Format
Creator	Adobe Systems Incorporated
Publisher	Adobe Systems Incorporated
Date	1999
Identifier	[not available]
Rights	Copyright Adobe Systems Incorporated. All rights reserved.
Description	A <i>de facto</i> standard for documents. The portable document format was designed by Adobe as a device independent way to represent document structure and lay-out. It is based on the Postscript page rendering standard. The pdf format is widely used to publish documents on the World Wide Web and CD-ROM.
Subject	<u>document rendering</u> document structure
Relation	Postscript RTF

Title	RTF • Rich Text Format
Creator	Microsoft Corporation
Publisher	Microsoft Corporation
Date	1999
Identifier	http://msdn.microsoft.com/en-us/library/aa140277.aspx [Version 6.1]
Rights	Copyright Microsoft Corporation.
Description	The RTF Specification provides a format for text and graphics interchange that can be used with different output devices, operating environments, and operating systems. RTF uses the American National Standards Institute (ANSI), PC-8, Macintosh, or IBM PC character set to control the representation and formatting of a document, both on the screen and in print. With the RTF Specification, documents created under different operating systems and with different software applications can be transferred between those operating systems and applications.
Subject	document rendering document structure
Relation	PDF

Title	SGML • Standard Generalized Markup Language
Creator	Goldfarb, Charles F
Publisher	International Organization for Standardization (ISO)
Date	1986
Identifier	ISO 8879:1986
Rights	[Open Standard]
Description	SGML is the International Standard (ISO 8879) language for structured data and document representation, the basis of HTML and XML. In the cultural sector SGML itself is not widely in use. An early project was the CIMI CHIO demonstrator project. Later projects often use XML, which originally started as a simplified SGML.
Subject	document structure document encoding
Relation	XML HTML http://www.iso.org (ISO website)

WORD

Title	Microsoft Word Document Format
Creator	Microsoft Corporation
Publisher	Microsoft Corporation
Date	1983 onwards
Identifier	[none available]
Rights	Proprietary format of Microsoft
Description	<i>A de facto</i> standard for documents. The .doc format was designed by Microsoft for their Word program and has gone through several version changes. The format is widely in use for exchange of documents. The specification of the format has not been made public.
Subject	document rendering document structure

text

Title	XML • eXtensible Markup Language
Creator	World Wide Web Consortium (W3C)
Publisher	World Wide Web Consortium (W3C)
Date	1996 onwards
Identifier	http://www.w3.org/TR/xhtml1/ [revised 1 August 2002]
Rights	[Open Standard]
Description	XML started as ‘SGML light’, to overcome some of its complexities and terseness. Soon it overhauled its ancestor in popularity and is now the most widely used structuring language for electronic documents. XML structures a document by ‘tagging’ texts. The tags can be freely defined, but can be controlled by a Document Type Definition (DTD) or an XML-schema. XML uses the Unicode character set, so that it is very usable in multi-lingual and international applications. Several XML derivatives have been standardized, such as XSLT (eXtensible Style Language and Transformation) and the XPath query syntax.
Subject	document structure document encoding
Relation	SGML HTML Unicode

2.3.2 *Image*

Still image encoding deals with how images are represented in digital form.

In this section we are dealing with 'raster images' rather than 'vector images'.

Raster images are usually 'photographic' images, where the image is represented by a number of horizontal and a number of vertical picture elements (pixels). Each element has its own colour representation. Raster graphics are produced by scanners and digital cameras. Eventually the individual pixels become visible if raster images are enlarged.

