Project no. 033104

MultiMatch

Technology-enhanced Learning and Access to Cultural Heritage
Instrument: Specific Targeted Research Project
FP6-2005-IST-5

D2.1 First Analysis of Metadata in the Cultural Heritage Domain

Start Date of Project: 01 May 2006
Duration: 30 Months

Netherlands Institute for Sound and Vision

Version: Final

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)

Document Information

Deliverable number:	D2.1
Deliverable title:	First Analysis of Metadata in the Cultural Heritage Domain
Due date of deliverable:	October 2006
Actual date of deliverable:	23 October 2006
Author(s):	Johan Oomen MA, Hanneke Smulders
Participant(s):	Alinari, ISTI-CNR, Netherlands Institute for Sound and Vision, University of Sheffield
Workpackage:	WP2
Workpackage title:	Content Selection and Preparation
Workpackage leader:	Netherlands Institute for Sound and Vision
Est. person months:	7.9
Dissemination Level:	PU
Version:	Final
Keywords	[metadata, metadata schemas, controlled vocabularies, knowledge representation, cultural heritage]

Abstract

This deliverable provides an overview of current practice regarding knowledge representation in the cultural heritage domain. It does so by providing an overview of the metadata schemas and controlled vocabularies that are widely used in the cultural heritage sector.

An overview of current practice is gathered from:

- the cultural heritage partners in the project;
- cultural heritage institutes throughout Europe;
- work done within other (research) projects.

More generic knowledge representation standards and the use of the Semantic Web within the project are outlined. This overview provides insight into the metadata schemas and controlled vocabularies MultiMatch might have to deal with and build upon.

The deliverable concludes with a first analysis of the most important schemas and reference models together with a preliminary outline of their possible usability in the MultiMatch project.

Note: in the Description of Work, the title of this deliverable is listed as "First Analysis of Ontologies in the CH domain". This title was too narrow to cover the work and thus was amended slightly.

Table of Contents

Executive Summary

This deliverable provides an overview of current practice regarding knowledge representation in the cultural heritage domain and defines the basis for the approach towards maximum interoperability that will be adopted within the MultiMatch project. The focus is thus on descriptive metadata; in other words, the metadata that identify and describe the object and what it expresses (see further section 1.3). This first analysis is intended to be general, with more specific analysis in later deliverables.

In Chapter 1, the cultural heritage domain is divided into the six sub-domains to be targeted in this study. The methodology used in gathering the information is explained, as well as selection criteria used. A scheme or vocabulary is included only if the following criteria are met:

- it is constructed and maintained by a renowned institute in one of the sub-domains *and*,
- available in electronic form *and*,
- publicly available; in other words, there may be financial but no copyright hindrances to apply them in MultiMatch *and*,
- it is proven an international standard *or* a local standard, in use nationwide.

Chapters 2, 3 and 4 give an insight into the metadata schemas and controlled vocabularies MultiMatch might have to deal with. Chapter 2 provides a descriptive overview of the metadata schemas and the semantic resources (i.e. thesauri, controlled vocabularies) widely used within the organizations belonging to the specific sub-domains. Forty have been identified and analyzed in a structured fashion.

	Schema	**Controlled vocabularies**
Archives	2	4
Libraries	3	7
Museums	3	5
Educational sector	2	-
Audiovisual sector	7	2
Geospatial sector	5	2

Chapter 3 provides information on the metadata used by some of the cultural heritage institutions within the consortium and the Advisory Board. It also lists seventeen European projects and initiatives that are closely related to MultiMatch, including the MICHAELplus and The European Library projects. Furthermore, it includes data from a relevant inventory on multilingualism conducted by the MINERVA Plus project and provides a summary of the use of controlled vocabularies in the cultural heritage domain.

From this survey it became clear that the uptake of international established controlled vocabularies is quite limited. Local and nationally established/managed vocabularies are therefore predominant. Part of the reason for this is that the available international controlled vocabularies are still not available in every European language (currently there are 20 official languages in the European Union).

We can note, however, that certain controlled vocabularies are particularly popular and have already been used in many European countries:

- Getty Arts and Architecture Thesaurus
- The UNESCO thesaurus
- Library of Congress Subject Headings (LCSH)
- The HEREIN thesaurus
- The NARCISSE vocabulary and the EROS project
- ICONCLASS (in the field of iconographic description).

Chapter 4 describes some generic knowledge representations and several metadata schemas, ontologies and reference models that are used in various contexts, not only within the cultural heritage domain.

Generic identification standards and reference models	• CIDOC Conceptual Reference Model • Digital Object Identifier • Functional Requirements for Bibliographic Records • SKOS Simple Knowledge Organisation System • RDF Resource Description Framework
Generic Metadata Schema	• Dublin Core Metadata Initiative • MPEG-7 • MPEG-21

Chapter 4 concludes explaining the relationship between the goals of MultiMatch and the Semantic Web (SW). Here it is noted how much of the technology examined in MultiMatch will consider issues relevant to the development of the Semantic Web. Thus the project should both add to and benefit from SW technologies and research, and provide tools and materials which are exploitable in the context of the Semantic Web.

As part of MultiMatch, documents, within the Cultural Heritage domain, will be marked-up with semantic information (or metadata) from a common vocabulary. One criticism leveled at the SW is the cost associated with providing this markup; the project will examine the use of classification and information extraction techniques to alleviate this problem. The SW is also concerned with the interoperability between different vocabularies (and ontologies); an issue which will have to be addressed within MultiMatch as well. There are also issues which relate to the SW, such as "trust" and the provenance of information, privacy and censorship and the provision of Web services which, whilst not central, will be examined in the project.

The fifth and final chapter of this deliverable summarises the most relevant standard(s) for each sub-domain.

	Schema	Controlled vocabularies
Archives	EAD and ISAD(G)	IPTC thesaurus, ISAAR (CPF), Thésaurus architecture et patrimoine, UK Archival Thesaurus
Libraries	FRBR, MARC, MODS and METS	DDC, UDC, LCSH and RAMEAU
Museums	CDWA, Object ID, VRA	AAT, ULAN, TGN
Educational sector	IEEE LOM	ERIC thesaurus
Audiovisual sector	P_META and SMEF-DM	-
Geospatial sector	CSDGM and ISO 19115:2003	-

It also gives a preliminary indication of the possible usability of these popular standards for MultiMatch. In the sections following, the most relevant generic schemas (Dublin Core, MPEG-7, MPEG-21) and reference models (FRBR, CIDOC-CRM) are analysed.

Next, the metadata schemas possibly relevant for MultiMatch are analysed according to a number of criteria (applying the analysis methodology from De Sutter et. al. [Sutter, 2006]), to provide a first typology of these schemas in a tabular overview.

The concluding paragraph outlines further research issues concerning knowledge representation within the project. In D2.2 the approach for knowledge representation in MultiMatch will be defined and described in detail. This deliverable, D2.1, thus represents the starting point for the further research needed to decide on the knowledge representation within MultiMatch.

1. Introduction

This 'First Analysis of Ontologies in the Cultural Heritage domain' will feed into the specification of the first prototype. The final approach regarding content interoperability will be defined in conjunction with work in WP1 and WP3 and (after internal papers) will form the core of Deliverable 2.2 (Content interoperability: metadata and file formats), to be released at PM10. This first analysis is thus intended to be general, with more specific analyses to be provided in later deliverables.

Workpackage 2 'Content Selection and Preparation' is closely linked to WP1 (User Requirements) and WP3 (System Architecture Design and System integration).

- **WP1** defines the user requirements after conducting interviews, log analyses and performing desk research. These requirements will provide pivotal input to arrive at a definitive approach regarding interoperability.
- Initial work in **WP3** deals with the detailed specifications of the first prototype. WP2, and more specifically task 2.1, will provide necessary input regarding issues connected with metadata, thesauri/ontologies, and semantic web encoding.

1.1 Outline of this Document

Deliverable 2.1 provides an overview of current practice regarding knowledge representation in the cultural heritage domain. As metadata standards enable interoperability between systems and organisations that information can be exchanged and shared, the overview in this deliverable provides the basis for the approach towards interoperability that will be adopted within the MultiMatch project.

The primary focus is on descriptive metadata, representing the conceptually meaningful aspects of an object, but some technical dimensions are also into account. Current practice in the diverse areas into which the cultural heritage domain can be broken down is investigated.

- In Chapter 1, the cultural heritage domain is divided into the six sub-domains to be targeted in this study. The methodology adopted and the terminology used are also explained in this introductory chapter.
- Chapter 2 provides a descriptive overview of the metadata schemas and the semantic resources (i.e. thesauri, controlled vocabularies) widely used within the organizations belonging to the specific sub-domains.
- Chapter 3 provides information on the metadata used by some of the cultural heritage institutions within the consortium and within related European projects. Chapter 3 also includes data from a relevant inventory multilingualism conducted by the MINERVA Plus project and provides a summary of the use of controlled vocabularies in the cultural heritage domain.
- Chapter 4 describes some generic knowledge representations and several metadata schemas, ontologies and reference models that are used in various contexts, not only the cultural heritage domain. These knowledge representations can play a role within the MultiMatch project. This chapter also explains the relationship between the goals of MultiMatch and the Semantic Web.
- The fifth and final chapter of this deliverable summarises the most relevant standard(s) for each sub-domain. This is done by looking at the uptake of standards in section 5.1. This section also gives a preliminary indication of the possible usability of these popular standards for MultiMatch. The most relevant generic schemas (Dublin Core, MPEG-7, MPEG-21) and reference models (FRBR, CIDOC-CRM) are then analysed.

Furthermore, the metadata schemas possibly relevant for MultiMatch are analysed according to four criteria (the analysis methodology from De Sutter et. al. [Sutter, 2006]), to provide a first typology of these schemas in a tabular overview in paragraph 5.2. The concluding paragraph of this deliverable outlines further research issues concerning knowledge representation within MultiMatch. In D2.2 (PM 10), the approach that MultiMatch will adopt for knowledge representation will be defined and described in detail.

1.2 Methodology

The focus of this deliverable is the current practice of knowledge representation in the cultural heritage sector. This survey provides the technical partners of the MultiMatch project with a clear view of the dimensions of the data they will have to deal with. It will feed into different other tasks, notable the functional specification of the prototype.

Furthermore, this deliverable will provide input for the decision on knowledge representation in the MultiMatch project (to be reported in D2.2).

The methodological approach can be broken down in three parts:

- Defining cultural heritage
- Information gathering process
- Selection Criteria

1.2.1 Defining Cultural Heritage

The concept Cultural Heritage can be defined in many ways. Here are just three examples.

> "It is the legacy of physical artefacts and intangible attributes of a group or society that are inherited from past generations, maintained in the present and bestowed for the benefit of future generations. Physical or "tangible cultural heritage" includes buildings and historic places, monuments. Natural heritage is also an important part of a culture, encompassing the countryside and natural environment. Smaller objects that are considered part of our cultural heritage are stored in libraries, museums and galleries. Cultural heritage objects are studied by academics and enjoyed by tourists; making it hard to draw boundaries." (Definition of Wikipedia)

> Europe's collective memory includes print (books, journals, newspapers), photographs, museum objects, archival documents, audiovisual material (hereinafter 'cultural material'). (Definition of Digicult[1])

> The term cultural heritage collections is intended to cover all types of material collected and displayed by museums and related institutions, as defined by ICOM. This includes collections, sites and monuments relating to natural history, ethnography, archaeology, historic monuments, as well as collections of fine and applied arts. (Definition of the International Council of Museums - ICOM[2])

[1] http://www.digicult.info/pages/index.php

[2] http://icom.museum/

In order to systematically study current practice we use the sub-domain definition advocated by the DEN (Digital Heritage Netherlands) and ePSINet (the European Public Sector Information Network[3]):

1. Archives
2. Libraries
3. Museums
4. Educational sector
5. Audiovisual sector
6. Geospatial sector

Clearly, there is significant overlap between these domains. In those cases in which it was unclear in which category an activity should be placed, a judgement was made based on a close examination of the schemas and semantic elements. Those controlled vocabularies that are used across these domains, are listed under the category 'generic' and described in Chapter 4.

1.2.2 Information gathering process

The methodology adopted for this first analysis of knowledge representation consisted of:

1. thorough desk research conducted on special interest groups and organisations working on this topic, as well as personal contacts provided us with the overview and insight presented below.
2. a questionnaire (see Appendix 1) to our target group: libraries, museums, archives and other cultural institutions participating in related European projects. We have sent the questionnaire to:
   - 17 partners of the BRICKS community
   - 6 members of the steering board of the Culture Mondo network
   - 14 partners of the Digital Heritage Network
   - 31 partners or members of the MINERVA project
3. consultation with experts in- and outside the consortium by telephone interviews.

1.2.3 Selection criteria

The selection of the knowledge representations in use is based on several criteria. A scheme or vocabulary is included if:

- it is constructed and maintained by a renowned institute in one of the sub-domains *and*,
- available in electronic form *and*,
- publicly available; in other words, there may be financial but no copyright hindrances to apply them in MultiMatch *and*,
- it is proven an international standard *or*
- a local standard, in use nationwide.

[3] http://www.epsigate.org/

1.3 Domain Terminology

Knowledge representation

This is a two sided concept:

1. Knowledge on cultural heritage objects is represented in metadata schemas (mainly in the semantic description of a cultural heritage object, not in the technical or administrative part of a metadata schema). Synonym: metadata model.

2. Knowledge on cultural heritage object is also represented in 'controlled vocabularies' or 'knowledge organization systems' of all kinds, therewith controlling the content of several metadata elements or attributes of a metadata schema.

Synonym: authority files.

Metadata

“a cloud of collateral information around a data object” Clifford Lynch (director of the Coalition for Networked Information).

A metadata record is a file of information, compiled (automatically and/or manually) in the format of the metadata schema concerned, which captures the basic characteristics of a data or information resource (e.g. a cultural heritage object). Metadata refers to “data about data”, in other words, information that describes information sources or objects, e.g. a Dublin Core record or a record from the catalogue of an archive. The format and structure of metadata is often dictated in a set of rules, called metadata schema.

Indirectly, the European Commission stressed the importance of metadata for online accessibility, in the 'Communication of 30 September 2005' on the Digital Libraries Initiative that deals with cultural heritage and its online preservation and accessibility.

"Questions of online accessibility are not limited to intellectual property rights. Putting material online does not mean it can be found easily by the user, still less that it can be searched and used. Appropriate services allowing the user to discover and work with the content are necessary. This implies structured and quality description of the content, both the collections and the items in them, and support for its use (e.g. annotation)." [4]

1. **Descriptive** metadata - mainly information to identify and describe the object or information source and what it expresses. These metadata include the author/title cataloguing as well as the subject indexing. In other words, the descriptive metadata include the subgroup of the objective elements that formally describe the object (e.g. identification number, title, creation date, creator name, the language of the object, physical media).

And the subgroup of semantic elements (also called analytical metadata) that contain information on the subject of the object to enhance access to the resource's contents (e.g. subject keywords, classification codes, abstract). Note, that the descriptive metadata, and especially the semantic elements are the scope of D2.1. Note also: descriptive metadata can be of a technical character, think of for instance 'compression schema' (this is the algorithm used to compress the audiovisual essence), the number of pages (book), black and white/colour (photograph, film) or specific information on the storage medium or carrier.

[4] http://europa.eu.int/information_society/activities/digital_libraries/doc/communication/en_comm_digital_libraries.pdf Last viewed September 14, 2006

2. **Technical** metadata - describe the technological characteristics of the related object (e.g. data that must be available to be able to use out the material, file locations, authentication and security information, characteristics needed for computer programming and database management)

3. **Administrative** metadata – metadata used in managing and administering the objects concerned (e.g. content provider name, acquisition information, copyrights, location information, language of record, record number).

Metadata schema

"Full, logically organised structure of relations between defined (groups) of metadata and the information objects they describe."[5]

“a set of rules for encoding information that supports specific communities of users.”[6]

A metadata schema consists of several metadata elements. For some elements the input is free (e.g. Title), for other elements the input is guided by syntactical rules or guidelines or even restricted by controlled vocabularies of all kinds (e.g. thesaurus for subject keywords or closed term list for object type).

Metadata element

A metadata element is an item, or an editorial part of metadata. A semantic metadata element is an element from the descriptive metadata that describes the cultural heritage object.

A metadata element name is given to a data element in, for example, a data dictionary or metadata schema or registry. In a formal data dictionary, there is often a requirement that no two data elements may have the same name, to allow the data element name to become an identifier, though some data dictionaries may provide ways to qualify the name in some way, for example by the application system or other context in which it occurs.

A data element definition is a human readable phrase or sentence associated with a data element within a data dictionary that describes the meaning or semantics of a data element.

Controlled vocabulary

A limited set of terms that must be used to index | represent | tag the subject matter | content of documents | objects (indexing tools in use to describe a cultural heritage object).

Examples: Alphabetic lists of “approved” words or phrases, thesauri, subject heading systems, classification schemes, ontologies, taxonomies.

These examples illustrate that controlled vocabularies are largely applied for subject keywords or generic concept identification. However, controlled vocabularies or lists of preferred terms are also applied for other metadata elements, e.g. person names like author or creator, names of historical people and corporate bodies on the cultural heritage object or as its subject of the cultural heritage object, geographic places (actual location of the cultural heritage object / place of creation / place where the cultural heritage object was found / place as subject of the cultural heritage object) and organisation names. *See also: Authority files in this table.*

[5] Metadata in the audiovisual production environment : an introduction / Annemieke de Jong. – Hilversum: Nederlands Instituut voor Beeld en Geluid, 2003

[6] Murtha Baca, Getty Research Institute

Classification schemes, Taxonomies and Categorization schemes[7]

These terms are often used interchangeably. Although there may be subtle differences from example to example, in general these types of knowledge representation provide ways to separate entities into buckets or relatively broad topic levels. Some examples provide a hierarchical arrangement of numeric or alphabetic notation to represent broad topics. These types of knowledge representation may not follow the strict rules for hierarchy required in the ANSI NISO Thesaurus Standard (Z39.19) (NISO), and they lack the explicit relationships presented in a thesaurus.

Examples of classification schemes include the Library of Congress Classification Schedules (an open, expandable system), the Dewey Decimal Classification (a closed system of 10 numeric sections with decimal extensions), and the Universal Decimal Classification (based on Dewey but extended to include facets). Subject categories are often used to group thesaurus terms in broad topic sets, outside the hierarchical scheme of the thesaurus. Taxonomies are increasingly being used in object oriented design and knowledge management systems to indicate any grouping of objects based on a particular characteristic. "Taxonomy" may also refer to a scheme that presents subject elements in a hierarchical arrangement based on some characteristic.

Thesauri

These knowledge organization systems are based on concepts, and they show relationships between terms. Relationships commonly expressed in a thesaurus include hierarchy, equivalence, and associative (or related).
These relationships are generally represented by the notation BT (broader term), NT (narrower term), SY (synonym), and RT (associative or related). There are standards for the development of monolingual thesauri (NISO, 1998; ISO, 1986) and multi-lingual thesauri (ISO, 1985).

It should be noted that the definition of a thesaurus in these standards is often at variance with schemes that are actually called thesauri. There are many thesauri that do not follow all the rules of the standard, but are still generally thought of as thesauri. Many thesauri are very large (more than 50,000 terms). Most were developed for a specific discipline, or to support a specific product or family of products.

Subject headings

This scheme provides a set of controlled terms to represent the subjects of items in a collection. Subject heading lists can be extensive, covering a broad range of subjects. However, the subject heading lists structure is generally very shallow, with a limited hierarchical structure. In use, subject headings tend to be pre-coordinated, with rules for how subject headings can be joined to provide more specific concepts. Examples include the Medical Subject Headings (MeSH) and the Library of Congress Subject Headings (LCSH).

Authority files

Authority files are lists of terms that are used to control the variant names for an entity or the domain value for a particular field. Examples include names for countries, individuals, and organizations. Non-preferred terms may be linked to the preferred versions. This type of knowledge organization generally does not include a deep organization or complex structure. The presentation may be alphabetical or organized by a shallow classification scheme.

There may be some limited hierarchy applied in order to allow for simple navigation, particularly when the authority file is being accessed manually or is extremely large.

---

[7] For the definitions of the several types of controlled vocabularies the following source is used: Taxonomy of Knowledge Organization Sources/Systems (1). - Draft June 7, 2000 (revised July 31, 2000)

http://nkos.slis.kent.edu/KOS_taxonomy.htm Last viewed 2006-09-14.

Specific examples of authority files include the Library of Congress Name Authority File and the Getty Geographic Authority File.

Semantic network

With the advent of natural language processing, there have been significant developments in the area of semantic networks. These knowledge organization systems structure concepts and terms not as hierarchies but as a network or a Web. Concepts are thought of as nodes with various relationships branching out from them.

The relationships generally go beyond the standard BT, NT and RT. They may include specific whole-part relationships, cause-effect, parent-child, etc. One of the most noted semantic network is Princeton's WordNet, which is now used in a variety of search engines.

Ontology

An ontology is a data model that represents the existing knowledge within a domain and is used to reason about the objects in that domain and the relations between them. Ontologies are used as a form of knowledge representation about the world or some part of it. Ontologies generally describe: Individuals (the basic or "ground level" objects); Classes (sets, collections, or types of objects); Attributes (properties, features, characteristics, or parameters that objects can have and share); Relations (ways that objects can be related to one another).[8]

Therefore thesauri and classification schemes can be regarded as ontologies with a relatively little number of relationships.

Ontologies can represent complex relationships between objects, and include the rules and axioms missing from semantic networks. Ontologies that describe knowledge in a specific area are often connected with systems for data mining and knowledge management.

Upper Ontology (top-level ontology, or foundation ontology). An attempt to create an ontology which describes very general concepts that are the same across all domains. The aim is to have a large number on ontologies accessible under this upper ontology.