Title	BMP • BitMap
Creator	Microsoft Corporation
Publisher	Microsoft Corporation
Date	1987
Identifier	[not available]
Rights	Copyright Microsoft Corporation
Description	The bmp image format is the Microsoft Windows standard format. It holds black and white-, 16-colour, 256-colour and Truecolor (24 bits) images. The palletized 16-colour and 256-colour images may be compressed via run length encoding.
Subject	image format raster graphics

Title	DjVu
Creator	AT&T Labs
Publisher	Lizardtech
Date	1996 onwards
Identifier	http://djvu.org/docs/DjVu3Spec.djvu [a DjVu plug-in is needed]
Rights	[Open Standard]
Description	Designed mainly to store scanned images, especially those containing text and line drawings. It offers advanced compression technology.
Subject	<u>document rendering</u> document structure
Relation	PDF http://djvu.org [DjVu website]

images

Title	GIF • Graphical Interchange Format
Creator	CompuServe Interactive Services Incorporated
Publisher	CompuServe Interactive Services Incorporated
Date	1987
Identifier	[not available]
Rights	Patent on LZW compression by Unisys.
Description	GIF was created by Compuserve for their online service, but the specifications were made publicly available. GIFs can hold multiple bitmaps of up to 256 colours, each using LZW compressed raster data to minimize file sizes. The format uses Lempel Ziv Welch (LZW) compression that has been patented by Unisys.
Subject	image format raster graphics

Title	JPG • JPEG
Creator	Joint Photographic Expert Group Independent JPEG Group
Publisher	International Organization for Standardization (ISO)
Date	1990
Identifier	ISO/IEC 10918-1:1994
Rights	[Open Standard]
Description	JPEG is a still image compression algorithm based on the fact that the human eye cannot detect subtle differences in colour or contrast. JPEG is a lossy algorithm: the higher the compression factor the more information gets lost. An image that has been compressed using the JPEG algorithm cannot be completely reconstructed. The file format for JPEG compressed images is called JFIF. This file format is what people generally mean when they refer to “JPEG”.
Subject	image format raster graphics
Relation	http://www.iso.org (ISO website)

Title	PNG • Portable Network Graphics
Creator	World Wide Web Consortium (W3C)
Publisher	International Organization for Standardization (ISO)
Date	2003
Identifier	ISO/IEC 15948:2003 (E)
Rights	[Open Standard]
Description	PNG is an extensible format for the lossless, portable, well-compressed storage of raster images. PNG provides a patent-free replacement for GIF and can also replace many common uses of TIFF. Indexed-colour, grayscale, and Truecolor images are supported, plus an optional alpha channel for transparency. Sample depths range from 1 to 32 bits.
Subject	image format raster graphics
Relation	http://www.w3.org/TR/PNG/ http://www.iso.org (ISO website)

Title	TIFF • Tagged Image File Format
Creator	Aldus Corp
Publisher	Adobe Systems Inc.
Date	1992
Identifier	http://partners.adobe.com/public/developer/en/tiff/TIFF6.pdf [Version 6.0]
Rights	[Open Standard]
Description	TIFF provides a general purpose data format and is compatible with a wide range of scanners and image-processing applications. It is device independent and is used in most operating environments. This non-proprietary industry standard for data communication has been implemented by most scanner manufacturers and desktop publishing applications.
Subject	image format raster graphics

2.3.3 *Audio*

Digitisation of sound is achieved through a sound sampling process.

During sampling the analogue signal strength is digitally measured at given time intervals.

The obtained signal values are then stored in compressed or uncompressed form.

The open standard for the use of compressed sound in this area is MP3. The proprietary standards tend to offer better performance, in terms of file size (speed of transmission) and quality of sound.

Title	AIFF • Audio Interchange File Format
Creator	Apple Computer Incorporated Electronic Arts
Publisher	Apple Computer Incorporated
Date	1988-1989
Identifier	http://www-mmsp.ece.mcgill.ca/Documents/AudioFormats/AIFF/Docs/AIFF-1.3.pdf [Version 1.3]
Rights	Apple Computer Incorporated [?]
Description	A non-compressed audio format most widely found on Apple Macintosh computers. Lossless, it is commonly used with professional-level audio and video software and systems.
Subject	sound format

Title	MP3 • MPEG Layer 3 Coding of Moving Pictures and Associated Audio for Digital Storage Media
Creator	ITU-T (International Telecommunication Union Telecommunication Standardization Sector) Moving Pictures Expert Group (ISO/IEC JTC1/SC29 WG11)
Publisher	International Organization for Standardization (ISO)
Date	1993 onwards
Identifier	ISO/IEC 11172:1993, Part 3: Audio ISO/IEC 13818, Part 3: Audio ISO/IEC 14496, Part 3: Audio (Amendment 1: Audio extensions)
Rights	[Open Standard]
Description	An audio compression format common on the Internet. Part of the MPEG standards, it can take larger audio recordings and shrink them down to a fraction of their size while losing little if any fidelity of the sound.
Subject	sound format
Relation	MPEG-1; MPEG-2; MPEG-3 http://www.iso.org (ISO website)

Title	WAV • RIFF Resource Interchange Format
Creator	IBM Microsoft Corporation
Publisher	Microsoft Corporation
Date	1991
Identifier	http://partners.adobe.com/asn/developer/pdfs/tn/TIFF6.pdf
Rights	[?]
Description	Contain sampled audio. The sound information itself it stored in a container using the Resource Interchange File Format. The RIFF file stores data in chunks, including metadata. A WAV file can contain sound clips with different sample rates, number of channels etc.
Subject	sound format
Relation	AVI

audio

Title	WMA • Windows Media Audio
Creator	Microsoft Corporation
Publisher	Microsoft Corporation
Date	[?]
Identifier	[not available]
Rights	Copyright Microsoft Corporation
Description	Microsoft's proprietary competition to MP3. Optimised to deliver audio over the Web, particularly streaming, and using Microsoft products. Has integrated rights management, the advantages of a smaller file size and therefore transfer rates over other formats, including MP3. In addition the sound quality is said to be better.
Subject	sound format
Relation	WMV

REALAUDIO

Title	RealAudio
Creator	RealNetworks Incorporated
Publisher	RealNetworks Incorporated
Date	[?]
Identifier	[not available]
Rights	Copyright Real Networks
Description	Format optimised for delivery of audio over the Web.
Subject	sound format
Relation	RealVideo
Relation	RealMedia

audio

Title	AU
Creator	Sun Microsystems Incorporated
Publisher	Sun Microsystems Incorporated
Date	[?]
Identifier	[not available]
Rights	Copyright Sun Microsystems Incorporated
Description	A sound format for Unix systems. It is the 'standard' audio file format for Java.
Subject	sound format

2.3.4 Video

Digitisation of video adds the time dimension to the digitisation of still images. In principle the process is similar to producing raster images, but then produced in a very fast succession of images. The number of images per second is called the frame rate.

The quality of digitised video is defined by three factors: resolution, colour depth and frame rate. Digitised video can produce large quantities of data. Because of this compression is very important.

Compression relies on the fact that only small portions of the images tend to change between successive frames. With regards to encoding standards, the situation is similar to that with audio encoding, with the MPEG family of standards being open, and a set of proprietary standards, particularly on offer as well.

Title	AVI • Audio Video Interleave
Creator	Microsoft Corporation (for Intel)
Publisher	Microsoft Corporation
Date	[?]
Identifier	[not available]
Rights	Copyright Microsoft Corporation
Description	The earliest video format for PCs. The size of image that can be displayed is dependent on the hardware being used. As compression and decompression functionality is part of Microsoft's <i>Video for Windows</i> package, there is support for this format in a range of hardware and software configurations. File sizes are high and therefore this format ill-suited for delivery over the Internet.
Subject	video format

FLV (Flash Video Format)

Title	FLV • Flash Video Format
Creator	Macromedia (now Adobe)
Publisher	Adobe Systems Incorporated
Date	2002 onwards
Identifier	http://www.adobe.com/devnet/flv/pdf/video_file_format_spec_v10.pdf [Version 10]
Rights	Copyright Adobe Systems Incorporated
Description	Used for the delivery of video over the Internet. It is viewed either using separate 'player' software or using a web browser 'plug-in'. It is becoming the <i>de facto</i> standard of video embedded on web pages.
Subject	video format

video

MOV (Quicktime)