Markup ontology languages These languages use a markup scheme to encode knowledge, most commonly XML. (SHOE, XOL, DAML+OIL, OIL, RDF, RDF Schema, OWL)

Semantic Web

The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. It is based on the Resource Description Framework (RDF), which integrates a variety of applications using XML for syntax and URIs for naming.

The Semantic Web intent is to enhance the usability and usefulness of the Web and its interconnected resources. Within MultiMatch the use of a Semantic Web-compatible markup will guarantee a rich use (mainly in retrieval functionality) of the metadata on cultural heritage objects provided by the partners in combination with several **ontologies** related to the cultural heritage domain. A domain ontology (or domain-specific ontology) models a specific domain, or part of the world. An ontology on arts can be used to say, for instance that “Picasso” is a “Painter”, and that a “Painter” is an “Artist”. The combination of such ontologies together with the MultiMatch indexes automatically provides the end user with several extra ways to navigation through the MultiMatch collection. E.g. this combination can present all cultural heritage objects from museums in Spain,

[8] Definition taken from: www.wikipedia.org

without the need for the content providing partners to manually add extra metadata to the descriptions of their objects. See also paragraphsection 4.3.

XML schema

An XML schema is a description of a type of XML document, typically expressed in terms of constraints on the structure and content of documents of that type, above and beyond the basic syntax constraints imposed by XML itself. An XML schema provides a view of the document type at a relatively high level of abstraction. There are languages developed specifically to express XML schemas. The Document Type Definition (DTD) language, which is native to the XML specification, is a schema language.

Data model

"A data model is a model that describes in an abstract way how data are represented in a business organization, an information system or a database management system. This term is ambiguously defined to mean:

1. **how data generally are organized**, e.g. as described in Database management system. This is sometimes also called "database model"
2. **or how data of a specific business function are organized logically** (e.g. the data model of some business)

While simple data models consisting of few tables or objects can be created "manually", large applications need a more systematic approach. Within the relational database modelling community, the entity-relationship model method is used to establish a domain-specific data model. In computer science, an entity-relationship model (ERM) is a model providing a high-level description of a conceptual data model. Data modelling provides a graphical notation for representing such data models in the form of entity-relationship diagrams (ERD).

A conceptual schema, or high-level data model or conceptual data model, is a map of concepts and their relationships, for example, a conceptual schema for a karate studio would include abstractions such as student, belt, grading and tournament."[9]

In this deliverable, data models are referred to as reference models, see also paragraphsection 4.1. A data model, especially the concepts or entities and relationships of the model, dictate the metadata elements that are needed in the metadata schema that goes along with the data model.

[9] www.wikipedia.org

2 Knowledge Representation in the Cultural Heritage Domain

Knowledge representation in the cultural heritage domain includes metadata schemas on the one hand, and semantic element definitions (i.e. thesauri, controlled vocabularies) on the other. See also section 1.3 for further definitions.

In order to provide a descriptive overview of metadata in the cultural heritage domain, this chapter presents, for each sub-domain, a selection of metadata schemas and of controlled vocabularies. The selection of the knowledge representations in use is based on several criteria, listed in section 1.2.3. To start with, some generic standards are described. The subsequent descriptions of the selected knowledge representations appear in alphabetical order, for each sub-domain.

2.1 Generic Standards

The following tables provide an overview of generic metadata standards. The selection consists of: Friend Of A Friend, Wiktionary and WordNet.

Friend Of A Friend

Name	Friend of a Friend
Acronym	FOAF
Status / version	Not available
Type	Standard
Management	Edd Dumbill Editor and publisher, xmlhack.com
Short description	FOAF is a domain-specific vocabulary to support the social interactions of humans within the general Web. It provides a vocabulary for describing the kind of information that is found on people's home pages in a machine-understandable fashion, e.g. "My name is", "I am interested in" and "You can see me in this picture". This allows queries to be made over communities of people, e.g. "Show me pictures of people who are interested in Marilyn Manson who live near me."
URL(s) documentation	http://rdfweb.org/topic/FAQ Available at 2006-06-21 http://www.foaf-project.org/ Available at 2006-06-21
URL guidelines for application	http://www-106.ibm.com/developerworks/xml/library/x-foaf.html Viewed 2006-09-26
XML encoding available	Yes (also RDF, Semantic Web)

Wiktionary

Name	The EnglishWiktionary
Acronym	Wiktionary
Status / version	20060704
Type	Standard
Management	Wikimedia
Short description	A collaborative project to produce a free, multilingual dictionary with definitions, etymologies, pronunciations, sample quotations, synonyms, antonyms and translations. Wiktionary is the lexical companion to the open-content encyclopedia Wikipedia. The English Wiktionary aims to describe all words of all languages, with definitions and

	descriptions in English only. For example, see Wörterbuch (a German word). In order to find a German definition of that word, visit the equivalent page in the German Wiktionary.
Number of elements	290,688 entries
Available in language	124 languages
XML encoding available	No
URL(s) documentation	http://en.wiktionary.org/wiki/Main_Page Viewed 2006-10-19

WordNet

Name	WordNet
Acronym	WordNet
Status / version	Version 2.1
Type	Semantic lexicon
Management	Princeton University
Short description	WordNet is not a controlled vocabulary in the sense of a set of preferred terms, but it is an online lexical reference system whose design is inspired by current psycholinguistic theories of human lexical memory. English nouns, verbs, adjectives and adverbs are organized into synonym sets, each representing one underlying lexical concept. Different relations link the synonym sets. WordNet is considered to be the most important resource available to researchers in computational linguistics, text analysis, and many related areas.
Number of elements	155,327 unique strings ; 207,016 word-sense pairs
Available in language	English only. However, the Mimida Project[10], developed by Maurice Gittens, is a WordNet-based mechanically-generated multilingual semantic network for more than 20 languages based on dictionaries found on the Web.
XML encoding available	No
Extra information on application	**MultiWordNet**[11], developed by Luisa Bentivogli and others is a multilingual lexical database, developed at ITC-irst, in which the Italian WordNet is strictly aligned with Princeton WordNet 1.6. The current version includes around 44,400 Italian lemmas organized into 35,400 synsets which are aligned, whenever possible, with their corresponding English Princeton synsets. The MultiWordNet database can be freely browsed through its on-line interface, and is distributed both for research and commercial use. Information on the distribution licence is available at the web site. **EuroWordNet**[12] is a multilingual database with wordnets for several European languages (Dutch, Italian, Spanish, German, French, Czech and Estonian). The wordnets are structured in the same way as the American WordNet for English (Princeton WordNet, Miller et al 1990) in terms of synsets (sets of synonymous words) with basic semantic relations between them.
URL(s) documentation	http://wordnet.princeton.edu/ Viewed 2006-10-02.
URL guidelines for application	http://wordnet.princeton.edu/man/wnintro.3WN (the API documentation) http://wordnet.princeton.edu/doc (reference manual WordNet 2.1) Viewed 2006-10-19

[10] http://www.gittens.nl/SemanticNetworks.html

[11] http://multiwordnet.itc.it/english/home.php

[12] http://www.illc.uva.nl/EuroWordNet/

2.2 Archives

An archive refers to a collection of records, and also refers to the location in which these records are kept. Archives are made up of records which have been created during the course of an individual or organization's life. In general an archive consists of records which have been selected for permanent or long-term preservation. Records, which may be in any media, are normally unpublished, unlike books and other publications.

2.2.1 Metadata schemas

The following tables provide an overview of the selected metadata schemas used by archives. The selection consists of: Encoded Archival Description and General International Standard Archival Description.

Encoded Archival Description

Name	Encoded Archival Description
Acronym	EAD
Status / version	Version 2002
Type	International standard
Management	The standard is maintained in the Network Development and MARC Standards Office of the Library of Congress (LC) in partnership with the Society of American Archivists.
Short description	The EAD Document Type Definition (DTD) is a standard for encoding archival finding aids using Extensible Markup Language (XML). Finding aids are indexes used to catalogue detailed information about collections within an archive. They are used by researchers to determine whether information within a collection is relevant to their research. Finding aids often describe the scope of the collection, biographical and historical information related to the collection, and access details. Finding aids can created in various electronic and print formats. The standard format for finding aids is Encoded Archival Description. EAD defines the structural elements and designates the content of descriptive guides to archival and manuscript holdings. It is intended to provide standardized, digital description of archival and manuscript collections and facilitate uniform, on-line, Web-based access to the detailed information about primary research materials held in repositories worldwide. It provides tools for a detailed, multilevel description, structured display, navigation, and searching. Archives and libraries can use EAD to XML-encode the information in their finding aids for greater online access.
Syntaxes	In principle, encoded finding aids consist of three parts, the first describing the information about the finding aid itself (<**eadheader**>), the second describing the prefatory matter useful for the display or publication of the finding aid (<**frontmatter**>), and the third one containing the description of the archival records or manuscript papers (<**archdesc**>). The Document Type Definition defines document structure, while elements constitute informational units. Elements can be modified with attributes. EAD presentation (display) is prescribed using style sheets - separate files controlling presentation of data (text layout and format). Style sheets can also supply default text and images.
Extra information on application	Effectively an organized presentation of a collection of documents (typically in an archive or manuscript collection) ➢ EAD header carries metadata for the finding aid

	➢ Provides for simple or complex mark-up to support varying levels of indexing Well-suited for interweaving narrative with links to specific objects in a collection (either directly to the object or via a record for the object that may link to the object).[13] Dublin Core is mapped to EAD, USMARC and ISAD(G) are mapped to EAD, EAD is also mapped to ISAD(G).[14]
Applied by the following organizations e.g.	Widely used by academic institutions and archives in North America. EAD was the basis of the Research Library Group (RLG) Archival Resources database, which included close to 50,000 finding aids together with briefer collections cataloguing.
URL(s) documentation	http://www.loc.gov/ead/ http://www.loc.gov/ead/ead2002a.html
URL guidelines for application	http://www.rlg.org/en/page.php?Page_ID=411 http://www.rlg.org/rlgead/guidelines.html Viewed 2006-10-19
XML encoding available	Yes

General International Standard Archival Description

Name	General International Standard Archival Description
Acronym	ISAD(G)
Status / version	2nd edition 1999
Type	Recommendation
Management	International Council on Archives
Short description	Rules in order to make standardised multilevel descriptions for archives. Aiming at presenting the context and hierarchical structure of an archive.
Number of elements	26 elements grouped into 7 entities: identification, context, content & structure, conditions for consultation and lending, related material, notes and description management.
Vocabularies proposed	International Standard Archival Authority Record for Corporate Bodies, Persons and Families: ISAAR(CPF). This standard provides general rules for the construction of authority files for the metadata element 'archive builder' (a syntax for names of organisations, persons and families). See section 2.3.2 - ISAAR.
Extra information on application	ISAD(G) is mapped to EAD and vice versa. http://www.getty.edu/research/conducting_research/standards/intrometadata/crosswalks.html
Applied by the following organizations e.g.	Stadsarchief Antwerpen
URL(s) documentation	http://www.ica.org/biblio/isad_g_2e.pdf Viewed 2006-10-19
XML encoding available	No

[13] Metadata standards / Eric Childress. Presentation for FEDLINK OCLC Users Group Meeting. November 18th 2003.

[14] http://www.getty.edu/research/conducting_research/standards/intrometadata/crosswalks.html

2.2.2 Controlled vocabularies

The following tables provide an overview of the selected controlled vocabularies used by archives. The selection consists of: IPTC thesaurus, International Standard Archival Authority Record, Thésaurus architecture et patrimoine and UK Archival thesaurus.

IPTC thesaurus

Name	IPTC Newscodes – subjectcode
Acronym	IPTC thesaurus
Status / version	Version 17, 2006-08-21
Type	International standard
Management	The International Press Telecommunications Council
Short description	A tree-structured list of thematic keywords. The IPTC Subject Reference System was developed to allow information providers access to a universal language independent coding system for indicating the subject content of news items.
Number of elements	Approximately 1,200 terms on all subject areas.
Available in language	Dutch, English, French, German
XML encoding available	Yes
Extra information on application	A three-level hierarchy where the top level is **Subject;** the second level is **Subject Matter** and the third level is **Subject Detail**. There are 17 top-level **Subjects,** and the IPTC has developed secondary **Subject Matter** lists for each of these. To date, there are third-level **Subject Detail** lists for three Subjects: Economy, Business and Finance, Politics, and Sport.
Applied by the following organizations e.g.	• News agencies and independent journalists worldwide • BIRTH project (http://www.birth-of-tv.org/birth/)
URL(s) documentation	http://www.iptc.org/NewsCodes/nc_ts-table01.php?TsByName=iptc-subjectcode Viewed 2006-10-19
URL guidelines for application	http://www.iptc.org/std/NewsCodes/0.0/documentation/SRS-doc-Guidelines_3.pdf Viewed 2006-10-19

International Standard Archival Authority Record

Name	International Standard Archival Authority Record for Corporate Bodies, Persons and Families
Acronym	ISAAR (CPF)
Status / version	Second edition, 2004
Type	International standard
Management	International Council on Archives
Short description	This standard provides guidance for preparing archival authority records which provide descriptions of entities (corporate bodies, persons and families) associated with the creation and maintenance of archives. The elements of description for an archival authority record are organized into four

	information areas: 1. Identity Area (where information is conveyed which uniquely identifies the entity being described and which defines standardized access points for the record) 2. Description Area (where relevant information is conveyed about the nature, context and activities of the entity being described) 3. Relationships Area (where relationships with other corporate bodies, persons and/or families are recorded and described) 4. Control Area (where the authority record is uniquely identified and information is recorded on how, when and by which agency the authority record was created and maintained).
Number of elements	Not available
Available in language	Dutch, English, French, Italian, Portuguese, Spanish, Welsh.
XML encoding available	No, but see below.
Extra information on application	This standard addresses only part of the conditions needed to support the exchange of archival authority information. Successful automated exchange of archival authority information over computer networks is dependent upon the adoption of a suitable communication format by the repositories involved in the exchange. Encoded Archival Context (EAC) is one such communications format which supports the exchange of ISAAR(CPF) compliant archival authority data over the World Wide Web. EAC has been developed in the form of Document Type Definitions (DTDs) in XML (Extensible Markup Language) and SGML (Standard Generalized Markup Language).
Applied by the following organizations e.g.	Widely
URL(s) documentation	http://www.ica.org/biblio.php?pdocid=144 Viewed 15-9-2006.

Thésaurus architecture et patrimoine

Name	Thésaurus architecture et patrimoine
Acronym	Thésaurus
Status / version	2000
Type	National standard, France
Management	Ministère de la culture et de la communication – La Médiathèque de l'architecture et du patrimoine
Short description	Monolingual thesaurus on the subject areas: urbanism, all sorts of architecture (religious, public, housing, industrial, artistic and commercial), parks and gardens, furniture (including religious furniture), music instruments ; scientific instruments and production machines and engines.
Number of elements	5,000 (June 2000)
Available in language	French
XML encoding available	No
Applied by the following organizations e.g.	The MICHAELplus project and many institutions in France.

URL(s) documentation	http://www.culture.gouv.fr/culture/inventai/presenta/bddinv.htm Viewed 2006-10-19

UK Archival Thesaurus

Name	UK Archival Thesaurus
Acronym	UKAT
Status / version	August 2004
Type	National standard, UK
Management	A Management Board consisting of personnel from the National Archives and the University of London Computer Centre (ULCC)
Short description	A subject thesaurus which has been created for the archive sector in the United Kingdom. It is a controlled vocabulary which archives can use when indexing their collections and catalogues. The backbone of UKAT is the UNESCO Thesaurus (UNESCO), a high-level thesaurus with terminology covering education, science, culture, the social and human sciences, information and communication, politics, law and economics. The UNESCO thesaurus is significantly enhanced to include terms of relevance to the archive community and its users.
Number of elements	19,698 terms: 6,356 inherited from the UNESCO Thesaurus, and 13,342 incorporated following editing.
Available in language	English
XML encoding available	Yes : UKAT data marked up using the SKOS-Core 1.0 RDF schema.
Applied by the following organizations e.g.	the MICHAELplus project
URL(s) documentation	http://www.ukat.org.uk/index.html Viewed 2006-10-19

2.3 Libraries

In the scope of this document, a library is defined as a collection of books and periodicals. It can refer to an individual's private collection, but more often it is a collection of information resources and services that is funded and maintained by a city or institution.

2.3.1 Metadata schemas

The following tables provide an overview of the selected metadata schemas used by libraries. The selection consists of: Machine Readable Cataloguing, Metadata Object Description Schema and Metadata Encoding and Transmission Language.

Machine Readable Cataloguing

Name	Machine Readable Cataloguing
Acronym	MARC21
Status / version	1996
Type	International standard
Management	Library of Congress
Short description	The MARC formats are standards for the representation and communication of bibliographic and related information in machine-readable form. Widely used within the Library domain, but rarely in other domains.
Number of elements	> 200 elements
Vocabularies proposed	• the MARC Code List for Organizations contains short alphabetic codes used to represent names of libraries and other kinds of organizations that need to be identified in the bibliographic environment (27.719 elements). • the country code list is made up of three parts: Part I: Name Sequence, Part II: Code Sequence, and Part III: Regional Sequence (12 regions). Furthermore the following controlled vocabularies are mentioned: • For names, one of the most widely used authority files is the Library of Congress Name Authority File (or LCNAF; http://authorities.loc.gov/). • For topics or geographic names, the most used subject authority file is the LCSH. There are many other subject heading lists, such as the *Sears List of Subject Headings* and the *Art and Architecture Thesaurus*.
Extra information on application	MARC 21 has been mapped to the following metadata standards: MODS ; Dublin Core; MARC Character Sets to UCS/Unicode ; Digital Geospatial Metadata (FGDC) and vice versa. Unimarc is mapped to MARC21. The structure of MARC records is an implementation of national and international standards, e.g., *Information Interchange Format* (ANSI Z39.2) and *Format for Information Exchange* (ISO 2709).
Applied by the following organizations e.g.	Libraries worldwide
URL(s) documentation	http://www.loc.gov/marc/ http://www.loc.gov/cds/marcdoc.html
URL guidelines for application	Understanding MARC Bibliographic http://www.loc.gov/marc/umb/
XML encoding	Yes : a framework for working with MARC data in a XML environment is being

available	developed: http://www.loc.gov/marc/marcxml.html A list of some tools that work with HTML, SGML and XML applications is at http://www.loc.gov/marc/marctools.html Viewed 2006-10-19

Metadata Object Description Schema

Name	Metadata Object Description Schema
Acronym	MODS
Status / version	Version 3.2
Type	Recommendation
Management	Library of Congress
Short description	An XML schema for descriptive metadata, library-oriented, compatible with the MARC 21 bibliographic format, in other words: optimized for from-MARC conversion of legacy records. Well-suited as a metadata format for OAI harvesting. This schema may be used to carry selected data from a subset of existing MARC21 records as well as to enable the creation of original resource description records.
Vocabularies proposed	Lists for use with MODS: • Sources • Authority File • Classification • Form • Genre • Subject • Organizations • Target Audience • Relators and Roles Value lists • Relators and Roles (MARC) • Form (MARC) • Form (SMD) • Genre (MARC) • Target Audience (MARC) • Organization (MARC)
Extra information on application	There are crosswalks available to MARC and to Dublin Core and vice versa.
Applied by the following organizations e.g.	• OpenOffice Bibliographic Project • Minerva project • University of Chicago Press • California Digital Library • Library of Congres is planning to convert 100K American Memory records
URL(s) documentation	http://www.loc.gov/standards/mods http://www.loc.gov/standards/mods/v3/mods-3-2.xsd

URL guidelines for application	http://www.loc.gov/standards/mods/v3/mods-userguide.html Viewed 2006-10-19
XML encoding available	Yes

Metadata Encoding and Transmission Language

Name	Metadata Encoding and Transmission Language
Acronym	METS
Status / version	Version 1.5, April 2005
Type	Encoding standard
Management	Library of Congress
Short description	An XML document format for encoding metadata necessary for both management of (compound) digital library objects within a repository and exchange of such objects between repositories (or between repositories and their users). Depending on its use, a METS document could be used in the role of Submission Information Package (SIP), Archival Information Package (AIP), or Dissemination Information Package (DIP) within the Open Archival Information System (OAIS) Reference Model.
Number of elements	Six modules define descriptive, administrative, structural, rights and other metadata; 36 elements in total. A METS document in XML contains 7 sections.
Extra information on application	METS is an XML Schema designed for the purpose of creating XML document instances that express the hierarchical structure of digital library objects, the names and locations of the files that comprise those objects, and the associated metadata. METS can, therefore, be used as a tool for modelling real world objects, such as particular document types. METS is a standard “shell” for encoding data essential for retrieving, preserving, and serving up digital resources; it can be seen as a "wrapper", like MPEG-21. The need for METS was identified at Digital Library Federation metadata experts meetings, as varied local approaches to non-descriptive metadata are not scaling well & offering little interoperability between agencies. The value of METS is that it offers a standard mode for object “packaging” for preservation, institutional repositories, other activities.[15]
Applied by the following organizations e.g.	British Library, OCLC DCPS, RLG, Harvard, Stanford, UC Berkeley, National Library of Wales are exploring or using for variety of projects. Library of Congress is planning to use with selected moving images, audio recordings, folk life mixed media collections.
URL(s) documentation	http://www.loc.gov/standards/mets/ http://www.loc.gov/standards/mets/docs/mets.v1-5.html
URL guidelines for application	http://www.loc.gov/standards/mets/mets-schemadocs.html Viewed 2006-10-19
XML encoding available	Yes

[15] Metadata standards / Eric Childress. Presentation for FEDLINK OCLC Users Group Meeting. November 18th 2003.