Title	QuickTime
Creator	Apple Computer Incorporated
Publisher	Apple Computer Incorporated
Date	1991 onwards
Identifier	[not available]
Rights	Copyright Apple Computer Incorporated
Description	Apple's proprietary video (and virtual reality) format and system. Built into the Mac's operating system, can be viewed on a PC by using a free to download player. Some of its technology was used in the development of MPEG-4
Subject	video format
Relation	Quicktime VR
	MPEG-4
	http://www.apple.com (Apple website)

video

Title	MPEG-4 • Very-low bitrate audio-visual coding
Creator	Moving Pictures Expert Group (ISO/IEC JTC1/SC29 WG11)
Publisher	International Organization for Standardization (ISO)
Date	1999 (version 1) 2001 (version 2)
Identifier	ISO/IEC 14496 (Parts 1 to 10)
Rights	[Open Standard]
Description	A high compression version of MPEG-2. Version 2 has data protection and IPR.
Subject	video format
Relation	MPEG-2 MP3 QuickTime [used in development] http://www.iso.org (ISO website)

Title	MPEG-1 • Coding of Moving Pictures and Associated Audio for Digital Storage Media
Creator	Moving Pictures Expert Group (ISO/IEC JTC1/SC29 WG11)
Publisher	International Organization for Standardization (ISO)
Date	1993
Identifier	ISO/IEC 11172:1993 (Parts 1 to 5)
Rights	[Open Standard]
Description	Designed to be the equivalent of a video recorder format in the digital world, and to make use of the early model CD-ROMs as a delivery method. Standard television quality images, with a compression ratio of 50 to 1.
Subject	video format
Relation	MP3 http://www.iso.org (ISO website)

MPG-2

Title	MPEG-2 • Coding of Moving Pictures and Associated Audio for Digital Storage Media
Creator	ITU-T (International Telecommunication Union Telecommunication Standardization Sector) Moving Pictures Expert Group (ISO/IEC JTC1/SC29 WG11)
Publisher	International Organization for Standardization (ISO)
Date	2000
Identifier	ISO/IEC 13818:2000 (Parts 1 to 11)
Rights	[Open Standard]
Description	An improvement to MPEG-1, with encoding techniques to allow for higher quality video and audio, and delivery from DVDs. High definition television quality images, with a compression ratio of 150 to 1.
Subject	video format
Relation	MPEG-1 MP3 http://www.iso.org (ISO website)

video

RM (Real Media)

Title	RM • Real Media
Creator	RealNetworks Incorporated
Publisher	RealNetworks Incorporated
Date	[?]
Identifier	[not available]
Rights	Copyright Real Networks
Description	Format optimised for delivery of audio over the Web
Subject	multimedia
Relation	RealAudio
Relation	RealVideo

video

SWF (Flash Movie)

Title	SWF • Small Web Format Flash Movie
Creator	Macromedia (now Adobe)
Publisher	Adobe Systems Incorporated
Date	1996 onwards
Identifier	http://www.adobe.com/devnet/swf/pdf/swf_file_format_spec_v10.pdf [version 10]
Rights	Copyright Adobe Systems Incorporated
Description	Originally just for the delivery of animated vector graphics it is now used for interactive audio and video. It can be viewed using a standalone player or via a web browser plug in.
Subject	animated vector graphics Interactivity

video

WMV (Windows Media Video)

Title	WMV • Windows Media Video
Creator	Microsoft Corporation
Publisher	Microsoft Corporation
Date	[?]
Identifier	[not available]
Rights	Copyright Microsoft Corporation
Description	Microsoft's proprietary competition to MPEG-4. Optimised to deliver video over the Web, particularly streaming, and using Microsoft products.
Subject	video format
Relation	WMA

video

Title	ASF • Advanced Streaming Format
Creator	Microsoft Corporation RealNetworks Incorporated
Publisher	Microsoft Corporation
Date	2004
Identifier	http://www.microsoft.com/windows/windowsmedia/forpros/format/asfspec.aspx [Revision 01.20.03]
Rights	Copyright Microsoft Corporation. All rights reserved.
Description	A proprietary format, designed to deliver, compressed, streaming video / audio content over the Internet.
Subject	video format