2.3.2 Controlled vocabularies

The following tables provide an overview of the selected controlled vocabularies used by libraries. The selection consists of: Dewey Decimal Classification, Functional Requirements on Authority Records, Library of Congress Authority Files, Library of Congress Classification, Library of Congress Subject Headings, RAMEAU and Universal Decimal Classification code.

Dewey Decimal Classification

Name	Dewey Decimal Classification
Acronym	DDC
Status / version	DDC 22, 2003; updated quarterly
Type	International standard
Management	The system is developed and maintained in the Library of Congress: the Dewey editorial office. Copyrights are owned by OCLC (mailto:DeweyLicensing@oclc.org).
Short description	A universal classification schema, i.e. describing all subject areas. At the broadest level, the DDC is divided into ten *main classes*, which together cover the entire world of knowledge. Each main class is further divided into ten *divisions*, and each division into ten *sections* (not all the numbers for the divisions and sections have been used). This general knowledge organisation tool has a s*tructural hierarchy:* all topics (aside from the ten main classes) are part of all the broader topics above them.
Available in language	English, French and more than 30 other languages.
XML encoding available	No, web-based
Extra information on application	The notation is expressed in Arabic numerals. Thousands of Library of Congress Subject Headings (LCSH) have been statistically mapped to Dewey numbers from records in WorldCat (the OCLC Online Union Catalogue) and intellectually mapped by DDC editors for WebDewey. http://www.oclc.org/dewey/versions/webdewey/default.htm Last viewed 2006-09-15.
Applied by the following organizations e.g.	The DDC is the most widely used classification system in the world. Libraries in more than 135 countries use the DDC to organize and provide access to their collections, and DDC numbers are featured in the national bibliographies of more than sixty countries. Libraries of every type (especially public libraries and small academic libraries in the U.S.) apply Dewey numbers on a daily basis and share these numbers through a variety of means (including WorldCat, the OCLC Online Union Catalogue). Dewey is also used for other purposes, e.g., as a browsing mechanism for resources on the web. For instance, the subject gateway Renardus has assigned DDC for organizing and accessing electronic resources.
URL(s) documentation	http://www.oclc.org/dewey http://www.oclc.org/dewey/versions/ddc22print/glossary.pdf
URL guidelines for application	http://www.oclc.org/dewey/versions/ddc22print/intro.pdf Viewed 15-9-2006

Functional Requirements on Authority Records

Name	Functional Requirements on Authority Records
Acronym	FRAR
Status / version	Draft, June 2005
Type	Conceptual model
Management	IFLA UBCIM Working Group on Functional Requirements and Numbering of Authority Records (FRANAR)
Short description	A conceptual model to set up authority records for metadata elements like person name, family name and organization name according to a predefined structure. Like the rules for a thesaurus, there are 14 relationship types acknowledged, for instance pseudonym relationship and alternative linguistic form relationship.
XML encoding available	No
Applied by the following organizations e.g.	Many
URL(s) documentation	http://www.ifla.org/VII/d4/wg-franar.htm Viewed 04-10-2006

Library of Congress Authority Files

Name	Library of Congress Authority Files
Acronym	LCAF
Status / version	Updated weekly
Type	International standard
Management	Library of Congress
Short description	A set of controlled vocabularies (authority files) for the following metadata elements: subject (see: LCSH), names (person names, corporate names, meeting names and geographic names), series and uniform title and name/title.
Number of elements	Approximately: 265,000 subject authority records 5.3 million name authority records (ca. 3.8 million personal, 900,000 corporate, 120,000 meeting, and 90,000 geographic names) 350,000 series and uniform title authority records 340,000 name/title authority records (numbers date from January 2003).
Available in language	English
XML encoding available	No
Applied by the following organizations e.g.	Widely
URL(s) documentation	http://authorities.loc.gov/ Viewed 2006-10-19

Library of Congress Classification

Name	Library of Congress Classification
Acronym	LCC
Status / version	Not available
Type	International standard
Management	Library of Congress
Short description	LCC is a classification system designed for the Library of Congress collection, covering all subject areas. It has been adopted by many large academic libraries in the U.S.
Number of elements	21 basic classes
Available in language	English
XML encoding available	Yes, the LCC records are available in MARCXML format.
Applied by the following organizations e.g.	It is used by most research and academic libraries in the U.S. and several other countries. Recommended by VRA.
URL(s) documentation	http://www.loc.gov/catdir/cpso/lcco/lcco.html Viewed 2006-10-19 http://www.loc.gov/catdir/cpso/lcc.html Viewed 2006-10-19

Library of Congress Subject Headings

Name	Library of Congress Subject Headings
Acronym	LCSH
Status / version	29th edition, 2006 (the online version is updated weekly)
Type	International standard
Management	Library of Congress
Short description	A thesaurus on all subject areas. A structured vocabulary designed to represent the subject and form of the books, serials, and other materials in the Library of Congress collections, with the purpose of providing subject access points to the bibliographic records contained in the Library of Congress catalogues. More broadly, LCSH is used as a tool for subject indexing of library catalogues and other materials (including visual materials). Available in print (annual) and microfiche (updated quarterly). Also available on line from various vendors and bibliographic utilities, and as part of the Library of Congress CD-ROM product *Classification Plus*.
Number of elements	> 280,000
Available in language	English, Greek, Hungarian
XML encoding available	No
Extra information on application	There is a version of this thesaurus where Hungarian is the first language (> 10,000 items) and English the second language. The terms are stored in MARC21 format. webpac.lib.unideb.hu/corvina/nagy/term_search
Applied by the following organizations e.g.	LCSH are widely used in library catalogues in North America and around the world. The National Library of Canada worked with LCSH representatives to create a complementary set of Canadian Subject Headings (CSH) to access and express the topic content of documents on Canada and Canadian topics. Recommended by VRA.

URL(s) documentation	http://www.loc.gov/cds/lcsh.html Viewed 2006-10-19

RAMEAU

Name	RAMEAU (Répertoire d'autorité-matière encyclopédique et alphabétique unifié)
Acronym	RAMEAU
Status / version	2005, updated continuously
Type	National standard, France
Management	Bibliothèque Nationale de France - Service de coopération bibliographique - Centre national RAMEAU.
Short description	A thesaurus on all subjects, built on the basis of the Canadian controlled vocabulary 'RVM Laval' (in French, with English synonyms), which took the LCSH as a starting point.
Number of elements	More than 256,000; including 46,000 geographic keywords and 88,000 common names.
Available in language	French
XML encoding available	No
Extra information on application	http://rameau.bnf.fr
Applied by the following organizations e.g.	Bibliothèque Nationale de France for the national catalogue BN-Opale Plus, many university, public and municipal libraries in France, Bibliothèque de l'Université Laval (Canada), Communauté française de Belgique.
URL(s) documentation	http://catalogue.bnf.fr/servlet/AccueilConnecte Viewed 2006-10-19

Universal Decimal Classification

Name	Universal Decimal Classification code
Acronym	UDC
Status / version	2006, updated continuously, published three times per year
Type	International standard
Management	The UDC Consortium
Short description	Multilingual classification scheme concerning all subject areas.
Number of elements	> 66,000
Available in language	Czech, Dutch, English, Russian, Spanish.
XML encoding available	No; Data from the Master Reference File (a CDS/ISIS database) can be exported in two ways: in ISO 2709 format or as plain text (ASCII).
Applied by the following organizations e.g.	Widely used within academic and special libraries.
URL(s) documentation	http://www.udcc.org/ Viewed 2006-09-22.
URL guidelines for application	http://www.udcc.org/guide.htm Viewed 2006-10-05.

2.4 Museums

A museum is typically a "permanent institution in the service of society and of its development, open to the public, which acquires, conserves, researches, communicates and exhibits, for purposes of study, education, enjoyment, the tangible and intangible evidence of people and their environment." This definition is taken from the International Council of Museums (ICOM) Statutes.

In addition to the standards described in this section, a number of other controlled vocabularies that are nationally applied, monolingual or handle a specific sub-domain were studied as well. They include:

- the schema of the Consortium for the Computer Interchange of Museum Information (CIMI)
- the UK museum documentation standard, SPECTRUM (Museum Libraries Archives council - MLA)
- the controlled vocabularies published by the British Museum Document Association (MDA)
- the controlled vocabularies published by the English Heritage
- the British Museum thesaurus
- RKDartists, a Dutch controlled list of names of artists
- the thesaurus for graphic materials.

2.4.1 Metadata schemas

The following tables provide an overview of the selected metadata schemas used by museums. The selection consists of: Categories for the Description of Works of Art, Object ID and Visual Resources Association Core Categories.

Categories for the Description of Works of Art

Name	Categories for the Description of Works of Art
Acronym	CDWA
Status / version	Revised version 2005-11-16 CDWA Lite version 1.1 2006-07-17
Type	Recommendation
Management	Getty Art Institute
Short description	CDWA describes the content of art databases by articulating a conceptual framework for describing and accessing information about works of art, architecture, other cultural material, groups and collections of works, and related images. CDWA provides broad, encompassing guidelines for the information elements needed to describe an art object from a scholarly or research point of view. A small subset of categories are considered core in that they represent the minimum information necessary to identify and describe a work. CDWA Lite is an XML schema to describe core records for works of art and material culture based on CDWA and the VRA guidelines for Cataloguing Cultural Objects (CCO). Lite records are intended for contribution to union catalogues and other repositories using the Open Archives Initiative (OAI) harvesting protocol.
Number of elements	512 categories and subcategories E.g. FOR PLACE/LOCATION AUTHORITY : Place Name, Source, Place Type, Broader Context.
Vocabularies proposed	The purpose of the CDWA is described as follows: "The *Categories* provide a framework to which existing art information systems can be mapped and upon which new systems can be

	developed. In addition, the discussions in the CDWA identify vocabulary resources and descriptive practices that will make information residing in diverse systems both more compatible and more accessible."
Extra information on application	Several recommendations on the syntax. E.g. Subject Matter – Display "Use sentence case. You may use complete sentences and/or phrases. Begin the first word of the note with an uppercase letter, and end the note with a period. Follow rules for standard English grammar (if the record is in another language, use grammar rules appropriate to that language). If you rely upon information from a published source, cite the source in SUBJECT MATTER - CITATIONS." The CDWA is mapped to several other standards and metadata element sets: CCO, CIMI, Dublin Core, EAD, FDA Guide, MARC, Object ID and VRA 3.0.[16] The following cultural heritage metadata standards also map to CDWA: the AMICO (Art Museum Image Consortium) data dictionary, SPECTRUM, a standard developed for museums in the UK; the CIDOC Guidelines for Museum Object Information; and the International Council of Museums AFRICOM data standard. See also under: Vocabularies proposed.
Applied by the following organizations e.g.	Recommended by VRA. Art Museum Image Consortium (AMICO, operated from 1997-2005) http://www.amico.org/ : the AMICO Catalogue Record is based on the CDWA. AMICO was a project of the College Art Association and the Getty Art History Information Program.
URL(s) documentation	http://www.getty.edu/research/conducting_research/standards/cdwa/ Viewed 2006-09-20 http://www.getty.edu/research/conducting_research/standards/cdwa/cdwalite.html Viewed 2006-09-20
XML encoding available	Only CDWA Lite: Yes (the core categories)[17] Viewed 2006-10-19

Object ID

Name	Object ID
Acronym	Object ID
Status / version	1994
Type	International standard
Management	Council for the Prevention of Art Theft (info@object-id.com)
Short description	Object ID is an international standard for describing cultural objects developed from a subset of the CDWA and in collaboration with the museum community, police and customs agencies, the art trade, insurance industry, and valuers of art and antiques. This metadata schema codifies the minimum set of data elements needed to protect or recover an object from theft and illicit traffic, it includes the information needed to describe objects for purposes of identification. In 1999 a UNESCO committee endorsed Object ID “as the international standard for recording minimal data on movable cultural property”.
Number of elements	10 metadata elements: Type of Object , Materials & Techniques , Measurements,

[16] http://www.getty.edu/research/conducting_research/standards/intrometadata/crosswalks.html Viewed 2006-09-20

[17] http://www.getty.edu/CDWA/CDWALite/CDWALite-xsd-public-v1-1.xsd

	Inscriptions & Markings, Distinguishing Features, Title, Subject, Date or Period, Maker and Short Description.
Extra information on application	The metadata schema is available in 11 languages: Arabic, Chinese, Czech, English, French, German, Hungarian, Italian, Korean, Russian and Spanish.[18]
Applied by the following organizations e.g.	The use of this standard is promoted by International Council of Museums (ICOM), the Inspectorate of Museums in The Netherlands and the Museum Documentation Association in the UK. The Object ID metadata schema (called: checklist) is compatible with the majority of art theft databases, so a number of insurance companies in Europe and North America are now promoting the standard.
URL(s) documentation	http://www.object-id.com/ Viewed 2006-10-19
XML encoding available	No

Visual Resources Association Core Categories

Name	Visual Resources Association Core Categories
Acronym	VRA Core
Status / version	4.0 beta, September 2006
Type	Recommendation
Management	Visual Resources Association Data Standards Committee
Short description	A metadata schema for describing works of visual culture as well as the images that document them. The Core Categories were designed with the awareness that there are often multiple representations of a work of art, such as the original painting and a slide of the painting used in teaching. The elements that comprise the VRA Core are designed to facilitate the sharing of information among visual resources collections about works and images (= visual representations of a work). Based on Categories for the Description of Works of Art. VRA Core 4.0 also provides an initial blueprint of how those elements can be hierarchically structured.
Number of elements	18 elements: work, collection or image; agent; date; description; inscription; location; material; measurements; relation; rights; source; stateEdition; stylePeriod; subject; technique; textref; title and worktype.
Vocabularies proposed	The VRA Data Standards Committee promotes the use of the following controlled vocabularies: • Art and Architecture Thesaurus (Getty AAT): required for the Type, Material, and Style/Period elements; recommended for the Culture and Subject elements • Controlled vocabularies (ALA) • British Museum Object Names Thesaurus • ICONCLASS • Multilingual Glossary for Art Librarians (IFLA) • Provenance Index (Getty) • Library of Congress Catalogue (LC) • Thesaurus of Geographic Names (Getty TGN)

[18] http://icom.museum/object-id/checklist.html Viewed 2006-09-20

	• Thesaurus for Graphic Materials I (Library of Congress TGM1) • Thesaurus for Graphic Materials II (Library of Congress TGM2) • Union List of Artists Names (Getty ULAN)
Extra information on application	CCO: Cataloguing Cultural Objects is a guide to describing cultural works and their images (draft Feb. 2005). It provides a set of rules surrounding various elements from CDWA (which contains elements and rules) and VRA Core (which contains elements); it is more directly analogous to AACR and DACS.[19] Published by ALA in June 2006. CCO provides guidelines for selecting, ordering, and formatting data used to populate catalogue records based on core categories in CDWA and VRA Core. Each category of VRA Core 3.0 is mapped to CDWA and to Dublin Core 1.1[20] A crosswalk of VRA Core 3.0 to MARC21[21] "The VRA template provides a specialization of the Dublin Core set of metadata elements, tailored to the needs of art images."[22] VRA provides guidance, e.g. core category **Measurements**:[23] 1. Qualifiers • Measurements.Dimensions • Measurements.Format • Measurements.Resolution 2. Description: The size, shape, scale, dimensions, format, or storage configuration of the Work or Image. Dimensions may include such measurements as volume, weight, area or running time. The unit used in the measurement must be specified. 3. Data Values: formulated according to standards for data content (e.g., AACR, etc.) 4. VRA Core 2.0: W3 Measurements; V2 Visual Document Format; V3 Visual Document Measurements 5. CDWA: Measurements-Dimensions; Measurements-Shape; Measurements-Format; Related Visual Documentation-Image Measurements 6. Dublin Core: FORMAT
Applied by the following organizations e.g.	Many
URL(s) documentation	http://www.vraweb.org/datastandards/VRA_Core4_Welcome.html Viewed 2006-09-20
XML encoding available	Yes

[19] http://www.vraweb.org/ccoweb// Viewed 2006-09-20.

[20] http://www.vraweb.org/vracore3.htm#record%20type Viewed 2006-09-20.

[21] http://php.indiana.edu/~fryp/marcmap.html Viewed 2006-09-20.

[22] An integrated multimedia approach to cultural heritage e-documents / Arnold W.M. Smeulders (University of Amsterdam, UvA), Lynda Hardman (CWI), Guus Schreiber (UvA) and Jan-Mark Geuzebroek (UvA), 2003. www.MultimediaN.nl

[23] http://www.vraweb.org/vracore3.htm

2.4.2 Controlled vocabularies

The following tables provide an overview of the selected controlled vocabularies used by museums. The selection consists of: Art and Architecture Thesaurus, ICONCLASS, Thesaurus of Geographic Names, Union List of Artists Names and UNESCO thesaurus**.**

Art and Architecture Thesaurus

Name	Art and Architecture Thesaurus
Acronym	AAT
Status / version	Updated monthly
Type	International standard
Management	The Getty Institute
Short description	A thesaurus on the subject areas: fine art, architecture, decorative arts, archival materials, and material culture.
Number of elements	> 131,000 terms, approx. 34,000 concepts
Available in language	Dutch, English
XML encoding available	Yes
Extra information on application	Dutch version is available.[24] The data are also available as relational tables and in MARC format.
Applied by the following organizations e.g.	AAT is a worldwide applied indexing tool. Used by dozens of cultural institutions and individuals in Flanders (Belgium) and the Netherlands. The MultiMedian project applies for instance AAT. http://e-culture.multimedian.nl/demo/facet The portal provides access to a relatively large set of key culture-heritage collections in The Netherlands. The library as well as the archive of the Dutch national heritage organisation will be using the Dutch translation of the AAT thesaurus for subject indexing. This standard is recommended by VRA.
URL(s) documentation	http://www.getty.edu/research/conducting_research/vocabularies/aat/
URL guidelines for application	http://www.getty.edu/research/conducting_research/vocabularies/aat/about.html http://www.getty.edu/research/conducting_research/vocabularies/aat/faq.html Viewed 2006-10-19

ICONCLASS

Name	ICONCLASS
Acronym	ICONCLASS
Status / version	Version 2005
Type	International standard
Management	Royal Netherlands Academy of Arts and Sciences (KNAW).
Short description	**ICONCLASS** is a specialized library classification designed for iconographic research and documentation of images. It was originally conceived by Henri van de Waal, and was

[24] http://www.rkd.nl/aat/index.html

	further developed by a group of scholars after his death. The domain is the iconography of western art from the medieval period to the contemporary art. ICONCLASS offers ready-made definitions for objects, persons, events, situations, and abstract ideas that can be represented in the visual arts. The system is organized in ten broad divisions within which subjects are ordered hierarchically. The relations between the terms are like in a thesaurus: Narrower term / Broader term ; Narrower term abstract / Broader term abstract ; Narrower term partitiv / Broader term partitiv ; Narrower term casual / Broader term casual ; Related term (or 'See also') ; Use/Used for (or 'See') ; Use OR ; Use AND ; Top term ; Other relations ; Scope Note.
Number of elements	The classification system contains > 28,000 definitions.
Available in language	English, German, Finnish, French and Italian
XML encoding available	No
Applied by the following organizations e.g.	• Medieval Illuminated Manuscripts of the National Library of the Netherlands (Koninklijke Bibliotheek) and the Museum Meermanno-Westreenianum : a database application that uses the ICONCLASS server for multilingual searching and retrieval of digital image files over the Internet. • Bildarchiv zur Kunst und Architektur in Deutschland • Digitales Informations-System für Kunst- und Sozialgeschichte: numerous German museums, archives, offices for the preservation of historical monuments, research institutes and university institutes are co-operating in compiling a database of art in Germany (DISKUS project). • Marburger Index, a reference work and guide to art in Germany • Bodleian Library Broadside Ballads See further: http://www.iconclass.nl/texts/project01.htm
URL(s) documentation	http://www.iconclass.nl/ http://www.iconclass.nl/downloads/iconclass.pdf
URL guidelines for application	http://www.iconclass.nl/texts/pub01.htm Viewed 2006-10-19

Thesaurus of Geographic Names

Name	Thesaurus of Geographic Names
Acronym	TGN
Status / version	Updated monthly.
Type	International standard
Management	The Getty Institute
Short description	The TGN is a structured vocabulary that contains names and other information about geographic places. TGN is focused on, but not limited to, places important to the study of art and architecture. Contains other interesting information, including vernacular and historical names, coordinates and place types.
Number of elements	> 1,000,000
Available in language	English
XML encoding available	Yes
Extra information on	The data are also available as relational tables.

application	
Applied by the following organizations e.g.	The MultiMedian project applies TGN. http://e-culture.multimedian.nl/demo/facet The portal provides access to a relatively large set of key culture-heritage collections in The Netherlands. This standard is recommended by VRA and by DCMI.
URL(s) documentation	http://www.getty.edu/research/conducting_research/vocabularies/tgn Viewed 2006-10-19

Union List of Artist Names

Name	Union List of Artists Names
Acronym	ULAN
Status / version	Updated monthly.
Type	International standard
Management	The Getty Institute
Short description	ULAN is a structured vocabulary containing around 120,000 records, including 293,000 names and biographical and bibliographic information about artists and architects, including a wealth of variant names, pseudonyms, and language variants.
Number of elements	Around 120,000 records on artists and architects
Available in language	English
XML encoding available	Yes
Extra information on application	The data are also available as relational tables and in MARC format.
Applied by the following organizations e.g.	The MultiMedian project. http://e-culture.multimedian.nl/demo/facet The portal provides access to a relatively large set of key culture-heritage collections in The Netherlands. Recommended by VRA.
URL(s) documentation	http://www.getty.edu/research/conducting_research/vocabularies/ulan/ Viewed 2006-09-22

UNESCO thesaurus

Name	UNESCO thesaurus
Acronym	UNESCO thesaurus
Status / version	CD-ROM 2005: 13th edition; web edition is updated continuously.
Type	Recommendation
Management	The UNESCO library (contact: Meron Ewketu)
Short description	This thesaurus is built according to the standards ISO 2788 & ISO 5964 and covers the following domains: education; culture; natural sciences; social and human sciences; communication and information; politics, law and economics; countries and country groupings. It also includes the names of countries and groupings of countries: political, economic, geographic, ethnic and religious, and linguistic groupings. There is a version on paper, on CD-ROM and on internet.
Number of elements	About 21,000 terms (7,000 in English, 8,600 in French, 6,800 in Spanish)
Available in	English, French, Russian, Spanish

language	
XML encoding available	No
Extra information on application	Softwares: Winisis, BASIS; wwwisis (web version)
Applied by the following organizations e.g.	UK National Digital Archive of Datasets (NDAD)
URL(s) documentation	http://databases.unesco.org/thesaurus/ Viewed 2006-10-19 http://www2.ulcc.ac.uk/unesco/ Viewed 2006-10-19

2.5 Educational Sector

The educational domain includes primary, secondary, and higher education. The latter can be broken down further into education provided by universities, vocational universities (community colleges, liberal arts colleges, and technical colleges, etc.) and other collegial institutions that award academic degrees, such as career colleges.

In addition to the ones described in this section, we also looked at a number of other metadata schemas and controlled vocabularies that are applied in the educational sector, namely:

- Dublin Core Education
- MEG, a standard from the UK's Metadata for Education Group
- the ERIC thesaurus, 13th edition, contains an alphabetical listing of terms used for indexing and searching in the ERIC database (online bibliographic database).
- Sharable Content Object Reference Model (SCORM), a collection of standards and specifications for web-based e-learning and
- the controlled vocabularies of the IMS Global Learning Consortium**.**

2.5.1 Metadata schemas

The following tables provide an overview of the selected metadata schemas used by the educational sector. The selection consists of: Gateway to Educational Material and IEEE Standard for Learning Object Metadata.

Gateway to Educational Material

Name	Gateway to Educational Material
Acronym	GEM
Status / version	2.0 draft, 2001
Type	International standard
Management	GEM consortium
Short description	GEM describes, manages, organizes education resources. It is an extension of Dublin Core, optimized to provide on-target, efficient searching by grade and subject keyword.
Number of elements	10 elements, namely: dc-ed.audience ; dc-ed.mediator ; dc-ed.level (Grade) ; gemq.age ; gemq.beneficiary ; gem.duration ; gem.resources ; gem.pedagogy ; gem.competency (Standards) ; gem.cataloguing.
Applied by the following organizations e.g.	• Federal Resources for Education Excellence (www.ed.gov) • CHIN Teachers' Centre (www.virtualmuseum.ca) • The Gateway to Educational Material (http://thegateway.org/)
URL(s) documentation	http://64.119.44.148/about/documentation/metadataElements RDF/XML schema for the qualified GEM element set http://purl.org/gem/qualifiers/
XML encoding available	Yes

Learning Object Metadata

Name	IEEE Standard for Learning Object Metadata
Acronym	LOM
Status / version	1484.12.1-2002
Type	IEEE standard
Management	IEEE Learning Technology Standards Committee (LTSC)
Short description	This Standard is a multi-part standard that specifies Learning Object Metadata. This part specifies a conceptual data schema that defines the structure of a metadata instance for a learning object. For this Standard, a learning object is defined as any entity--digital or non-digital-- that may be used for learning, education or training. Learning Objects are defined here as any entity, digital or non-digital, which can be used, re-used or referenced during technology supported learning. The standard focuses on the minimal set of attributes needed to allow these Learning Objects to be managed, located, and evaluated. Including: the ability for locally extending the basic fields and entity types (field status obligatory or optional). Relevant attributes of Learning Objects to be described include: type of object, author, owner, terms of distribution, and format, as well as (where applicable): pedagogical attributes, such as teaching or interaction style, grade level and mastery level.
Number of elements	Approximately 75 elements. Some of which can be used in more than one way. E.g. 'LifeCycle.Contribute.Entity' can be the metadata element for creator, when 'LifeCycle.Contribute.Role' has a value of "Author", but the same metadata element can also contain the publisher name, if 'LifeCycle.Contribute.Role' has a value of "Publisher".
Extra information on application	Standard for ISO/IEC 11404 binding for Learning Object Metadata data model to provide precise data model semantics, as permitted by the 11404 notation. (1484.12.2) Crosswalk to DC: http://db1-www.sub.uni-goettingen.de/servlets/metaformList1?Table=LOMDC&Head=LOM
Applied by the following organizations e.g.	The LOM data model standard (IEEE 1484.12.1-2002) has been widely adopted and adapted, including being fully incorporated into the most recent and stabilized version of the ADL SCORM (2004) and the IMS Global Learning Consortium's Meta-Data Specification.
URL(s) documentation	http://ltsc.ieee.org/wg12/
URL guidelines for application	IMS Meta-data Best Practice Guide for IEEE 1484.12.1-2002 Standard for Learning Object Metadata. Version 1.3 Public Draft. http://www.imsglobal.org/metadata/mdv1p3pd/imsmd_bestv1p3pd.html CanCore Guidelines for the Implementation of Learning Object Metadata (LOM) 2.0. http://www.cancore.ca/documents.html
XML encoding available	Yes, Standard for XML binding for Learning Object Metadata data model (1484.12.3)

2.6 Audiovisual Sector

A description of the domain could be: "Production, distribution and archiving of digital audiovisual material". The application fields are: radio and television broadcasting; audio and video (post)production; audiovisual archives; multimedia library systems; image banks; news agencies; WEB TV. Due to the convergence of television, computer and communication technologies, the separate areas within the digital media domain are approaching one another in many respects.

2.6.1 Metadata schemas

The following tables provide an overview of the selected metadata schemas used by the audiovisual sector. : FIAF Cataloguing Rules, MusicBrainz, Material Exchange Format, P_META Exchange scheme, Standard Media Exchange Framework Data Model, SMPTE Metadata Dictionary and TV-Anytime.

FIAF Cataloguing Rules

Name	FIAF Cataloguing Rules
Acronym	FIAF
Status / version	1991
Type	Recommendation
Management	International Federation of Film Archives (FIAF)
Short description	The FIAF Cataloguing Rules specify requirements for the description and identification of archival moving image materials, assign an order to the elements of the description, and specify a system of punctuation for that description.
Applied by the following organizations e.g.	Many
URL(s) documentation	http://www.fiafnet.org/uk/publications/catrules.cfm. 3 parts in PDF Viewed 2006-10-19
XML encoding available	No

MusicBrainz

Name	The MusicBrainz Metadata Initiative
Acronym	MusicBrainz
Status / version	2003
Type	Recommendation
Management	MetaBrainz Foundation
Short description	MusicBrainz is a user maintained community music metadatabase. Music metadata is information such as the name of an artist, the name of an album and list of tracks that appear on an album. MusicBrainz collects this information about music and makes it available to the public. The MusicBrainz metadata initiative is a model for describing digital audio and video tracks.
Number of elements	The database structure consists of seven parts: Artists, Albums, Tracks, TRMs, The search tables (track words and album words), Moderators and Moderations.

Applied by the following organizations e.g.	Widely by broadcasting companies and radio stations.
URL(s) documentation	http://musicbrainz.org/products/server/docs/index.html Viewed 2006-10-19
XML encoding available	No

Material Exchange Format

Name	Material Exchange Format
Acronym	MXF
Status / version	SMPTE 377M
Type	Recommendation
Management	SMPTE
Short description	MXF is a File Format optimized for the interchange of material for the content creation industries. MXF is a wrapper format intended to encapsulate and accurately describe one or more "clips" of Essence. These Essence "clips" may be Pictures, Sound, Data or some combination of all of these. The core requirement for the design and development of MXF was to be able to bundle the essence and an "EDL" in an unambiguous way that was both essence agnostic and metadata aware. In order for an application to do anything, the file must contain data about the essence i.e. Metadata. This particular sort of metadata is called “structural metadata” and allows many applications and devices to process content without knowing a-priori what the content is. The accurate description of the underlying content is one of the key strengths of the MXF format.
Extra information on application	MXF (Media Exchange Format) is a standard which defines the data structure for audio and visual material (essence) at a point of exchange (that is over networks, not internally within a system) it defines a header and footer as well as the manner in which metadata is packed, however, it does not set a standard for essence metadata's format (that is, information about the material from it's original acquisition through all the steps up to the present form), although it does define how the metadata must be written to "plug in" to MXF. MXF is more of an “end of the chain” wrapper, designed for completed content which will not be further reworked, or, at most, a cuts only assembly of material. This is especially important in a broadcast environment or for broadband content delivery.[25]
Applied by the following organizations e.g.	Not available
URL(s) documentation	http://www.smpte.org/smpte_store/standards/ Viewed 2006-10-19
XML encoding available	Yes, for the metadata part.

[25] http://www.creativecow.net/articles/fronc_marisu/aaf_mxf_nab/

P_META

Name	P_META Exchange scheme
Acronym	P_META
Status / version	version 1.2
Type	standard - EBU Tech 3295
Management	EBU
Short description	A standardised metadata exchange scheme that offers a way of sharing the meaning of electronic information necessary or useful for the business-to-business exchange of content. The P_META Scheme is intended for use in a business-to-business scenario where the participating organisations, mainly in the professional broadcast industry, may retain their internal data structures, workflows, and concepts. The P_META Scheme is basically a set of definitions that provide a semantic framework for the information that is typically exchanged along with audio-visual material (i.e. TV programs). It includes the identification of concepts (simple or complex) that are referenced by P_META names and P_META Identifiers as well as the description of the individual programs, transmission or publication metadata, metadata concerning editorial objects and media objects, technical metadata (audio and video specification, compression schemes, and so on), and rights metadata (contract clauses, rights list, and copyright holders).
Number of elements	The P_META standard presents a five-layered hierarchical model: the brand, the program group, the program, the program item, and the media object. Furthermore five different transactions are acknowledged, e.g. Producer to Distributor, Archive to Distributor. More than 220 elements are divided over these five entities/layers and transactions.
Vocabularies proposed	Classification of TV programs (for instance genre) according to the Escort 2.5 system. http://www.ebu.ch/en/technical/metadata/specifications/ Viewed 2006-10-02.
Extra information on application	The syntax is defined by an XML schema.
Applied by the following organizations e.g.	RAI (Italian Broadcasting Agency)
URL(s) documentation	http://www.ebu.ch/en/technical/metadata/specifications/notes_on_tech3295.php http://www.ebu.ch/CMSimages/en/tec_doc_t3295_v0102_tcm6-40957.pdf Viewed 2006-10-19
XML encoding available	Not applicable. It is technology-independent, and can be used in applications to create XML documents, embed metadata in file formats such as MXF or BWF, or simple Word templates.

SMEF-DM

Name	Standard Media Exchange Framework Data Model
Acronym	SMEF-DM
Status / version	1.5
Type	Standard
Management	BBC
Short description	SMEF-DM is the part of SMEF, an internal metadata model developed for the British Broadcasting Corporation (BBC) that is released to the public. It is a semantic, logical, hierarchical data model: • defining the meanings of items of data (attributes), of logical clusters of these items

	(entities), and of the relationships between the clusters • recording all information that becomes available during the whole production cycle, from a program concept over media and editorial objects to the actual publication. The definitions are organization independent and should be usable for any other broadcasters. The rights metadata are extensive.
Number of elements	The numerous elements are organised per entity. The main entities are: Brand, Programme Group, Programme (Editorial Object Programme) and Programme Item (Editorial Object Item). Other entities exist e.g. for Distribution-Channels and related issues, Mediatypes, Contracts and other right-issues, and many, many more. The material/essence is linked to the entity "MEDIA_OBJECT_INSTANCE" (MOI).
Extra information on application	"SMEF was a very committed and complex attempt towards the "perfect schema", going far beyond core TV-business but taken into account future goals. It started in an era when the distress-calls for metadata-standards "fulfilling all needs, situations and business" swept over the Media communities worldwide. The complex structure, relations and composition of SMEF seem to be the main obstacle to a worldwide success and usage of SMEF. Introducing Open-SMEF and its successors was a good attempt to boost the achievement of SMEF. SMEF-DM is technology agnostic, which ensures wide applicability. It makes use of standard attributes and definitions where possible, e.g. SMPTE, ISO and EBU, and is not in competition with these organisations but seeks co-operation. BBC used SMEF for input in other international Taskforces like SMPTE, MPEG7 and MPEG21, but also the EBU-Projects P/FTA, P/FRA,P /CHAIN, etc. Due to the fact that SMEF was not only meant for "BBC-in-house"-use, the exchange with other parties/systems is part of the system. The large number of entities and relations may be tricky to be handled when data is exchanged with other systems and the extensive use of manual intervention is to be taken into account." [26]
Applied by the following organizations e.g.	BBC
URL(s) documentation	http://www.bbc.co.uk/guidelines/smef/ Viewed 2006-10-19
XML encoding available	No

SMPTE MD

Name	SMPTE Metadata Dictionary
Acronym	SMPTE MD
Status / version	SMPTE 335M
Type	Recommendation
Management	SMPTE
Short description	A metadatadictionary structured in 7 distinct classes: Identification, Administration, Interpretation, Parametric, Process, Relational and Spatio-Temporal. With 457 nodes and 1363 leafs, the dictionary provides elements for

[26] PRESTOSPACE WP15_ORF_D15.1_DOCUMENTATION_MODELS_V1.6

	nearly every occasion and need in the A/V-world, acting perfectly as a rich source ("repository") for every DM-developing taskforce.
Number of elements	> 1800 elements
Applied by the following organizations e.g.	Many broadcasting companies either use this schema, or have applied it in a proprietary version.
URL(s) documentation	http://www.smpte-ra.org/mdd/RP210v8-final-040810MC.xls http://www.smpte.org/smpte_store/standards/index.cfm?scope=0&stdtype=smpte&CurrentPage=15 Viewed 2006-10-19
XML encoding available	No

TV-Anytime

Name	TV-Anytime
Status / version	2nd phase of specifications was accepted 2005
Type	Specification
Management	The TV-Anytime Forum[27]
Short description	The global TV-Anytime Forum is an association of organisations which seeks to develop specifications to enable audio-visual and other services based on mass-market high volume digital storage in consumer platforms. The TV-Anytime metadata specification (SP003v13) contains two main parts: • Part A the schemas. The TV-Anytime Forum has adopted the XML-based MPEG-7 Description Definition Language (DDL) [ISO/IEC 15938-2] as its representation format for metadata. Part A specifies the format of metadata to be exchanged, e.g. between content / information / metadata providers and the consumers including service, content and user description schemas and classification schemes • Part B system aspects. This part contains a recommended binary format (MPEG-7 BiM [ISO/IEC 15938-1]), a fragmentation model, a mode of encapsulation of these fragments and an indexing method. All the XML files necessary to implement the specification are provided in SP003v13. TV-Anytime is transport agnostic and can be adapted to different environments. During the fall of 2005, the second phase of the TV-Anytime specifications was accepted by ETSI as a proper specification for TV content metadata. And the underlying reference mechanism, the CRID, will be turned into an Internet specification by IETF.
Applied by the following organizations e.g.	While there seems to be a strong interest in adopting the TV-Anytime specifications, among cable and satellite broadcasters, this specification has yet to be deployed in a practice.
URL(s) documentation	http://www.ebu.ch/CMSimages/en/online_33_e_TV_anytime_tcm6-4050.pdf?display=EN http://www.ebu.ch/en/technical/trev/trev_295-evain.pdf?display=EN Both viewed: 2006-20-10
XML encoding available	Yes

[27] http://www.tv-anytime.org/

2.6.2 Controlled vocabularies

The following tables provide an overview of the selected controlled vocabularies used by the audiovisual sector. The selection consists of: FIAF subject headings and EBU System of Classification of Radio or Television Programmes.

FIAF subject headings

Name	International Index to Film Periodicals - Subject Headings
Acronym	FIAF subject headings
Status / version	January 2006
Type	Recommendation
Management	International Federation of Film Archives
Short description	Thesaurus used for the International Index to Film Periodicals. Consists of 3 parts: 1 General Subject Headings; 2 Corporate Bodies; 3 Personal names. Preferred terms and non-preferred terms in alphabetical order.
Number of elements	> 2,200 subject headings > 1,500 corporate bodies > 130 names of persons in films
Available in language	English
XML encoding available	No
Applied by the following organizations e.g.	Filmmuseum (The Netherlands) and many others.
URL(s) documentation	http://www.fiafnet.org/uk/publications/theasaurus.cfm Viewed 2006-10-05.

EBU System of Classification of Radio or Televison Programmes

Name	EBU System of Classification of Radio or Television Programmes
Acronym	ESCORT
Status / version	Version 2.5, 2006
Type	Recommendation
Management	EBU
Short description	The focus of ESCORT 2006 is the definition of television and radio programme and service concepts and genres. ESCORT 2006 information can be used within broadcasting organisations for a number of purposes, including finance & accounting and marketing. It can also be used to share and exchange information between broadcasters, for example audience research and statistics. ESCORT 2006 offers the means of exchanging comparable data between broadcasters or third parties. ESCORT 2006 related applications require concept and genre descriptions significantly simpler than otherwise used in broadcast applications such as Electronic Programme Guides (EPGs) where "genre" is used to attract or target viewers. For that reason, ESCORT

	applications may use one or more of the dimensions proposed in this document. ESCORT 2006 is a multi-dimensional classification scheme. The use of several dimensions allows the separation of genres into different categories while these different dimensions can be combined to develop richer content and genre descriptions. Preferably, each dimension should be applicable to every programme or service. ESCORT 2006 can be used to classify both radio and television programmes.
Elements	The ESCORT classification is organised in seven dimensions: • Intention • Format • Content • Participation • Intended Audience / Target Group • Origination • Content Alert
Available in language	English
XML encoding available	No
Applied by the following organizations e.g.	EBU
URL(s) documentation	http://www.ebu.ch/metadata/pmeta/WIP/ESCORT/ESCORT2006.htm http://www.newssummit.org/2005/presentations/Radio%20and%20TV%20Metadata.pdf Both viewed 2006-10-05.

See also the case studies of Alinari and the Netherlands Institute for Sound and Vision in chapter 3.

2.7 Geospatial Sector

The geospatial sector includes natural heritage, archaeology and open cultural heritage (i.e. managing historic buildings and monuments). Geospatial information in the context of this deliverable, is information that locates, objects from the domains archaeology and natural or open cultural heritage on/to the earth. The form in which the information is presented can be: metadata on the monuments themselves or photographs, audiovisual documents, textual documents, technical drawings, maps etcetera of the monuments.

Geospatial information can therefore be managed by all sub-domains mentioned. The distinction between geospatial information and other cultural heritage information lies in the fact that there is a spatial aspect to the geospatial data. In other words, the tangible cultural heritage objects can be related to a specific space on or (up to 100 metres) on planet earth. Again, in other words: geographic objects are buildings etcetera that can be located on the earth with coordinates such as for example latitude and longitude.

Various forms into which geospatial information can be presented[28]:

Form	**Examples**
atlas	boundary atlas: geological atlas; historical atlas; plat book; road atlas; statistical atlas (collections of maps, geospatial illustrations, and other information)
diagram	block diagram; fence diagram; reliability diagram; triangulation diagram (illustrations of specific relationships)
globe	terrestrial globe; celestial globe (physical models of celestial bodies)
map	aeronautical chart; base map; cadastral map; chart; index map; orthophotomap; plan; plat; relief map; thematic map
model	relief model (other physical models of geospatial data)
profile	(an illustration showing a vertical section of the ground)remote-sensing image, aerial photograph; photomosaic; infrared scanning image; multispectral scanning image; Sidelooking Airborne Radar (SLAR) image; SPOT image
section	geologic section
view	panorama; perspective view

An additional application of geospatial metadata is to document geographic digital resources such as Geographic Information System (GIS) files, geospatial databases, and earth imagery. A geospatial metadata record includes core library catalogue elements such as Title, Abstract, and Publication Data; geographic elements such as Geographic Extent and Projection Information; and database elements such as Attribute Label Definitions and Attribute Domain Values.[29]

More than 20 standards and/or guides on different aspects of geographic information are published by Technical Committee 211.[30] This includes a formal schema for geospatial metadata that is intended to apply to all types of information, ISO 19115:2003 (see section 2.7.1).

[28] FGDC Metadata Workbook - Version 2.0 5/1/00

[29] http://www.fgdc.gov/metadata

[30] ISO/TC211, http://www.isotc211.org/

2.7.1 Metadata schemas

The following tables provide an overview of the selected metadata schemas used by the geospatial sector. The selection consists of: Standard for Digital Geospatial Metadata, ISO 19115:2003, Monument Inventory Data Standard and Open Geospatial Consortium Specifications.