Title	VRML97 • Virtual Reality Modelling Language
Creator	VRML Consortium Incorporated Joint Technical Committee ISO/IEC JTC 1, Information technology, Subcommittee 24, Computer graphics and image processing
Publisher	International Organization for Standardization (ISO)
Date	1997
Identifier	ISO/IEC 14772-1:1997
Rights	[Open Standard]
Description	A format for describing 3D objects and worlds. These are designed to be interactive and accessible over the Internet as well as in other scenarios (e.g. standalone computers). Is capable of representing static or dynamic 3D and allowing links to other multimedia 'object' such as text, still images, audio and video.
Subject	virtual reality
Relation	X3D http://www.web3d.org (Web3D Consortium website) http://www.iso.org (ISO website)

X3D / EXTENSIBLE 3D

Title	X3D • Extensible 3D
Creator	Web3D Consortium
Publisher	International Organization for Standardization (ISO) [in process]
Date	2007
Identifier	ISO/IEC FDIS 19775-1.2:2008
Rights	[Open Standard]
Description	An XML-based format expressing the functionality of VRM97. Currently being developed as an ISO standard.
Subject	virtual reality
Relation	VRML97
	XML
	http://www.web3d.org (Web3D Consortium website)
	http://www.iso.org (ISO website)

virtual reality

QUICKTIME VR

Title	QuickTime VR
Creator	Apple Computer Incorporated
Publisher	Apple Computer Incorporated
Date	1995 onwards
Identifier	[not available]
Rights	Copyright Apple Computer Incorporated
Description	Apple's proprietary virtual reality format and system for creation of content. Built into the Mac's operating system, can be viewed on a PC by using a free to download player. Part of the QuickTime system, it 'stitches' individual photographs together to create panoramas, 3-D views of objects, and allows the linking of them into 'scenes'.
Subject	virtual reality panoramas
Relation	QuickTime http://www.apple.com (Apple website)

virtual reality

Title	EPS • Encapsulated Postscript
Creator	Adobe Systems Inc.
Publisher	Adobe Systems Inc.
Date	1992
Identifier	http://partners.adobe.com/public/developer/en/ps/5002.EPSF_Spec.pdf [Version 3.0]
Rights	[Open Standard]
Description	Although strictly speaking EPS is not an image format is often used to represent line drawings. EPS is based on the PostScript language, a page layout language often used in printers. The EPS version allows PostScript encoded layouts to be “encapsulated” in other documents.
Subject	page layout language vector graphics

vector graphics

Title	SVG • Scalable Vector Graphics
Creator	World Wide Web Consortium (W3C)
Publisher	World Wide Web Consortium (W3C)
Date	2003
Identifier	http://www.w3.org/TR/SVG/index.html [Version 1.1]
Rights	[Open Standard]
Description	SVG is a language for describing two-dimensional graphics in XML.SVG allows for three types of graphic objects: vector graphic shapes (e.g., paths consisting of straight lines and curves), images and text. Graphical objects can be grouped, styled, transformed and composited into previously rendered objects. Although primarily intended for vector applications SVG documents can also contain bitmap images.
Subject	vector graphics
Relation	XML

vector graphics

2.4 Other technical standards

These deal with various areas:

- **Search and retrieval** – The intricacies of formulating a query for a database system and obtaining (and handling) the search results.
- **Transmission** – To achieve the end-to-end delivery of digital data. They can be low-level (such as TCP/IP) or be more application specific, such as http. Today TCP/IP is the predominant low level protocol. Application specific protocols are constructed ‘on-top-of’ TCP/IP.
- **Character encoding** – Defining how separate characters in text are stored in a computer system.

Title	OAI-PMH • Open Archives Initiative Protocol for Metadata Harvesting (Version 2.0)
Creator	Sompel, Herbert van de Lagoze, Carl
Publisher	Open Archives Initiative
Date	2002
Identifier	http://www.openarchives.org/OAI/openarchivesprotocol.html
Rights	[Open Standard]
Description	The Open Archives Initiative metadata harvesting protocol provides access for harvesting programs to data stored in databases or repositories that cannot be harvested using 'standard' http/html parsing. The protocol is based on http requests and responds in the form of XML. Different record syntaxes can be used, but Dublin Core support is mandatory.
Subject	harvesting protocol
Relation	XML Dublin Core