Standard for Digital Geospatial Metadata

Name	Standard for Digital Geospatial Metadata
Acronym	CSDGM
Status / version	Version 2, 1998
Type	US Federal Metadata standard
Management	The Federal Geographic Data Committee (FGDC)
Short description	Provides common standard for publishing metadata about geospatial resources. The standard establishes the names of data elements and compound elements (groups of data elements) to be used for these purposes, the definitions of these compound elements and data elements, and information about the values that are to be provided for the data elements. The major uses of metadata are: • to maintain an organization's internal investment in geospatial data • to provide information about an organization's data holdings to data catalogues, clearinghouses, and brokerages, and • to provide information needed to process and interpret data to be received through a transfer from an external source. The information included in the standard was selected based on four roles that metadata play: • availability - data needed to determine the sets of data that exist for a geographic location; • fitness for use - data needed to determine if a set of data meets a specific need; • access - data needed to acquire an identified set of data; • transfer - data needed to process and use a set of data.
Extra information on application	Many systems and applications support the standard. Crosswalk of FGDC to ISO 19115:2003(E) *Geographic information - Metadata* available; ANSI technical amendment for ISO-FDGC harmonization in progress. Crosswalk to MARC21 and vice versa available.
Applied by the following organizations e.g.	Widely used by government and business.
URL(s) documentation	http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/base-metadata/index_html CSDGM Workbook Viewed 2006-10-19
URL guidelines for application	CSDGM Standard - This is the official technical specification of the CSDGM Standard. This document is recommended for those familiar with the nomenclature of standards technical documentation and those developing metadata creation and publication software applications.
XML encoding available	Yes

ISO 19115:2003

Name	ISO 19115:2003
Acronym	ISO 19115:2003
Status / version	Published and approved in 2003
Type	ISO
Management	ISO
Short description	ISO 19115 provides an abstract or logical model for the organization of geospatial metadata. It defines the schema required for describing geographic information and services. It provides information about the identification, the extent, the quality, the spatial and temporal schema, spatial reference, and distribution of digital geographic data. ISO 19115:2003 is applicable to the cataloguing of datasets, clearinghouse activities, and the full description of datasets and to geographic datasets, dataset series, and individual geographic features and feature properties. Though ISO 19115:2003 is applicable to digital data, its principles can be extended too many other forms of geographic data such as maps, charts, and textual documents as well as non-geographic data.
Applied by the following organizations e.g.	RACM, the Dutch national heritage organisation. Recommended by Open GIS.
URL(s) documentation	http://www.iso.ch/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=26020&I%20CS1=35&ICS2=240&ICS3=70 Viewed 2006-10-19
XML encoding available	Yes; A companion specification, ISO 19139, standardises the expression of 19115 metadata using the Extensible Markup Language (XML) and includes the logical model (UML) derived from ISO 19115. [31]

Monument Inventory Data Standard

Name	Monument Inventory Data Standard
Acronym	MIDAS
Status / version	Version 1, 1998 (2nd edition to be published in 2007)
Type	UK standard
Management	Forum on Information Standards in Heritage (FISH)
Short description	MIDAS is a content standard that sets out what sort of information should be recorded, for instance to describe the character or location of a monument. MIDAS sets out an agreed list of the items or 'units' of information that should be included in an inventory or other systematic record of the historic environment. These units of information are grouped together under broad headings or 'information schemas'. They cover areas such as Monument Character, Events, People and Organisation etc. It is a 'content' standard or 'metadata' standard for historic environment information.
Applied by the following organizations e.g.	Many
URL(s)	http://www.jiscmail.ac.uk/cgi-

[31] The SDI Cookbook : developing spatial data infrastructures / GSDI. – version 2.0, 25 January 2004. http://www.gsdi.org/docs2004/Cookbook/cookbookV2.0.pdf

documentation	bin/filearea.cgi?LMGT1=FISH&a=get&f=/web_midasintro.htm
URL guidelines for application	http://www.english-heritage.org.uk/upload/pdf/MIDAS3rdReprint.pdf http://www.english-heritage.org.uk/server/show/nav.8331 Viewed 2006-10-19
XML encoding available	No

Open Geospatial Consortium Specifications

Name	Open Geospatial Consortium Specifications
Acronym	OGC Specifications (formerly: Open GIS)
Status / version	Not applicable
Type	Recommendation
Management	The Open GeoSpatial Consortium Inc.(formerly the Open GIS Consortium, now OGC)
Short description	OGC develops specifications because they acknowledge the requirement for geospatial standards and strategies to be an integral part of business processes and enterprise architectures. There are 19 specifications currently available. Possibly relevant: OpenGIS® Catalogue Service Implementation Specification (CAT) en OpenGIS® Implementation Specification for Geographic information - Simple feature access - Part 1: Common architecture (SFA)
Extra information on application	There are a number of OGC Specifications that can be used immediately to enhance the sharing and integration of CAD/GIS content and services. How these interface standards are used and what CAD/GIS interoperability issues they will solve is dependent on the requirements of the users and the applications. [http://www.directionsmag.com/article.php?article_id=687&trv=1]
Applied by the following organizations e.g.	Many
URL(s) documentation	http://www.opengeospatial.org/specs/?page=specs
URL guidelines for application	http://www.directionsmag.com/article.php?article_id=687&trv=1 Viewed 2006-10-19
XML encoding available	No

2.7.2 Controlled vocabularies

The following tables provide an overview of the selected controlled vocabularies used by the geospatial sector. The selection consists of: ISO 3166 and INSCRIPTION.

ISO 3166

Name	International Standard for Country Codes
Acronym	ISO 3166
Status / version	Updated regularly.
Type	ISO standard
Management	International Organization for Standardization (ISO)
Short description	ISO 3166 is the International Standard for country codes. The purpose of ISO 3166 is to establish codes for the representation of names of countries, territories or areas of geographical interest, and their subdivisions. ISO 3166 contains three parts: • 1:1997 Codes for the representation of names of countries and their subdivisions - Part 1: Country codes which is what most users know as ISO's country codes. First published in 1974, it is has since then become one of the world's most popular and most widely used standard solution for coding country names. It contains a two-letter code which is recommended as the general purpose code, a three-letter code which has better mnemonic properties and a numeric-3 code which can be useful if script independence of the codes is important. • 2:1998 Codes for the representation of names of countries and their subdivisions - Part 2: Country subdivision code which gives codes for the names of the principal subdivisions (e.g provinces or states) of all countries coded in ISO 3166-1. This code is based on the two-letter code element from ISO 3166-1 followed by a separator and a further string of up to three alphanumeric characters. • 3:1999 Codes for the representation of names of countries and their subdivisions - Part 3: Code for formerly used names of countries which contains a four-letter code for those country names which have been deleted from ISO 3166-1 since its first publication in 1974. The code elements for formerly used country names have a length of four alphabetical characters (alpha-4 code elements). ISO 3166-1 is by far the most important of the three standards.
Number of elements	Not available.
Available in language	English, French
XML encoding available	No
Applied by the following organizations e.g.	Widely used Recommended by the Library of Congress to use with MARC descriptions.
URL(s) documentation	http://www.iso.org/iso/en/prods-services/iso3166ma/04background-on-iso-3166/what-is-iso3166.html
URL guidelines for application	http://www.iso.org/iso/en/prods-services/iso3166ma/04background-on-iso-3166/implementations-of-iso3166-1.html Viewed 2006-10-19

INSCRIPTION

Name	INSCRIPTION
Acronym	INSCRIPTION
Status / version	Not available
Type	National standard
Management	Forum on Information Standards in Heritage (FISH)
Short description	It is a collection of 'wordlists' for comprehensive and consistent indexing of different aspects of the built and buried heritage. Wordlist available per MIDAS Unit of Information (*= Candidate unit): Archive / Source Type ; Artefact Type*; Civil Parish; Component*; Condition; Constructional Material; County; Currency; Date Range Qualifier; District; Event Type; Evidence; Internal Cross-reference; Qualifier; Land Use; Management Proposal Outcome; Management Proposal Type; Management; Proposal Work Proposed; Monument Type; National Grid Reference Precision; Non Parish Area; Period; Protection Grade; Protection Status; Scientific Date Method; Topology; Unitary Authority.
Available in language	English
XML encoding available	No
Extra information on application	Recommend by FISH.
Applied by the following organizations e.g.	Many
URL(s) documentation	http://www.fish-forum.info/i_lists.htm Viewed 2006-10-19

For the description of the Getty Thesaurus of Geographic Names (TGN) see section 2.4.2.

3 Case Descriptions

This chapter provides information on the knowledge representations used by the some of the cultural heritage institutions within the consortium and the Advisory Board. It also lists seventeen European projects and initiatives that are closely related to MultiMatch. Furthermore, it includes data from a relevant inventory on multilingualism conducted by the MINERVA Plus project and provides a summary of the use of controlled vocabularies in the cultural heritage domain.

3.1 Alinari - Italy

Founded in Florence in 1852, Fratelli Alinari is the oldest firm in the world working in the field of photography, the image and communication. The birth of photography and the story of the Firm go hand in hand in their development and growth: the Alinari Archives contain 3.5 million photographs, collected in the Alinari Archives. Alinari also manages books, collotypes, slides, fine arts; glass plates and delicate graphics.

Alinari is a leading traditional and multimedia photographic publisher and is a synonym for the highest quality in the production of artistic prints. Alinari represents an irreplaceable landmark for preserving, cataloguing, making known and handing down, through photography, Italian and European history, society, art and culture. The unique heritage of the Alinari collections gives life to one of the biggest international centres of photographic and iconographic documentation with over 3.5 million vintage images from the 19th and 20th century from all over the world.

The Digital Archive was inaugurated in 2001. It continues to grow and progress constantly with images that can be consulted on line. Today there are over 200,000 pictures available on the Alinari Archives business site and 100,000 in the EDUCATION section.

Description of the **metadata schema** in use at Alinari

Name	1-Alinari internal schema based on the ICCD and University of Florence guidelines; 2-Dublin Core; 3-Open Archive Initiative (from Oct 2006)
Acronym	1-none; 2-Dublin Core; 3-OAI
Status / version	Not applicable
Type	1-a proprietary metadata schema based on a national standard, namely: ICCD (Istituto Centrale del Catalogo) 2-a proprietary metadata schema: Dublin Core with additional qualifiers 3-an international standard, namely: Open Archive Initiative
Management	Alinari
Short description	The business.alinari.it website presents an initial selection of around 200,000 pictures including historical nineteenth and twentieth century vintage prints in sepia and black and white, and colour photographs from 1920 to the present, chosen by teams of experts in the fields of traditional and electronic publishing, audiovisual, television and cinema communication, history of photography.
Number of elements	1. Internal schema: 30+ elements 2. Dublin Core: 21 elements (18 Dublin Core elements and 3 additional elements). Please see annex 5 for the element definition. 3. OAI schema: under construction
Applied by the	1-none

following organizations e.g.	2-see www.dublincore.org 3-see www.openarchives.org and Italian Ministry for Culture
URL(s) documentation	1- www.alinari.com (documentations are not open to the general public) 2- www.dublincore.org 3- www.openarchives.org/
XML encoding available ( )	Yes

Description of the **controlled vocabularies** in use at Alinari

Name	1- The thesaurus is built as a hierarchical tree with 61 primary classes and 8000 key words 2- Authority lists with structured categories and controlled vocabulary for person names, period names, geographic locations 3-MPEG visual descriptors (only for R&D) in CBR system
Acronym	None
Status / version	Not applicable
Type	Not applicable
Management	Alinari
Short description	GEOGRAPHICAL: functional in the search for sites. For Italy the thesaurus allows searching by region, province and municipality. For other countries, the search can be by nation. ICONOGRAPHIC: a dictionary of selected terms allows a search via subject, ordered in 61 iconographic classes from Agriculture to Zoology. PERIODS AND STYLES regards searches involving pictures of art and allows for a search by art style and/or historical period. TYPE OF WORK OF ART offers a subdivision of subjects of artistic and historical-artistic importance and therefore allows searches by type of object and/or architectural complex.
Number of elements	200,000
Available in language	200,000 in EN, IT; 50,000 in EN, IT, PL, GE, SP, FR
XML encoding available	Yes
Extra information on application	Main catalogue metadata Photographer information: • ID number of the image (archive) • Photographer Surname • Photographer Name • Date of shoot • Complete place location Work information: • Title (Italian) • Title (English) • Date of work of art • Date ISO (start, end) Support information: • Colour-B/W

	• Positive/Negative • Type of object • Technique • Dimensions • Format Other information: • Key words • Conservation state • Description • Copyright information
Applied by the following organizations e.g.	Not applicable
URL(s) documentation	Not applicable

3.2 Netherlands Institute for Sound and Vision

The Netherlands Institute for Sound and Vision (Nederlands Instituut voor Beeld en Geluid) curates, and provides access to, 70 per cent of the Dutch audiovisual heritage. In total, around 700,000 hours of television, radio, music and film, making Sound and Vision one of the largest audio-visual archives in Europe.

Sound and Vision is the 'working archive' of the national broadcasting corporations, a cultural history institute and also a unique media experience for its visitors. Media production professionals use the collections for new programmes and the archive is a unique source of information for research, not only for students and academics, but also for journalists.

Furthermore, the audiovisual material is a most valuable addition to traditional teaching methods, hence Sound and Vision also promotes the use of media in education.

Sound and Vision has always been at the forefront using new media technology. A large-scale project was launched a few years ago to keep anticipate the developments in media production and meet changing demands from end-users. The first stage has been finalized and as of 2006, key parts of the internal workflows have been automated:

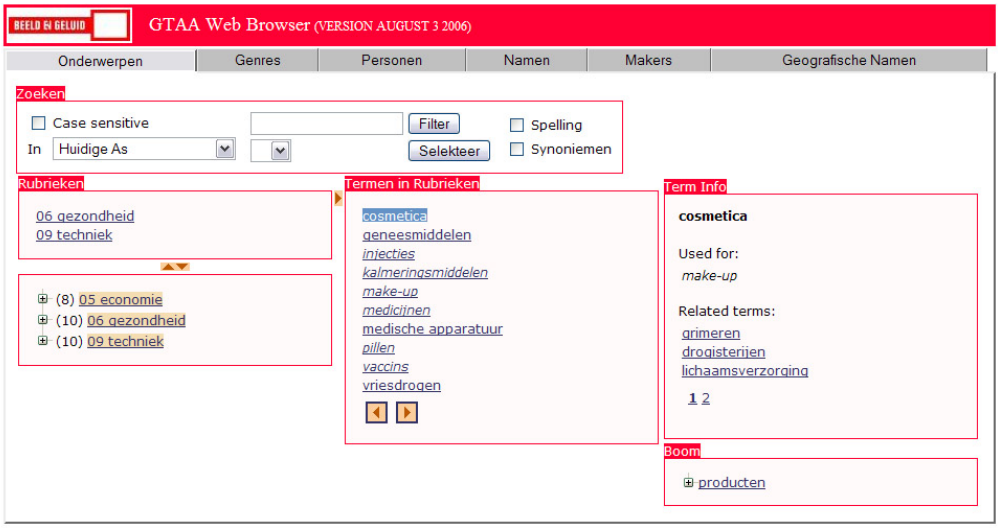
- ingest of material and metadata straight from the broadcast facilities
- installment of large storage facility
- tape less handling of requests (new material)
- encoding legacy material

Description of the **metadata schema** in use at Sound and Vision

Name	iMMix schema, based on International Federation of Library Associations and Institutions (IFLA)
Acronym	IMMix
Status / version	Not applicable
Type	Proprietary standard. References for the form and content of the relevant metadata are Dublin Core, SMPTE and P_META. The IFLAmodel FRBR has been the most important reference for modelling the metadata. XML (and AXF) are used for the exchange format.
Management	local
Short description	The metadata model defines the way the metadata should be structured. It is roughly divided in four stages: concept, actual realisation, physical embodiment, and carrier. Those four stages represent different layers in the model. • Work: the name of the intellectual and artistic concept or idea which is the foundation of one or more realisations. • Realisation: a realisation is an elaboration of a concept: a specific single or multiple productions. Every realisation has a clear structure and form of content compared to other realisations of the same work. In the case of several productions the realisation contains all data that are valid for the underlying expressions. • Series: a series is a group of expressions, usually decided by the makers or producers. A series has a beginning and an end. • Expression: an expression is a concrete result of a realisation or series. It is the actual production. • Selection: a selection is a part of an expression. For example a news item. • Publication: a publication of an expression; i.e. a television broadcast. • Position: the position of the publication on a carrier. • Carrier: description of a carrier.
Number of elements	85 (will be extended)
Extra information on application	In addition to the metadata model which defines the way the metadata are to be structured, two other means have been developed: Formats: these define in which way the fields and metadata are to be presented to the documentalist. Intentions: these define which metadata should be available to fulfil the specific information needs of a specific target group. The idea is that a broadcast professional prefers objective annotations whereas, for a visitor of the future Beeld en Geluid Media experience, a more catchy description is needed. This principle also facilitates the desire to add domain specific and other additional data to a description.
Applied by the following organizations e.g.	Sound and Vision
URL(s) documentation	http://www.prestospace.org/project/deliverables/D15-1_Analysis_AV_documentation_models.pdf
URL guidelines for application	http://www.prestospace.org/project/deliverables/D15-1_Analysis_AV_documentation_models.pdf
XML encoding available ( )	Yes

Description of the **controlled vocabularies** in use at Sound and Vision

Name	Common Thesaurus for Audiovisual Archives
Acronym	GTAA
Status / version	Not applicable
Type	Hierarchical thesaurus, proprietary
Management	Sound and Vision
Short description	The GTAA thesaurus is the controlled vocabulary used at The Netherlands Institute for Sound and Vision. GTAA stands for the Common Thesaurus for Audiovisual Archives; it is the result of the collaborative work of different institutions concerned with audiovisual documents indexing, including the Filmmuseum Amsterdam.[32]
Number of elements	It contains 159.831 preferred terms, 1.900 non-preferred terms, and 88 categories.
Available in language	Dutch
XML encoding available ( )	Yes / Also in RDF
Extra information on application	The GTAA is a general thesaurus with multiple facets: subjects, genres, persons, makers, names and locations. Only the subject facet, which contains the keywords, is structured. The terms in the subject facet are related to others via the related term, broader term and narrower term relations. The types of information we are looking for (keywords, persons, locations, names, makers and genre) are very closely related to the different facets of the GTAA.
Applied by the following organizations e.g.	Sound and Vision, Filmmuseum Amsterdam
URL(s) documentation	http://ems01.mpi.nl/CHOICE/ (the browser is also accessible online)
URL guidelines for application	http://www.cs.vu.nl/~guus/papers/Assem06b.pdf#search=%22GTAA%20malaise%22. See image below. Viewed 2006-10-19



[32] http://ems01.mpi.nl/CHOICE/

3.3 Metadata and the Institutes from the Advisory Board

The British Broadcasting Company (BBC)

The BBC has developed the SMEF metadata standard. See paragraph 2.7.1 for more details. The BBC is part of the Advisory Board and represents also FIAT/IFTA (International Federation of Television Archives).

Österreichische Nationalbibliothek (ONB)

The ONB manages the following cultural heritage objects:

- books
- manuscripts
- photographs
- slides
- glass plates
- delicate graphics
- three-dimensional objects.

ONB uses a proprietary metadata schema based on a standard, namely: Amico, for describing the objects this institute manages. See the description of AMICO Library in section 3.4.4 for further details on this standard.

There are three controlled vocabularies in use namely:

- Schlagwortnormdatei **SWD** for assigning keywords;
- Personennamendatei **PND** (German Name Authority);
- Gemeinsame Körperschaftsdatei **GKD** (German Corporate Headings Authority).

These knowledge organization systems are monolingual, German, and not available in XML.

ONB is involved in the following special project on the (digital) access of cultural heritage information:

Bildarchiv Austria (http://www.bildarchiv.at). This is a webportal of the Austrian National Library, offering a large stills collection. Metadata and preview thumbnails of ORF archive's stills are included in the CMS. On demand high-resolution digitization and ordering of historical photographs from the ONB and the ORF collections.[33]

National Archive of Poland (Archiwa Panstwowe, Polon)

The Polish archives use the metadata schema Dublin Core for the digital objects, both digitized analog documents and digital born documents. The recommendations of the International Council on Archives ISAD(G) and ISAR (CPF) are in use for the description of archival objects. Occasionally MARC21 is used for the manuscripts.

The National Archive of Poland (http://www.archiwa.gov.pl/) is developing a new national metadata standard at this moment for archiving the electronic documents from the records system of the current institutions and agencies. It is based on the Dublin Core. In practice it will be a Polish version of the chosen elements from the DC Metadata element Standard.

[33] The other CH member of the Advisory (Staatliche Museen zu Berlin) was also contacted and solicited for their input. Unfortunately, no reply was received in time for this deliverable. However, any input from this museum will be considered when making final decisions as to the metadata schema and ontologies to be adopted by MultiMatch

As to the controlled vocabularies:

- Polish state archival record use all recommended ISO standards according to the ISAD(G)
- Some local lists of keywords are used in the electronic inventory to the fonds of the Polish state archives. The inventory is available online as the SEZAM database: http://baza.archiwa.gov.pl/sezam/index.eng.php.

3.4 Selection of Related European Projects

This section provides descriptions of seventeen related (primarily European) projects that are working on scientific problems that are closely connected to MultiMatch. They are the best known (primarily European) projects looking at manipulating/representing cultural heritage data in some way. MultiMatch keeps track of their research results and will participate in workshops as they might be of interest for the project. Either to:

- disseminate project results
- keep track (and possibly influence) standardization efforts
- collaborate on solving research questions and
- reuse technology if appropriate (maybe even co-design software)

The projects are loosely ordered by probable relevance to the MultiMatch project. MICHAELplus and The European Library are described in more detail, as they represent the largest and most prominent endeavors currently undertaken regarding the disclosure of European cultural heritage. Both projects have also gained considerable political support. Note that this paragraph is not intended to be comprehensive, but the domain is well covered by these sixteen initiatives.

3.4.1 The European Library (TEL)

The European Library[34] is a service on the World Wide Web which offers access to the resources of the 45 national libraries of Europe. The resources are both digital and non-digital, and include books, magazines, journals, and other resources. TEL offers free searching, and delivers digital objects: some gratis, some at cost. This service is developed and maintained by a consortium of these national libraries.