search and retrieval

Title	SQL • Structured Query Language
Creator	International Organization for Standardization (ISO)
Publisher	International Organization for Standardization (ISO)
Date	1987
Identifier	ISO 9075
Rights	[Open Standard]
Description	SQL defines a query language for relational databases. It was developed by IBM during the 1970's. SQL went through a standardisation process during the 1980's. The 2nd version of SQL (SQL2 was standardised in 1992). The syntax is based on operations on two-dimensional tables which form the basis for relational database systems.
Subject	query language relational databases
Relation	ODBC http://www.iso.org (ISO website)

search and retrieval

Title	Z39.50 • Information and documentation — Information retrieval (Z39.50) — Application service definition and protocol specification
Creator	ANSI/NISO
Publisher	International Organization for Standardization (ISO)
Date	1988
Identifier	ISO 23950:1998
Rights	[Open Standard]
Description	The Z39.50 protocol is protocol to implement search and retrieval client-server applications. Queries are expressed in Reversed Polish Notation (RPN) form and records can be returned in different formats, including MARC (Machine Readable Catalogue) format. The search component deals with the construction and execution of a query, the retrieval component of the standard deals with handling the search results.
Subject	search and retrieval protocol
Relation	MARC http://www.iso.org (ISO website)

search and retrieval

Title	TCP/IP • Transmission Control Protocol/Internet Protocol
Creator	Internet Engineering Task Force
Publisher	World Wide Web Consortium (W3C)
Date	1981
Identifier	RFC793 RFC1122
Rights	[Open Standard]
Description	The Transport Control Protocol (TCP) and Internet Protocol (IP) go hand in hand and form the basis of all data transport over the Internet. The IP protocol provides a basic 'datagram' service. No guarantees are built in regarding the correct delivery of the data. IP is a so-called connectionless service. TCP runs on top of IP to provide error free, guaranteed-delivery connections. TCP is (unlike IP) on connection oriented protocol.
Subject	Internet data transmission
Relation	HTTP FTP

transmission

Title	HTTP • HyperText Transfer Protocol
Creator	Berners-Lee, Tim
Publisher	World Wide Web Consortium (W3C)
Date	1991
Identifier	http://www.w3.org/Protocols/
Rights	[Open Standard]
Description	The hypertext transfer protocol runs 'on-top-of' TCP/IP and is intended for the implementation of requesting (hypertext) documents over an Internet connection. A set of parameters can be passed within an http request, making it useable for other high level protocols, such as the OAI protocol.
Subject	hypertext transfer
Relation	TCP/IP OAI

transmission

Title	FTP • File Transfer Protocol
Creator	Neigus, Nancy J.
Publisher	Internet Engineering Task Force
Date	1973
Identifier	http://www.ietf.org/rfc/rfc542.txt
Rights	[Open Standard]
Description	The FTP protocol is a machine / operating system independent protocol for the transfer of files. Its origins are from the beginning of the Internet and the FTP protocol can be considered one of the foundations of the net.
Subject	file transfer
Relation	TCP/IP

transmission

2.4.3 Character encoding

ASCII

Title	ASCII • American Standard Code for Information Interchange
Creator	ANSI American National Standards Institute
Publisher	ANSI
Date	1967
Identifier	[?]
Rights	[Open Standard]
Description	A 7-bit code to represent characters, such as letters and digits in computer systems. The original ASCII standard was created in 1963, but was replaced by its final version in 1967. The ASCII character set contains 128 different characters of which 95 are 'printable' and 33 are 'control characters'
Subject	character encoding
Relation	ISO 8859-1 Unicode

character encoding

ISO 8859-1 / ISO LATIN 1

Title	ISO 8859-1 • ISO Latin 1
Creator	International Organisation for Standardization (ISO)
Publisher	International Organization for Standardization (ISO)
Date	1998
Identifier	ISO-8859-1 ECMA-94
Rights	[Open Standard]
Description	For the representation of European languages the 95 printable characters from the ASCII character set were not sufficient. To solve this problem 8 bit versions were created notably by the European Computer Manufacturers Association (ECMA-8) and IBM. IBM introduced different <i>code pages</i> for the different languages. The 8-bit ECMA standard was later adopted by the International Organization for Standardization (ISO) under the registration ISO-8859-1.
Subject	character encoding
Relation	ASCII Unicode http://www.iso.org (ISO website)