The Vision of The European Library is: *"Provision of equal access to promote world-wide understanding of the richness and diversity of European learning and culture."*[35]

Development

The service was created by a cooperation of 9 national libraries and CENL (Conference of European National Librarians) under the TEL (The European Library: Gateway to Europe's Knowledge) project. TheEuropeanLibrary.org service was launched on 17 March 2005. There are searchable collections from 19 national libraries available at the portal, with access to further collections of 21 national libraries as hyperlinks. The national libraries of Austria, Croatia and Serbia joined in 2005.

In the timeframe of TEL-ME-MOR (The European Library: Modular Extensions for Mediating Online Resources) project that lasts from 2005 to 2007, the national libraries of 10 New Member States of the European Union will also join the service. On 1 January 2006, the national libraries of Estonia and Latvia as well as the Danish Royal Library became the full members of the consortium. On 1 July 2006, the national libraries of the Czech Republic and Hungary joined The European Library.

In September 2006, the Digital Library project (EDL project) was launched; enabling a further 9 national libraries to be brought into the network. Countries involved are either members of the European Union or the European Free Trade Association: Belgium, Greece, Iceland, Ireland, Liechtenstein, Luxembourg, Norway, Spain and Sweden. The action will enrich The European Library with up to 100 new collections.

[34] www.europeanlibrary.org

[35] http://libraries.theeuropeanlibrary.org/aboutus_en.html

Search and retrieval

The European Library uses SRW (Search/Retrieve Web service) for search and retrieval. SRW is a Web Service, providing a SOAP interface to queries, to augment the URL interface provided by its companion protocol SRU (Search/Retrieve via URL). Queries in SRU and SRW are expressed using the Common Query Language (CQL).

SRW was announced as an alternative to Z39.50 under the umbrella of the Z39.50 international Next Generation [ZiNG]: two access mechanisms were proposed:

- SRW (Search and Retrieve via the Web), based on the use of SOAP2 and
- SRU (Search and Retrieve via URLs), based on the use of URLs. Both approaches offer a lower implementation barrier than Z39.50 and are more amenable to implementation in modern systems.

SRU is the simpler of the two mechanisms. With SRU, a search request takes the form of a base-URL and associated parameters, such as the query, the start record, the maximum number of returned records and the record schema.[36]

Metadata and TEL

The metadata model being used is a Dublin Core Application Profile. The project acknowledges that these functional requirements will change over time and therefore the metadata model will need to be able to evolve in a controlled way. A metadata registry was developed to fulfil this need.

The European Library is using the Dublin Core library profile[37]. This application profile defines the following:

- required elements
- permitted Dublin Core elements
- permitted Dublin Core qualifiers
- Permitted schemes and values (e.g. use of a specific controlled vocabulary or encoding scheme)
- library domain elements used from another namespace
- additional elements/qualifiers from other application profiles that may be used (e.g. DC-Education: Audience)
- refinement of standard definitions

To check whether the Library Application Profile met functional requirements, the elements from the Library Application Profile were mapped to a number of basic functions required by the TEL portal.

[36] http://www.dlib.org/dlib/february04/vanveen/02vanveen.html#ZiNG

[37] http://dublincore.org/documents/2002/09/24/library-application-profile/

Element	Qualifier/ Scheme/ Role	Search/resource discovery	Retrieval of metadata	Identification	Description	Link service	Multilinguality	Thesaurus service	Collection level	Authorisation	Administration	Hardware & Software	Navigation	Copy cataloguing	Miscellaneous	Comment
Identifier	Any		X	X		X				X					X	Identifies the metadata record
Identifier, Source, Relation	Base-URL		X	?		X			X				X		X	Encoding scheme for a URL with variable query
Identifier, Source, Relation	URN			X		X			X				X		X	Encoding scheme for URNs
Identifier	PURL			X		X			X				X		X	Encoding scheme for persistent URLs
Identifier, Source, Relation	OpenURL			X		X			X				X		X	Encoding scheme for queries with a variable base-URL
Collection level description namespace	All	X			X				X	X	X		X	?	?	

The end result of this process is that all the elements needed for specific functionality were revealed and the total behaviour of the portal that depends on the presence of specific metadata was identified. The mapping process produced two different application profiles, one for objects and one with additional elements for collections.

The most significant additional elements concerned collection level descriptions [CLD] is linking to objects and services and record identification.

3.4.2 MICHAELplus

The MICHAEL project has developed an electronic system to access, manage and update existing digital records of Europe's collections, including museum objects, archaeological and tourist sites, music and audiovisual archives, biographical materials, documents and manuscripts. MICHAEL culminates several progressive efforts under the eEurope Action Plan to harmonise EU Member States' programs to scan, photograph and otherwise enter cultural records into digital databases. It promotes standards, best practices and guidelines for digitisation that were originally proposed by the National Representatives Group's Lund Principles in Lund, Sweden, in 2001. It also follows up on ideas developed in the IST-funded MINERVA and MinervaPLUS projects.

Beginning in June 2004, the project assembled a consortium of public cultural institutions and private IT companies in France, Italy and the UK. It received 3 million euro from the eTen programme. MichaelPLUS, which began 1 June 2006, adds partners from 11 more countries that have since found funds to participate. The key objectives of MICHAELplus are to extend the number of countries involved in the MICHAEL project, set-up to launch an online trans-European service to enable European cultural heritage to be promoted to a worldwide audience.

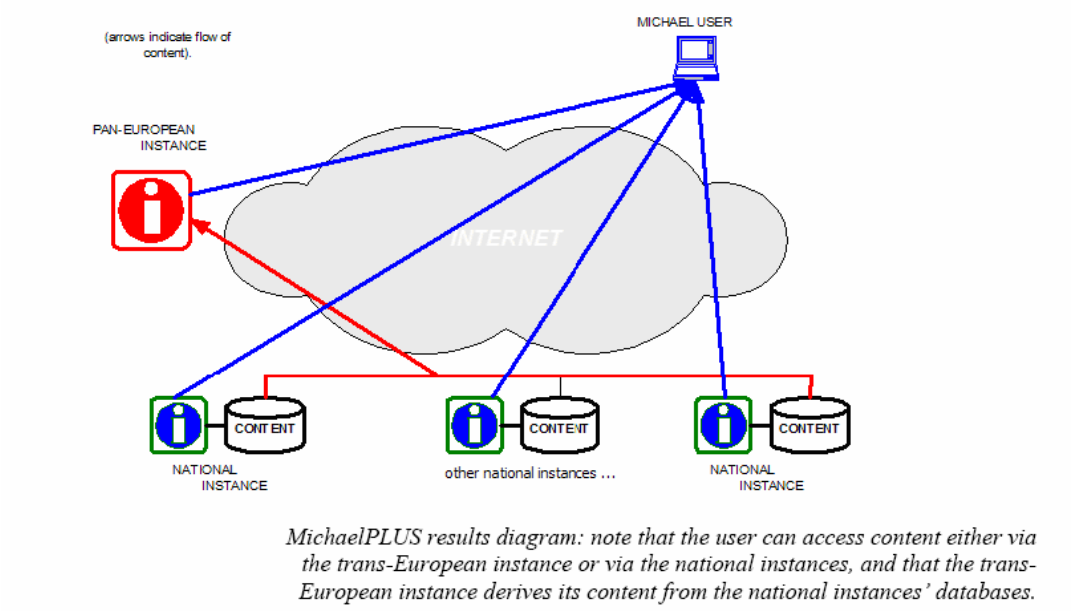
This inventory, set up by public institutions, will use a distributed and Open Source platform suitable to be extended to any other country. Due to the high political commitment of several Member States, the online service aims at becoming the European de facto standard for the exchange of information on digitised cultural contents.

To achieve these objectives, MICHAELplus is composed of five main activities:

1. Adaptation of the existing common open source platform and definition of its architecture to the characteristics of the new participating countries
2. Survey and gathering of information about existing digital cultural collections, harvesting of existing data, and the entry of new data onto the system
3. Definition of local and transEuropean communication and marketing plans to promote the service to the key target groups: education and research, cultural tourism, creative industries
4. Organisation of the technical infrastructure to manage the system, update content, and maintain the software, including training of its staff
5. Design and implementation of the organisation infrastructure of the delivery system, including the management structure to run the transnational service.

The technical results of the MICHAELplus project are:

- National inventories on a common meta-data model, data model and service model
- National portals running on a common open source technical platform, localized as necessary
- Trans-national inventory portal
- Sustainable, flexible extensible model based on XML technologies
- Open source solution built on Apache Tomcat, Cocoon, XtoGen, etc.
- Methodology and model, which is easy to deploy and replicate in additional countries.



MichaelPLUS results diagram: note that the user can access content either via the trans-European instance or via the national instances, and that the trans-European instance derives its content from the national instances' databases.

Metadata and MICHAELplus

The metadata set consists of three levels:

- Item-level metadata: The Dublin Core meta-data Element Set. This is by far the dominant meta-data set for item-level descriptions. It includes 15 core elements and the possibility to add extensions as needed.
- DC.Culture: enables the searching of item-level metadata through 4 key access points – Who, What, Where and When.
- Collection-level meta-data: The Research Support Libraries Programme (RSLP) collection-level metadata model forms the basis for the collection-level metadata to be used in MICHAEL. It

includes 28 basic elements, of which 13 are ‘general’; many of these are based on Dublin Core elements.[38]

The MICHAEL data model is meant to describe collections. The following information is covered:

- Descriptive information about the collection
- Information about the collector
- Information about the owner of the collection
- Information about the physical location or point of access to the collection
- Information about an administrator responsible for the collection[39]

More information on the data model can be found online.[40]

The database of MICHAEL[41] uses the following controlled vocabularies:

UK Archival Thesaurus	[http://www.ukat.org.uk]
DCMI Type Vocabulary	The DCMI Type Vocabulary provides a general, cross-domain list of approved terms that may be used as values for the Resource Type element to identify the genre of a resource. [http://dublincore.org/documents/dcmi-type-vocabulary]
ISO 639-2: languages Codes for the representation of names of languages.	[http://www.loc.gov/standards/iso639-2/]
ISO 3166: countries	Codes for the representation of names of countries and their subdivisions. [http://www.iso.org/iso/en/prods-services/iso3166ma/02iso-3166-codelists/ list-en1.html]
UK Local Authorities	[http://www.statistics.gov.uk/geography/geographic_area_listings/ad mi nistrative.asp#02]
UK Educational Levels	[http://www.ukoln.ac.uk/metadata/education/ukel/]

3.4.3 BRICKS

BRICKS aims at integrating the existing digital resources into a common and shared Digital Library, a comprehensive term covering “Digital Museums“, “Digital Archives“ and other kinds of the digital memory systems.

Its "bottom-up” approach, which is based on the interoperability of a dynamic community of local systems, maximizes the use of existing resources and know-how, and, therefore, national investments. BRICKS will contribute in:

- tuning the mission of memory institutions in the digital era,
- developing a shared vision for the exploitation of digital cultural content and
- encouraging cultural cooperation for the construction of an interoperable cultural capital.

BRICKS architecture will be decentralized, based on a peer-to-peer (P2P) paradigm, i.e. no central server will be employed. Every node represents a member institution, where the software for accessing the BRICKS

[38] http://cores.dsd.sztaki.hu/?class=http%3A%2F%2Fwww.cores-eu.net%2Fregistry%2Freg%2FElementSet;resource=http%3A%2F%2F

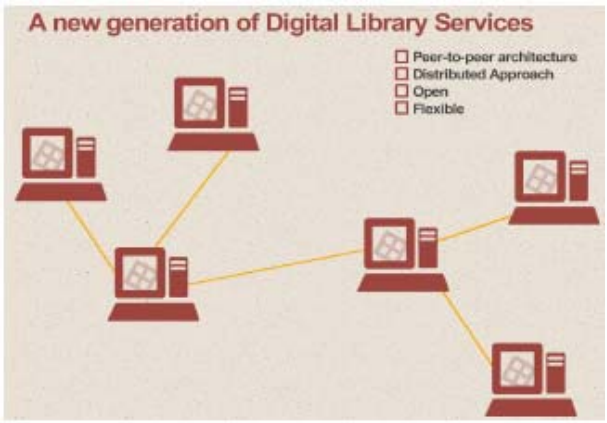
[39] http://www.michael-culture.org/doc.html

[40] http://www.michael-culture.org/documents/MICHAELDataModelv1_0.pdf

[41] In the follow-up project MICHAEL PLUS, this might be changed.

is installed. Such nodes are called BNodes. BNodes communicate among each-other and use available resources for content and metadata management. Every BNode knows directly only a subset of other BNodes in the system.

However, if a BNode wants to reach another member that is directly unknown to it, it will forward request to some of its known neighbour BNodes that will deliver the request to the final destination or forward again. To install a BNode, the member has just to download the software package and install it on the machine that will be used as a BNode.[42]



3.4.4 AMICO The Art Museum Image Consortium Library

The Art Museum Image Consortium (AMICO)[43] was a not-for-profit organization of institutions with collections of art, collaborating to enable educational use of museum multimedia. AMICO operated from 1997 to 2005. The AMICO Library had almost 40 members from Canada, the US and the UK. The Getty Institute, the Library of Congress and the Montreal Museum of Fine Arts

Each work of art in The AMICO Library is documented by:

- a Catalogue Record (based on the CDWA)
- an image file and
- an Image or Media Metadata Record (based on Dublin Core) for the related image or media file.
- Additional media files may also be present. Each of these has a metadata record.

Images in the AMICO Library include a broad range of works of the following genres: painting, sculpture, photography, print, drawing, ceramic, textiles, metalwork, furniture, books and scrolls, architecture, and archaeological finds. In 2002 the Library contained documents on over 100,000 works of art.

Several proprietary vocabularies are used, for instance for object type (the kind of work of art that is described), as well as AAT and the Library of Congress Thesaurus of Graphic Materials.

[42] http://www.brickscommunity.org/Partner/metaware/evamoscow05/file/#search=%22bricks%20community%20dlib%22

[43] http://www.amico.org/AMICOlibrary/dataspec.html

3.4.5 aceMedia

aceMedia is an IP project (IST-2002-2.3.1.7) - Integrating Knowledge, Semantics and Content for User Centred Intelligent Media Services, whose results will be taken in strong consideration for the development of integrated synergies with MultiMatch. aceMedia is an ongoing project targeted to extract and exploit meaning inherent to content in order to automate annotation and to add functionality that makes it easier for all users to create, communicate, find, consume and reuse audiovisual content. Within the framework of aceMedia, a multimedia ontology infrastructure has been constructed.

The multimedia ontology infrastructure consists of a Multimedia Structure Ontology (MSO), which models the top-level hierarchy structure of multimedia documents, a Visual Descriptor Ontology (VDO), which models the MPEG-7 visual part of MPEG-7 and a Visual Descriptor Extraction (VDE) tool for semi-automatically annotating multimedia documents.

3.4.6 DELOS Network of Excellence

DELOS Network of Excellence[44], coordinated by ISTI-CNR, defines unifying and comprehensive theories and frameworks over the life cycle of Digital Libraries information and studies interoperable multimodal/multilingual services and integrated content management. A strong relationship will be established with DELOS in order to disseminate and promote project results and to have a continuous view of the evolution of multimedia content management in Digital Libraries. ISTI-CNR is coordinator of this project.

For MultiMatch, the activities and results of DELOS WP5 “knowledge extraction and semantic interoperability” are of primary relevance.

3.4.7 Knowledge Web Network of Excellence

Knowledge Web[45] is a 4 year Network of Excellence project (started 2004) funded by the European Commission 6th Framework Programme. Knowledge Web began on January 1st, 2004. Supporting the transition process of Ontology technology from Academia to Industry is the main and major goal of Knowledge Web. The mission of KnowledgeWeb is to strengthen the European industry and service providers in one of the most important areas of current computer technology: Semantic Web enabled E-work and E-commerce.

In relation to MultiMatch, the most prolific Workpackages within Knowledge Web include:

- WP2.1: Scalability
- WP2.4: Semantic Web Services
- WP2.5: Semantic Web Language Extensions

[44] http://www.delos.info/

[45] http://knowledgeweb.semanticweb.org/

3.4.8 MACS (Multilingual Access to Subjects)

This CENL (Conference of European National Librarians) initiative aims to provide multilingual subject access to library catalogues.[46] It enables users to simultaneously search the catalogues of the project's partner libraries in the language of their choice (English, French, German). This multilingual search is made possible thanks to the equivalence links created between the three indexing languages used in these libraries: SWD (for German), RAMEAU (for French) and LCSH (for English). Topics (headings) from the three lists are analysed to determine whether they are exact or partial matches, of a simple or complex nature. The end result is neither a translation nor a new thesaurus but a mapping of existing and widely used indexing languages.

3.4.9 W3C Semantic Web Best Practices and Deployment Working Group

The aim of this Semantic Web Best Practices and Deployment (SWBPD) Working Group[47] is to provide hands-on support for developers of Semantic Web applications. With the publication of the revised RDF and the new OWL specification we expect a large number of new application developers. Some evidence of this could be seen at the last International Semantic Web Conference in Florida, which featured a wide range of applications, including 10 submissions to the Semantic Web Challenge. This working group will help application developers by providing them with "best practices" in various forms, ranging from engineering guidelines, ontology / vocabulary repositories to educational material and demo applications.

3.4.10 ECHO European Chronicles Online

ECHO[48] was funded by the European Community within the Fifth Framework Program and ran from 2000 to 2004. The Project aimed to develop a Digital Library service for historical films belonging to large national audiovisual archives.

> "To enable resource discovery on the ECHO film collections over the Web, this project defined content description standards or metadata standards for complex, multi-layered, time dependent, information rich data streams. The multi-layered description of audio-video documents makes it possible to describe different aspects of the same document. The hierarchical description of audio-video documents supports both (i) interoperability among different archives at the corresponding hierarchical levels and (ii) satisfying the needs of special interest user communities. This last point is obtained by refining the model by adding specialised descriptions.
>
> This multi-layered and hierarchical structuring is in accordance with the IFLA model: FRBR. The metadata elements of the ECHO metadata schema are divided over the seven entities of the ECHO metadata model[49]: AVDocument, Version, Video, Audio, Transcript, Media and Storage.
>
> Several proprietary controlled vocabularies are in use, e.g. Format (Media), Genre, Theme, Type of segment, as well as several controlled vocabulary standards, namely: Description language, Subtitle language and Audio language (ISO 639 Code for the representation of the names of languages) and (ISO 3166 Codes for the representation of names of countries)."

[46] http://www.ddb.de/eng/wir/projekte/macs.htm

[47] http://www.w3.org/2001/sw/BestPractices/

[48] http://pc-erato2.iei.pi.cnr.it/echo/

[49] http://pc-erato2.iei.pi.cnr.it/echo/public/deliv/D3-1-1%20ECHO%20Metadata%20Modelling.pdf

3.4.11 Renardus

The service (http://www.renardus.org/#DDC2) aims to provide a trusted source of selected, high quality Internet resources for those teaching, learning and researching in higher education in Europe. Renardus provides integrated search and browse access to records from individual participating subject gateway services (data providers) across Europe.

The Renardus service grew from a project funded 1 January 2000-30 June 2002 by the EU's IST 5th framework programme. Renardus allows users to find Internet resources selected according to quality criteria and carefully described by Subject Gateways from several European countries. The user can discover the individual resources and collections by searching and browsing these descriptions (metadata), not the full text of the resources themselves. Having selected the most relevant ones informed by the descriptions, the URL's provided direct the user to the original resources.

The Renardus partner gateways cover about 64000 predominantly digital web-based resources from within most areas of academic interest, mainly written in English. The Renardus Subject Gateways map their local browsing structures and classification systems to a common universal system, the Dewey Decimal Classification (DDC) in the French version.

3.4.12 SCULPTEUR

One of the goals of the European Project SCULPTEUR (http://www.sculpteurweb.org/ ; 2002-2005), was to use Semantic Web techniques to enable different institutions, namely museums and libraries, to preserve their specific needs and commitments, while ensuring the interoperability of the databases they produce.
This resulted in the development of a prototype for a query interface on heterogeneous databases, called "Concept Browser."[50] This tool is driven by an ontology adapted from the CIDOC CRM conceptual model that was developed by ICOM CIDOC (to be published as ISO standard 21127 in the near future). The data structure from each of the museum databases involved in the project was mapped to the CIDOC CRM. In addition, it should be possible to integrate bibliographic databases to the Concept Browser, as the library format called UNIMARC was mapped to the CIDOC CRM as well. The Concept Browser, which is based on TouchGraph technology, allows one to visualise the ontology itself.[51]

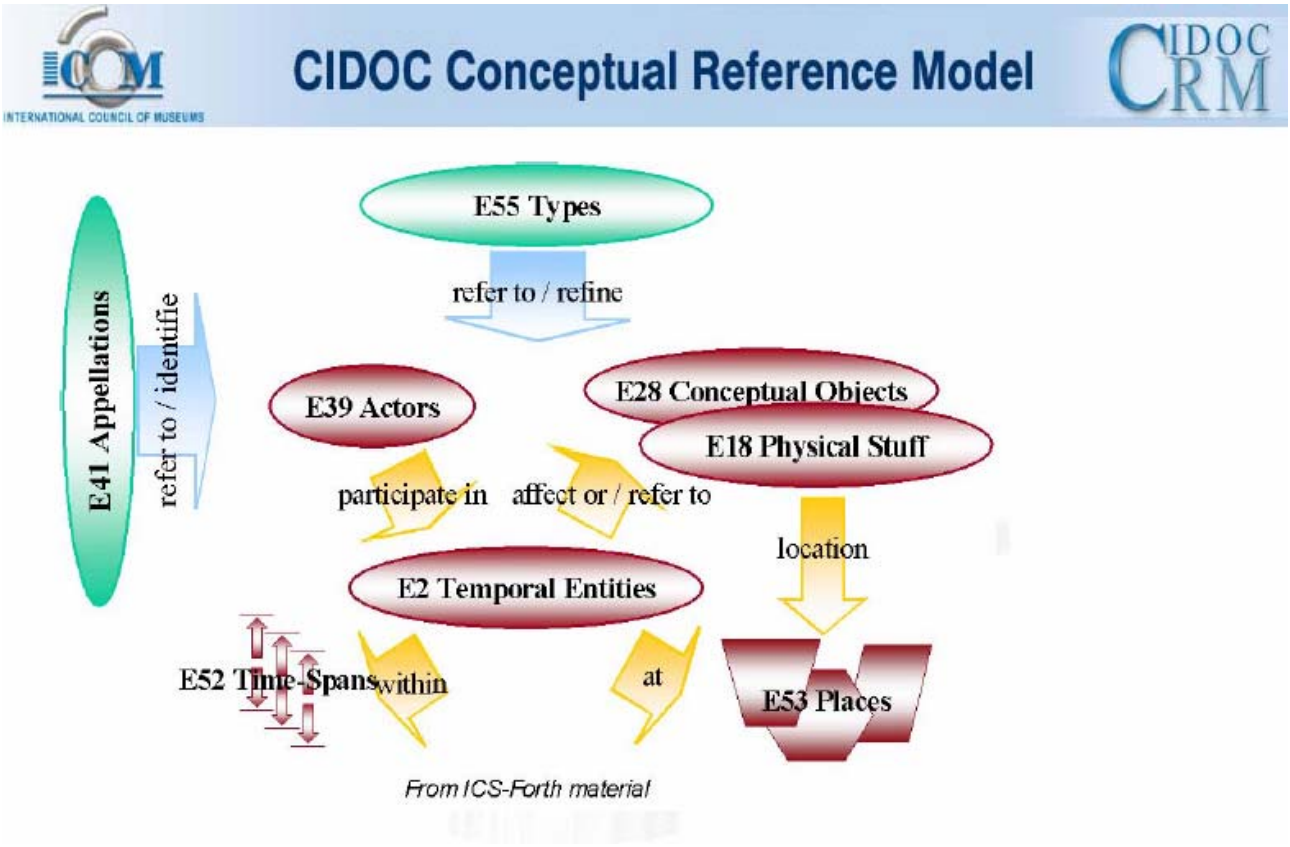
The ontology applied is focused on artefacts within the museum/gallery domain. It contains 81 classes and 139 properties.