character encoding

Title	Unicode
Creator	Unicode Incorporated
Publisher	International Organization for Standardization (ISO)
Date	1992
Identifier	ISO/IEC 10646 ECMA-94
Rights	[Open Standard]
Description	A 16-bit code to represents characters, such as letters and digits in computer systems. The Unicode standard was created to overcome the limitations of 8 bit character sets and to form one universally usable encoding scheme for characters from the Western languages, but also including non-Western scripts, such as Chinese, Arabic, Cyrillic and Hebrew. Version 3.0 of Unicode defines 49,194 characters. Different encoding schemes exist for Unicode including the widely used variable length UTF-8 encoding. Unicode is the default XML character set.
Subject	character encoding non-Western scripts
Relation	ASCII XML http://www.iso.org (ISO website)

character encoding

3. Conclusions

3.1 Cultural metadata standards

There are a limited number of key standards, and they are used extensively throughout Europe and indeed the world. These are often suggested as best practise but, from the evidence of the survey, there is still a long way to go to achieve interoperability. National standards in some countries are also a factor that needs to be taken into consideration for future work.

Dublin Core (DC) has been used extensively as a 'minimum' metadata standard for resource discovery, most notably with Europeana itself. However relationships between DC and the domain-specific metadata standards need to be defined and interchange enabled in some way. XML and in particular the XML-related XSLT (Extensible Stylesheet Language Transformations) may offer a solution to this.

XSLT is used for the transformation of one kind of XML document into a kind of XML document (or indeed to a human readable form). So it should be possible to transform data in a domain-specific metadata standard to Dublin Core. However it is unlikely that it will possible to transform data from DC to a domain-specific metadata form, except for a limited number of elements.

3.2 Technical standards

There are a great many technical standards for virtually all multimedia resources including many open standards.

For good an up-to-date general guide to which standards a cultural organisation should use one need not look further than the Minerva Project's:

Technical Guidelines for Digital Cultural Content Creation Programmes:

<http://www.minervaeurope.org/interoperability/technicalguidelines.htm>
[with links to various versions]

The Guidelines have many advantages.

They are:

- **Multilingual** – Having a guide in the working language of the organisation is a great advantage for their understanding and adoption. Further translation should be encouraged.
- **Written for a general cultural sector audience** – Purely technical guidelines are a barrier to the general audience, which most people in the cultural sector are in this area.
- **Updated** – Is it very important that technical guidelines are kept up-to-date, especially in the rapidly changing IT area.

4. Standards index

60	AIFF
89	ASCII
76	ASF
65	AU
67	AVI
53	BMP
25	CDWA
40	CIDOC-CRM
45	DJVU
54	DJVU
42	DUBLIN CORE
29	EAD
80	EPS
68	FLV (Flash Video Format)
32	FRBR
88	FTP
55	GIF
46	HTML
87	HTTP
31	ISAAR (CPF)
30	ISAD (G)

35	ISO 2709	48	RTF
90	ISO 8859-1 / ISO LATIN 1	49	SGML
56	JPG	28	SPECTRUM
33	MAB2	84	SQL
34	MARC	81	SVG
36	METS	74	SWF (Flash Movie)
38	MIDAS	86	TCP/IP
37	MODS	43	TEI
69	MOV (Quicktime)	58	TIFF
61	MP3	91	UNICODE
70	MP4	41	VRA
71	MPEG1	77	VRML97
72	MPEG2	62	WAV
26	MUSEUMDAT	63	WMA
83	OAI-PMH	75	WMV (Windows Media Video)
27	OBJECT ID	50	WORD
47	PDF	78	X3D / EXTENSIBLE 3D
57	PNG	51	XML
79	QUICKTIME VR	85	Z39.50
64	REALAUDIO		
73	RM (Real Media)		

Printed in the month
of October 2009
by La Tipografia
di Umbero Frisardi,
Roma, Italy