The following concepts play a role within the data model of SCULPTEUR:

- People (who: artist, curator, owner, restorer)
- Art objects and representations (what: painting, sculpture, paving stones, coins, Polynesian artefacts, films, digital representations)
- Events and activities (when: creation, acquisition, restoration, loan, birth, death, period)
- Places (where: gallery, archive, conservation centre, private collection, country, city, town, studio)
- Methods and techniques (how: oil, watercolour, carving, x-ray, restoration technique).

[50] http://sculpteur.it-innovation.soton.ac.uk/auth/login.jsp#

[51] Using an ontology-driven system to integrate museum information and library information / Patrick Le Bœuf (Bibliothèque nationale de France) Paper presented on the occasion of the Symposium on Digital Semantic Content across Cultures, Paris, the Louvre, 4-5 May 2006



The Concept Browser offers the following query functionalities:

- searching by content with the possibility to ask typical museum/gallery domain questions like:
  - Find objects that have a pattern / colour similar to this image.
  - Find vases that are a similar shape to this one.
  - Find paintings with cracks like this.
  - Which mould was used to make this figurine?
  - searching by 2-D content on the basis of colour, texture and shape; including sub-image matching.
  - searching by 3-D content on the basis of "Cord" histograms or shape distribution.

3.4.13 Web Gallery of Art

The Web Gallery of Art[52] is a virtual museum and searchable database of European painting and sculpture of the Gothic, Renaissance and Baroque periods (1100-1850), currently containing over 15.400 reproductions. Commentaries on pictures, biographies of artists are available. Guided tours, free postcard and other services are provided for the visitors.

The controlled vocabulary in use is monolingual: English. The proprietary metadata structure for the description of the works of art includes the following metadata elements: author, title, time-line, school, form, size, technique, type, location, owner and some data on the actual file (pixels, colour, kBytes). The biography of artist is a text file linked to the author/artist name.

[52] http://www.wga.hu/index.html

3.4.14 MEMORIES

Design of an audio semantic indexation system allowing information retrieval for the access to archive content.[53] This project will create a generic software library in order to facilitate the extraction of high level information from audio signals. The main expected innovations of the MEMORIES project are a user-friendly system that matches archivist needs for information retrieval in audio databases, the definition of a format for database structuring of information content descriptors and an efficient tool for audio restoration.

3.4.15 MUSCLE Network of Excellence

MUSCLE[54] is an EC-sponsored Network of Excellence that aims at establishing and fostering closer collaboration between research groups in multimedia datamining and machine learning. The Network integrates the expertise of over forty research groups working on image and video processing, speech and text analysis, statistics and machine learning. The goal is to explore the full potential of statistical learning and cross-modal interaction for the (semi-)automatic generation of robust meta-data with high semantic value for multimedia documents.

In particular, MUSCLE researchers are developing software tools and research strategies to enable users to move away from labor-intensive case-by-case modeling of individual applications, and allow them to take full advantage of generic adaptive and self-learning solutions that need minimal supervision. MUSCLE is one of the members of the MultiMatch Advisory board.

For MultiMatch the research in the workpackage “Representation and Communication of Data and Metadata” (WP5) is most relevant and will be monitored.

3.4.16 School of Information and Library (SILS) Metadata Research Center

The School of Information and Library (SILS) Metadata Research Center[55] at the University of North Carolina at Chapel Hill has been established to advance research in the area of metadata, semantics and ontologies.

Digital resource repositories representing nearly every domain (e.g., scientific, educational, commerce, medical, cultural information domains) are growing rapidly, particularly as people, organizations, and enterprises, increasingly turn to the Web for recording, preserving, and disseminating information. Commercial search engines using statistical algorithms facilitate resource discovery on the Web, often providing satisfactory results. As digital repositories and the global Web continue to grow, retrieval effectiveness and search engine scalability can decline. Metadata is an important means for enhancing resource discovery, and is a fundamental component of the Semantic Web.

This predicament invites a host of research questions requiring rigorous investigation: Who should create metadata? What metadata-generation techniques (e.g., automatic and/or manual) are the most efficient and effective? What features should be included in metadata generation applications? The Center has been established to lead and facilitate research efforts addressing these and other metadata questions, including research focusing on semantics and ontologies.

[53] http://www.memnon.be/FutureDevelopment/Lesd%C3%A9veloppementsfuturs/tabid/124/Default.aspx

[54] http://www.muscle-noe.org/

[55] http://ils.unc.edu/mrc/

3.4.17 BIRTH and Video Active

The BIRTH Television Archive is the world's first internet archive of vintage films from the early times of European television. (http://www.birth-of-tv.org/birth/) This unique archive was built up by five major European television archives from the British Broadcasting Corporation (BBC), Österreichischer Rundfunk (ORF), Nederlandse Instituut voor Beeld and Geluid, Radio Télévision Belge Francophone (RTBF) and Südwestrundfunk (SWR) together with the technical partners Joanneum Research and Noterik Multimedia.

The descriptions in the BIRTH database have the format of the Dublin Core metadata schema extended with two metadata elements (keywords and location) for the sake of multilingual searching. The follow up project Video Active (www.videoactive.eu) also uses this metadata schema.

BIRTH and Video Active use the following controlled vocabularies:

ISO 3166-1 for country codes

IPTC thesaurus for subject keywords (see also section 2.3.2 for further details).

3.4.18 CASPAR (Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval)

CASPAR is an Integrated Project co-financed by the European Union within the Sixth Framework Programme .

CASPAR intends to:

- Implement, extend, and validate the OAIS reference model (ISO:14721:2002)
- Enhance the techniques for capturing Representation Information and other preservation related information for content objects
- Design virtualisation services supporting long term digital resource preservation, despite changes in the underlying computing (hardware and software) and storage systems, and the Designated Communities.
- Integrate digital rights management, authentication, and accreditation as standard features of CASPAR.
- Research more sophisticated access to and use of preserved digital resources including intuitive query and browsing mechanisms
- Develop case studies to validate the CASPAR approach to digital resource preservation across different user communities and assess the conditions for a successful replication.
- Actively contribute to the relevant standardisation activities in areas addressed by CASPAR.
- Raise awareness about the critical importance of digital preservation among the relevant user-communities and facilitate the emergence of a more diverse offer of systems and services for preservation of digital resources.

More information can be found online: http://www.casparpreserves.eu/

3.4.19 DRIVER (Digital Repository Infrastructure Vision for European Research)

DRIVER sets out to build the testbed for a future knowledge infrastructure of the European Research Area. Aimed to be complimentary to GEANT2, the successful infrastructure for computing resources, data storage and data transport, DRIVER will deliver the content resources, i.e. any form of scientific output, including scientific/technical reports, working papers, pre-prints, articles and original research data. The vision, to be

accomplished in a second phase, is to establish the successful interoperation of both data network and knowledge repositories as integral parts of the E-infrastructure for research and education in Europe.

DRIVER meets the three key strategic objectives of the EC programme for research infrastructures: a) it optimises the use of the technical infrastructure GEANT by delivering all types of content resources, b) it contributes to the creation of a new Europe wide infrastructure for knowledge and c) it aggregates and presents the knowledge base of European research to the world. The knowledge infrastructure testbed, delivered by DRIVER, will be based on nationally organised digital repository infrastructures, similar to GEANT2 and the NREN's. The successful DARE network in the Netherlands, recently presented to the public by the project partner SURF, will serve as model to DRIVER.

DRIVER with its testbed will not build a specific digital repository system with pre-defined services, based on a specific technology and serving dedicated communities. The testbed will in its inception focus on the infrastructure aspect, i.e., open, clearly defined interfaces to the content network, which allow any qualified service-provider to build services on top of it. Like the data network GEANT, DRIVER's knowledge infrastructure offers mainly a well structured, reliable and trustworthy basis. DRIVER opens up knowledge to the communities, it does not prescribe how to use the knowledge.

3.5 Nationally Applied (Inter)national Standards in Europe

Recently, the Multilingualism and thesaurus Subgroup[56] of MINERVA Plus conducted a major survey of the situation concerning language usage in cultural websites. The aim of the survey was to see to what extent cultural websites and portals are available for users of different language communities and also whether websites use more languages than the language they were originally created in.

Furthermore the survey investigated whether cultural websites are using retrieval tools such as controlled vocabularies or thesauri and whether multilingual tools are available for use. In particular, this second inventory provides relevant input for this deliverable. Below, the usage of controlled vocabularies[57], applicable in the cultural heritage domain, is summarized.

Australia	Generally **Dublin Core** is well understood and used. In government the **AGLS** standard (AS 5044; an extension of the DC standard) is used to describe resources. For collection descriptions the **EAD** and **DC-Collection** are used in government archives. Libraries are using MARC as a metadata schema. The controlled vocabularies generally in use are at Libraries are: • Library of Congress subject headings (**LCSH**), • Getty **AAT** • Getty Thesaurus of Geographic Names (**TGN**) • Australian Pictorial Thesaurus[58] (**APT** Australian Historic Records Register Thesaurus. Smaller historical institutions tend to use the Historical collections classification scheme for small museums by Patricia Summerfield.[59]
Austria	The (ORF) uses a proprietary metadata model to describe video clips, movies and photographs. Proprietary controlled vocabularies, in German, are in use for the following metadata: classification (content related keywords), programme type (genres) and geography (countries only).
Belgium	The library of the Flemish Institute for tangible cultural heritage considers the application of Dublin Core. Both library and archive use an adapted version of the Art and Architecture Thesaurus (**AAT**).
Czech Republic	No multilingual thesauri in use. The institutions of the cultural heritage domain use **LCSH** and **Unesco** thesaurus, neither translated in Czech.
Estonia	No multilingual thesauri in use.
Finland	Finland is currently in the phase of building a national set of ontologies (project web site http://www.seco.tkk.fi/). The National Library of Finland maintains two

[56] http://www.minervaeurope.org/structure/workinggroups/inventor/multilingualism.htm

[57] Registered thesauri on the survey's website http://www.mek.oszk.hu/minerva/survey/contr_vocs2.htm.

[58] http://www.picturethesaurus.gov.au/),

[59] http://www.nla.gov.au/pict/survey1994/parta.html

	different thesauri, which are both also linked to / available in Swedish. Generally used in Finland is the Finnish General Thesaurus is called **YSA.**[60] The corresponding thesaurus translated in Swedish is called Allärs. Finnish Music Thesaurus (**MUSA**) has also a Swedish translation (CILLA). In the museums generally used thesauri are YSA and **MASA**.[61] The project Museum Finland (MuseoSuomi - http://www.museosuomi.fi) provides unified access to three different Finnish museum collections using semantic web technology.
France	Multilingual controlled vocabularies are still scarce. **HEREIN** t**hesaurus** on architecture and archaeology (more than 500 terms in seven languages (English, French, German, Spanish, Bulgarian, Polish and Slovenian) but eleven other languages will soon be available. This "first multilingual thesaurus in the cultural field at an international level " according to the Council of Europe is now available online[62] and is developed by the European Heritage Network (HEREIN). Several other monolingual thesauri in use: on architecture (**Thésaurus de l'architecture**), on archaeology and antiquity (**the "PACTOLS" thesauri:** PACTOLS is the acronym for "Peoples and cultures, Anthroponyms, Chronology, Toponyms, Works, Places, Subjects") and on religious objects. A multilingual database to manage museum laboratory documentation relating to painting materials built for the European **NARCISSE** project (Network of Art Research Computer Image SystemS) in the late 1980s. Partly still monolingual (French), about restoration and conservation. Art works and museum objects: **Museum images vocabularies** (covers art, architecture, sciences, technology, and history; available in five languages: English, German, Italian, French, Spanish). Post-medieval manuscripts and letters: the **Malvine vocabulary** (allows semantic interoperability and is available in five languages: German, English, French, Spanish, Portuguese). Culture field: the **UNESCO thesaurus** (English, French, Spanish). Libraries: the **MACS project** (Multilingual Access to Subjects). The MACS project aims at providing a multilingual access to subjects in the catalogues of the participants (Die Deutsche Bibliothek (SchagWortnormDatei), The British Library (Library of Congress Subject Headings), the Bibliothèque nationale de France (Répertoire d'Autorité-Matière Encyclopédique et Alphabétique Unifié), and the Swiss National Library).

[60] http://www.lib.helsinki.fi/english/libraries/thesauri/index.htm, http://vesa.lib.helsinki.fi/

[61] Special thesaurus for Museums http://www.nba.fi/fi/masaetusivu.[61]

[62] http://www.europeanheritage.net/sdx/herein/thesaurus/introduction.xsp

Germany	Three widely available and electronic Authority lists exist for cataloguing in German libraries: the Schlagwortnormdatei **SWD** (German Subject Headings Authority; also used by some museums), the Gemeinsame Körperschaftsdatei **GKD** (German Corporate Headings Authority) and the Personennamendatei **PND** (German Name Authority; linked to the virtual international authority list VIAF). Museums and archives: many different and individual solutions were created, mainly monolingual. Art museums: large number uses **Iconclass** in German. Only a few use the Getty vocabularies (**TGN, ULAN** and **AAT**).The project "Common Internet Portal for Libraries, Archives and Museums (BAM portal)", partner of MINERVA, uses a subset of Dublin Core for metadata-mapping.
Greece	Scarce use of controlled vocabularies, mainly proprietary. 3% multilingual thesauri, translations of international standards: **LCSH** and **Sears**.
Hungary	Two monolingual (**OSZK** Thesaurus, **WebKat** Thesaurus) are used. Two bilingual (Library of Congress Subject Headings List **LCSH**, Thesaurus of Library Information Science **LIS**) in English and Hungarian. The Hungarian Educational Thesaurus is available in French, English, and German. The UNESCO International Thesaurus of Cultural Development is available in Hungarian, but it has never been used. There is only one thesaurus for museums, but it has never been used.
Ireland	A variety of monolingual controlled vocabularies and thesauri. None multilingual.
Italy	The Central Institute for Catalogue and Documentation (**ICCD**) of the Italian Ministry for Cultural Heritage and Activities produces several mono- or multilingual controlled vocabularies for cataloguing purposes (national standards). The domains covered are: architecture, art-history, archaeological objects and sites, artistic objects, architectural areas. The **ICCD** presents 8 controlled vocabularies related to description of cultural areas, authors, artistic technique and artistic objects. **Artistic objects** (one of the most used) is available in Italian, English, German, French and Portuguese (with specific sections in other languages). The **architectural areas vocabulary** is available in Italian, English and French. The **Multilingual Thesaurus of Religious Objects**, which is available in English, French, and Italian, is also produced by the ICCD in cooperation with the Canadian Heritage Information Network (CHIN), the Getty Information Institute, and the French Ministry of Culture. Another important tool for multilingual classification for the iconography of western art, **ICONCLASS**, is available in Italian, English, German, French, and Finnish. **ThIST (Italian Thesaurus of Earth Sciences),** available in Italian and English, covers the earth science domain. An Italian to English **iconographic thesaurus** (translated into Spanish, German, and French)**,** is maintained by Alinari in cooperation with the University of Florence: about 8,000 entries organised in 61 classes ; includes a geographic thesaurus, thesauri for Periods and Styles, controlled lists for Events, People, Authors (artists) and Photographers. The Alinari thesaurus is a work in progress.

Latvia	Museums use the Dublin Core metadata schema as well as local developed classification schemes in Latvian and the Art & Architecture Thesaurus (**AAT**) and **MDA Spectrum** in English. Archives use the **UKCAT** thesaurus in English and the standards ISAD(G) and ISAAR(CPF). Libraries use four principal vocabulary tools: **UDC** classification in English (this is being translated into Latvian), **MeSH** in English and Latvian (part translation), **LCSH** is used as the basis for developing a partly adapted translation in Latvian, **AGROVOC** in English. Libraries use standards ISBD, AACR2, MARC, FRBR and Dublin Core.
The Netherlands	Dutch heritage community: mostly usage of **AAT-NL**, Ethnographical thesaurus, **RKDartists**[63] (200,000 names and details of artists) and **IconClass**[64]. The Dutch national heritage organisation, Rijksdienst voor Archaeologie, Cultuurlandschap en Monumenten (RACM), is responsible for the preservation and permanent development of archaeological values, monumental buildings and man-made landscape.[65] Their library as well as their archive will be using the Dutch translation of the **AAT** thesaurus for subject indexing. Another controlled vocabulary they use is the Name list of Dutch municipalities to indicate the city in which the object is located. The standard Geographic Markup Language (**GML**) is a new standard (based on XML), since 2005, which makes a previous exchange format (NEN 1878) redundant. In The Netherlands there is a recent standard for the description of geographic objects (i.e. buildings etcetera that can be located on the earth with coordinates like for example degrees of latitude and degrees of longitude): NEN 3610: 2005. This standard dictates the basis elements. On top of that several models for knowledge representation are developed or being developed for different sectors. For example, IMWA for water related geographic objects, IMRO for spatial planning, and IMKICH for the tangible cultural heritage. The Dutch national heritage organisation is setting up a new monument registry or database. The use of open architecture and of GML to communicate between the separate components of the information system guarantee interoperability with other information systems of e.g. the ministries. This new information system will apply **ISO 19115** as well as much specifications and standards of the **OGC** as possible The National Library of the Netherlands uses various vocabularies, including: • GOO: Joint Subject Indexing by scientific libraries (http://www.kb.nl/bst/goo/goointroen.html) • Iconclass: used for annotation and facetted access to ‘illuminated

[63] http://RKD.nl/rkddb/

[64] www.iconclass.nl

[65] http://www.racm.nl/content-en/rubriek-s3.asp?toc=s3

	manuscripts' on the KB website RKD (Netherlands Institute for Art History) has a long history working on thesauri and making them available. Recently they took responsibility for the IconClass thesaurus (www.iconclass.nl). This thesaurus currently represented in ASCII text, an XML version created by the Adlib company also exists.
Poland	No controlled vocabularies for the MultiMatch domain.
Russian Federation	**AAT** (translation in Russian of a part is in development); the **iconography thesaurus** (in English, French and Russian; controlled by the Ministry for Culture of France)
Slovak Republic	No multilingual thesauri in use. Libraries: UDC and monolingual thesauri; MARC21. Museums and galleries: local monolingual controlled vocabularies.
Slovenia	No controlled vocabularies found.
Sweden	The most used vocabularies are the ones that come with the **SOFIE museum** database system (used by approx 300 CH organisations in Sweden). Furthermore there is a mix of local vocabularies and subject specific vocabularies. The plan is to merge data from different Swedish sources into a technical infrastructure in order to create a more controlled set of vocabularies. The KMM project (Knowledge Management Systems in Museums) funded by the Swedish Council for Cultural Affairs the started in July 2005. The project is a long term national R&D initiative concerning "Knowledge Engineering in Museums", focusing on development, research, demonstration, implementation and evaluation of a Swedish state of the art Testbed and R&D platform based on the **CIDOC CRM**. The KMM project is the Swedish partner in the MICHAELplus project. In the Michael project vocabularies will be merged, standards etc used in that project into the infrastructure (ISO, RSLP, DCMI, UNESCO, MINERVA and MICHAEL specific terms). KMM is using the CIDOC CRM as a base model for almost everything inside the testbed combined from now on with the Michael Project Data Model for collections and organisations. The only list that might be considered as a standard is **OCM** (Outline of Cultural Materials, from Human Relations Area Files, New Haven[66] A Nordic multilingual version is available. For geographic purposes almost everyone uses either RAÄs (The National heritage Board, www.raa.se) Sockenkod" or/and the List of Counties of TheNordic Museum or/and data from Statistics Sweden (www.scb.se). Currently (fall 2006), data from the museums in the project and the "Nordic Outline" is merged into NameMaster and ClassMaster. The main vocabularies in ClassMaster will cover subject, technical terms, material and Outline plus any vocabulary that is used in the project and all other Swedish vocabularies that we can get hold on. In the GeoMaster we are merging data from the museums, the National Archives, the National Heritage Board and Statistics Sweden.

[66] http://www.yale.edu/hraf/

United Kingdom	Multilingual: **ARENA periods** (a simple vocabulary list in English, Danish, Norwegian, Icelandic, Polish and Romanian) plus **ARENA top level themes** (a simple vocabulary list covering the cultural heritage and sites and monuments and available in English, Danish, Norwegian, Icelandic, Polish and Romanian). Monolingual thesauri and terminology lists were registered by English Heritage, the Tate and by the Scottish Library and Information Council. The JISC IE Metadata Schema Registry (IEMSR)[67] project is funded by JISC through its Shared Services Programme. The IEMSR project is developing a metadata schema registry as a pilot shared service within the JISC Information Environment. Metadata schema registries enable the publication, navigation and sharing of information about metadata. The IEMSR will act as the primary source for authoritative information about metadata schemas recommended by the JISC IE Standards framework. Metadata within the JISC IE is based largely on two key standards: the **Dublin Core Metadata Element Set** (DCMES) and the **IEEE Learning Object Metadata** (LOM) standard. JISC also funded the project "**Metadata Generation for Resource Discovery**"[68] The first broad aim is to identify the metadata needs of the JISC Information Environment. It is axiomatic that resource-discovery metadata generation tools should be tailored to the resource profile of that domain and the needs of its stakeholders (particularly end users). The second broad aim is to identify and evaluate the metadata generation and creation processes that are currently used within the JISC IE, and particularly the Portals programmes. The third broad aim is to identify currently available tools that are not being used by the JISC portals and discover the reasons for the lack of uptake. The fifth broad aim is to identify the most promising new techniques and approaches emerging from recent experimental research into automated metadata generation.

[67] http://www.ukoln.ac.uk/projects/iemsr/

[68] http://ahds.ac.uk/about/projects/metadata-generation/Metadata_Generation_Project_Plan_Final2%5B1%5D.pdf

3.5.1. Conclusion: related project and current practice in Europe

The most prolific EC projects (MICHAEL plus, The European Library) playing a role in shaping the vision of i2010 'Digital Libraries Initiative'[69] are using Dublin Core. Research regarding (semantic) interoperability is taking place in various places; i.e. BRICKS, DELOS NoE and Knowledge Web are the most prominent initiatives in this respect. MultiMatch will follow the research conducted in the project closely.

With respect to the use of controlled vocabularies; from the Minerva survey, as well as from the case descriptions from the MultiMatch partners, it is clear that there are still a lot of proprietary metadata schemas and local or national controlled vocabularies in use. The latter if only because of the fact that international controlled vocabularies are still not available in every European language (currently there are 20 official languages in the European Union). Although it seems though, that AAT as well as LCSH are widely used.

The Subgroup of Minerva selected some thesauri which are available in more than two languages, and have already been used in many European countries and described them in detail, pages 84-90:

- The UNESCO thesaurus
- Library of Congress Subject Headings (LCSH)
- The HEREIN thesaurus
- The NARCISSE vocabulary and the EROS project
- ICONCLASS (in the field of iconographic description).

Their final recommendation was: " So our suggestion within European context would be, instead of supporting the creation of brand new thesauri, it would be more useful supporting the translations of the well-tried, European wide used thesauri: like UNESCO, HEREIN, ICONCLASS, Library of Congress Subject Heading List on the European Commission level."

[69] http://europa.eu.int/information_society/activities/digital_libraries/index_en.htm

4 Generic Knowledge Representations

Next to the domain-specific standards listed in Chapter 2, more 'generic' standards are also available. These are used in various domains. In the light of the heterogeneous content sources MultiMatch is covering; these generic standards will be studied in much detail.

The following types are distinguished:

- Generic identification standard
- Reference models
- Generic metadata schemas.

The first section presents a description of the identification standard Digital Object Identifier together with a description of the reference models, CIDOC CRM and Functional Requirements for Bibliographic Records (FRBR). These reference models can be viewed as conceptual data models, in other words as maps of concepts or entities and their relationships that describe in an abstract way how data are represented in an information system. Section 4.1 concludes with a description of the generic representation standards SKOS and RDF.

The second section describes three generic metadata schemas, namely Dublin Core (DCMI), MPEG-7 and MPEG-21. Chapter 4 concludes with a separate section on the relationship between the goals of the Semantic Web and MultiMatch.

4.1 Generic Identification Standards, Reference Models and Representation Languages

Digital Object Identifier

Name	Digital Object Identifier
Acronym	DOI
Status / version	NISO standard, proposed for ISO standard
Type	ANSI/NISO Z39.84-2000
Management	The International DOI Foundation (IDF).
Short description	A unique identification mechanism for content in all media. It provides a way to link users of the materials to the rights holders or their agents to facilitate automated digital commerce. Cross platform network identification. A DOI can be used to identify any resource involved in an intellectual property transaction.
Number of elements	Not applicable
Syntaxes	The DOI is composed of the *prefix* and the *suffix*. Within the prefix are the: • Directory Code <DIR> • and the Registrant Code <REG>. • The suffix is made up of the DOI Suffix String <DSS>. The syntax of the DOI string is: <DIR>. <REG> /<DSS> There is no limit on the length of a DOI string, or any of its components. A DOI prefix (for example, 10.1000/) enables a registrant to assign many DOIs, by building on the prefix to construct a range of unique identifiers (10.1000/abc, etc).

Vocabularies proposed	Not applicable
Applied by the following organizations e.g.	Several hundred different registrant organizations have so far allocated several million DOIs. Because the origins of the DOI were in the text sector, an initial large implementation covering half of these registrants was from traditional print-publishing companies that have already established major online publishing programs. However the fundamental design of the system is applicable to any media or content. The IDF is working closely with many businesses in other sectors of the "content industries" to extend the application of the DOI to many other types of intellectual property.
URL(s) documentation	The DOI Handbook (Version 4.3.0, released May 2006) is the primary source of information about the DOI. It discusses the components and operation of the DOI system, and provides a central point of reference for technical information.[70]
URL guidelines for application	The Appendices provide detailed technical documentation and supporting information which is not necessary for most users. http://www.doi.org/handbook_2000/app1
XML encoding available	Not applicable

CIDOC Conceptual Reference Model

Name	CIDOC Conceptual Reference Model
Acronym	CIDOC CRM
Status / version	Version 4.2
Type	Model recently became a standard: ISO/PRF 21127 in May 2006
Management	International Council of Museums (ICOM)
Short description	The CIDOC Conceptual Reference Model is an ontology for cultural heritage information. It describes, in a formal language, the implicit and explicit concepts and relations used in cultural heritage documentation. The model is specifically meant to integrate and exchange heterogeneous sources of information on cultural heritage in the context of the Semantic Web. In other words, CIDOC CRM is a basis for data exchange and for building integrated query tools. The CIDOC CRM is intended to promote a shared understanding of cultural heritage information by providing a common and extensible semantic framework to which any cultural heritage information can be mapped. It is intended to be a common language for domain experts and implementers to formulate requirements for information systems and to serve as a guide for good practice of conceptual modelling. In this way, it can provide the "semantic glue" needed to mediate between different sources of cultural heritage information, such as that published by museums, libraries and archives. The CRM is thought to be primarily a tool for the museum community, however it also enables an effective communication with the libraries and archives world. It is also applicable in the sub-domains archaeology and the preservation of monuments and historic buildings. "The CRM can be regarded as a model of history in the physical sense, as perceived by humans. As such, it contains very abstract concepts."[71]
Number of elements	An ontology of 80 classes and 132 properties for culture and more. Key concepts: Actor, Actor Appellation, Event, Physical thing, Appellation, Conceptual

[70] http://www.doi.org/hb.html Available at 2006-06-19

[71] http://delos-noe.iei.pi.cnr.it/activities/standardizationforum/ontology/ontology.html

	Object, Place, Place Appellation, Type, Time-span, Time-span Appellation.
Extra information on application	CRM maps to Dublin Core. Development: work on extension covering FRBR, FRAR and CRM.[72]
Applied by the following organizations e.g.	A detailed list of references to the CIDOC CRM can be found on the References page. A selection of references: • The design of the database employed by RLG in its Cultural Materials Initiative is based on the CIDOC CRM. • The European IST Project SCULPTEUR. The vision of SCULPTEUR is to develop both the technology and the expertise to help create, manipulate, manage and present these cultural archives, and make available cultural heritage to European people and the world. It employs the CIDOC CRM to develop a sophisticated semantic layer for distributed multimedia information management and a knowledge structure linking low and high-level multimedia representations. In other words, information is integrated through mapping to a common ontology: CIDOC CRM • The KMM project (Knowledge Management Systems in Museums) funded by the Swedish Council for Cultural Affairs started in July 2005. The project is a long term national R&D initiative concerning "Knowledge Engineering in Museums", focusing on development, research, demonstration, implementation and evaluation of a Swedish state of the art Testbed and R&D platform based on the CIDOC Conceptual Reference Model. (...)The KMM project is the Swedish partner in the MICHAELplus project. (...) We are using the CIDOC CRM as a base model for almost everything inside the testbed combined from now on with the Michael Project Data Model for collections and organisations. • Ec(h)o is an "augmented reality interface" utilizing spatialized soundscapes and a semantic web approach to knowledge. It uses the CIDOC CRM model to describe museum artefacts. • The European IST Project I-Mass. The basis for knowledge representation I-Mass is the ontology or ontological framework that is used to shape the knowledge landscape. Such an ontology models the domain of discourse and, as such, determines in a Wittgensteinian fashion what may [not] be spoken about. This implies that the ontology must in principle be able to cover the whole cultural domain, and not be restricted to either the so-called "high-culture" of the elite or to the western view on culture. I-Mass is based on the CIDOC-Conceptual Reference Model (CIDOC-CRM) for several 'practical' requirements, such as allowing for machine interpretation and minimising semantic completeness. References: Geert de Haan, Design for Global Access to Cultural Heritage • CIMEC, the Romanian Institute for Cultural Memory, is developing a data model based on FRBR and CRM. • the National Library of France also plays an active role in the International Working Group on FRBR/CIDOC CRM Harmonisation. Applications system & schema design based on CIDOC CRM: • RLG Cultural Materials • Finnish National Gallery Database • City of Geneva MusInfo Project • Germanische Nationalmuseum Nuremberg • Monument Inventory Data Standard

[72] The CIDOC CRM, a Standard for the Integration of Cultural Information / Martin Doerr (Institute of Computer Science Foundation for Research and Technology – Hellas) Presentation. Nurnberg, 14-15 November 2005.

	• Heritage Data Dictionary • CLIO Cultural Documentation System, ICS-FORTH/Benaki Museum.
URL(s) documentation	http://cidoc.ics.forth.gr/official_release_cidoc.html Viewed 2006-10-19
URL guidelines for application	International Guidelines for Museum Object Information: The CIDOC Information Categories http://www.willpowerinfo.myby.co.uk/cidoc/guide/ Available at 2006-06-19. A description of the Information Categories that can be used when developing records about the objects in museum collections:
XML encoding available	No, but it can be encoded in XML. Instructions on encoding in OWL or in RDF are available at http://cidoc.ics.forth.gr/official_release_cidoc.html .

Functional Requirements for Bibliographic Records

Name	Functional Requirements for Bibliographic Records
Acronym	FRBR
Status / version	1998
Type	Recommendation for a conceptual model
Management	International Federation of Library Associations and Institutions (IFLA)
Short description	FRBR is a recommendation by IFLA to restructure catalogue databases to reflect the conceptual structure of information resources. More technically, FRBR uses an entity-relationship model of metadata for information objects, instead of the single flat record concept underlying current cataloguing standards. FRBR conceptualizes three groups of entities: *Group 1* consists of the products of intellectual or artistic endeavour (e.g., publications). *Group 2* comprises those entities responsible for intellectual or artistic content (a person or corporate body). *Group 3* includes the entities that serve as subjects of intellectual or artistic endeavour (concept, object, event, and place). The internal subdivision of Group 1 entities is important as well. FRBR specifies that intellectual or artistic products include the following types of entities: **the work**, a distinct intellectual or artistic creation; **the expression**, the intellectual or artistic realization of a work; **the manifestation**, the physical embodiment of an expression of a work; **the item**, a single exemplar of a manifestation.
Number of elements	Not applicable
Applied by the following organizations e.g.	• Many libraries • The European project ECHO designed a metadata schema based on FRBR. • Netherlands Institute of Sound and Vision based their metadata model on FRBR. • OCLC research has resulted in the decision to converse WorldCat to FRBR standards. They use a work set algorithm to cluster related WorldCat records. • Several broadcasting companies System & schema design based on FRBR: • RLG RedLightGreen • Sound and Vision • ARTstor: The Illustrated Bartsch • AustLit Gateway
URL(s) documentation	http://www.ifla.org/VII/s13/frbr/frbr.htm Viewed 2006-10-19

XML encoding available	No

SKOS Simple Knowledge Organisation System

Name	Simple Knowledge Organisation System
Acronym	SKOS Core
Status / version	Draft 2, November 2005 Review proposals every 2-3 months: http://www.w3.org/2004/02/skos/core/proposals
Type	Standard representation language
Management	W3C SWBPD-WG
Short description	SKOS Core provides a model for expressing the basic structure and content of concept schemes (or knowledge organization systems) such as thesauri, classification schemes, subject heading lists, taxonomies, 'folksonomies', other types of controlled vocabulary, and also concept schemes embedded in glossaries and terminologies. The SKOS Core Vocabulary is an application of the Resource Description Framework (RDF) that can be used to express a concept scheme as an RDF graph. Using RDF allows data to be linked to and/or merged with other data, enabling data sources to be distributed across the web, but still be meaningfully composed and integrated. SKOS can be seen as a supplement to OWL Web Ontology Language (the semantic mark-up language for publishing and sharing ontologies on the WWW; http://www.w3.org/2004/OWL/ Viewed 2006-09-27).
URL(s) documentation	http://www.w3.org/TR/2005/WD-swbp-skos-core-spec-20051102/ Viewed 2006-09-27.
URL guidelines for application	http://www.w3.org/TR/2005/WD-swbp-skos-core-guide-20051102/ Viewed 2006-09-27.
XML encoding available	Yes

RDF Resource Description Framework

Name	Resource Description Framework
Acronym	RDF
Status / version	10 February 2004
Type	Standard representation language
Management	W3C
Short description	Graphing theory (i.e. arcs and nodes)-influenced, XML syntax-based metalanguage for expressing metadata about web resources. Designed to convey metadata for machine consumption.
Number of elements	Fundamental building block of RDF is the triple (subject + predicate + object).
Syntaxes	RDF/XML Syntax Specification (Revised)
Interoperability	See Discussion paper.
Extra information on application	• An 'RDF Vocabulary' is a set of RDF 'terms' for describing something in RDF... • E.g. DC for simple meta-properties, FOAF for social networks, OWL for ontologies.[73]

[73] SKOS Core and RDF : presentation at NKOS workshop 2004-09-16 /A.J. Miles http://www.w3.org/2004/02/skos/ Last viewed September 14, 2006.

Applied by the following organizations e.g.	RDF is an enabling technology for a wide variety of projects. Content Authoring, Resource Description, and General Purpose Cooperative Catalogues: • Altova SemanticWorks • Adobe's Extensible Metadata Platform (XMP) • Creative Commons • Friend of a Friend (FOAF) • Dublin Core Metadata Initiative • OCLC Connexion • Open Directory Project • xmlTree • DSpace Syndication, Aggregation, and Rating: • XMLNews-Meta • PRISM: Publishing Requirements for Industry Standard Metadata • Personal Collections: Music, Photos, Calendars, and Contacts : • MusicBrains Metadata Initiative • RDFPic, a tool to embed an RDF description of an image (digitized photograph) into the image itself.
URL(s) documentation	http://www.w3.org/RDF/#gen-col http://planetrdf.com/guide/#sec-apps Viewed 2006-10-19
URL guidelines for application	Quick Guide to Publishing a Thesaurus on the Semantic Web W3C. Alistair Miles ed. "This document describes in brief how to express the content and structure of a thesaurus, and metadata about a thesaurus, in RDF."[74] RDF Vocabulary Description Language 1.0: RDF Schema. W3C. Dan Brickley, R.V. Guha eds. "This specification describes how to use RDF to describe RDF vocabularies" [75] Other useful references include: • MITRE RDF and OWL tutorials by Roger L. Costello and David Jacobs: XML Design (A Gentle Transition from XML to RDF), Inferring and Discovering Relationships using RDF Schemas and OWL Web Ontology Language. 31 March 2003. • RDF Tutorial, Miloslav Nic, Zvon.org - a series of RDF examples with short descriptions. Available from the Zvon download area as zip or tar.gz archives. August 2000. • RDF Tutorial (HTML - PS and PDF also available), Pierre-Antoine Champin, University of Lyon, France, 9 March 2000. • Examples maintained by Andy Powell, UKOLN • Dublin Core Examples in RDF, Eric Miller and Renato Iannella, 6 March 1998 • Resource Description Framework Documents and Examples at DSTC RDU, Australia
XML encoding available	Yes

[74] http://www.w3.org/TR/swbp-thesaurus-pubguide/

[75] http://www.w3.org/TR/rdf-schema/

OAIS Open Archival Information System

Name	Open Archival Information System
Acronym	OAIS
Status / version	January 2002
Type	Standard representation language
Management	Consultative Committee for Space Data Systems (CCSDS)
Short description	The OAIS Reference Model is a useful vocabulary for discussing the preservation of digital objects in a repository context. The purpose is to establish a system for archiving information, both digitalized and physical, with an organizational scheme composed of people who accept the responsibility to preserve information and make it available to a designated community. This reference model addresses a full range of archival information preservation functions including ingest, archival storage, data management, access, and dissemination. It also addresses the migration of digital information to new media and forms, the data models used to represent the information, the role of software in information preservation, and the exchange of digital information among archives. It identifies both internal and external interfaces to the archive functions, and it identifies a number of high-level services at these interfaces. It provides various illustrative examples and some "best practice" recommendations. It defines a minimal set of responsibilities for an archive to be called an OAIS, and it also defines a maximal archive to provide a broad set of useful terms and concepts. The OAIS model may be applicable to any archive. It is specifically applicable to organizations with the responsibility of making information available for the long term. This includes organizations with other responsibilities, such as processing and distribution in response to programmatic needs.. In this reference model there is a particular focus on digital information, both as the primary forms of information held and as supporting information for both digitally and physically archived materials. Therefore, the model accommodates information that is inherently non-digital (e.g., a physical sample), but the modeling and preservation of such information is not addressed in detail.
Extra information on application	A number of follow-on standards were identified in the Reference Model, and work is being pursued for a number of these in the [Consultative Committee for Space Data Sysems] and other organisations: • standard(s) for the submission (ingest) of digital data sources to the archive • standard(s) for accreditation of archives
Applied by the following organizations e.g.	• Dutch Royal Library e-depot • The Integrated Project CASPAR - Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval CASPAR intends to Implement, extend and validate the OAIS Reference Model (ISO:14721:2002). http://www.casparpreserves.eu/ • Examples of Digital repository systems using the OAIS Reference Model include aDORe, DAITSS, DSpace and Fedora
URL(s) documentation	http://public.ccsds.org/publications/archive/650x0b1.pdf Viewed 2006-11-25 http://www.iso.org/ search for ISO 14721:2003
URL guidelines for application	http://www.ukoln.ac.uk/projects/grand-challenge/papers/oaisBriefing.pdf (Briefing Paper: the OAIS Reference Model) http://nost.gsfc.nasa.gov/isoas/ref_model.html http://www.ercim.org/publication/Ercim_News/enw66/giaretta.html

XML encoding available	There are several published metadata schemata which have been designed with the preservation of digital documents in mind, and these have varying degrees of correspondence with the OAIS Information Model. For example, PREMIS and the National Library of New Zealand both use their own data model and metadata structure, but both map fairly well on the OAIS Information Model, with PREMIS being a little more comprehensive.

4.2 Generic Metadata Schemas

Dublin Core Metadata Initiative

Name	Dublin Core Metadata Initiative
Acronym	DCMI
Status / version	Standard / Version 1.1
Type	ISO 15836; ANSI/NISO Z39.85
Management	The Dublin Core Metadata Initiative
Short description	Dublin Core is a metadata schema that describes content and context of a digital work such as video, sound, image, text and composite media like web pages. An implementation of Dublin Core is currently XML and Resource Description Framework based. Dublin Core is one standard for a set of descriptors (such as the title, publisher, subjects, etc.) that are used to catalogue a wide range of networked resources, such as digitized text documents, photographs and audiovisual media. This information about the item, or metadata, is embedded within the electronic item itself, and enables the documents/objects to be found using controlled vocabulary and keyword searching. The Dublin Core standard includes two levels: Simple and Qualified. Simple Dublin Core comprises fifteen elements; Qualified Dublin Core includes three additional elements (Audience, Provenance and RightsHolder), as well as a group of element refinements (also called qualifiers) that refine the semantics of the elements in ways that may be useful in resource discovery.
Number of elements	The Simple Dublin Core Metadata Element Set (DCMES) consists of 15 metadata elements. Each Dublin Core element is optional and may be repeated. The Dublin Core Metadata Initiative (DCMI) has established standard ways to refine elements and encourage the use of encoding and vocabulary schemes. There is no prescribed order in Dublin Core for presenting or using the elements. Subsequent to the specification of the original 15 elements, an ongoing process to develop exemplary terms extending or refining the Dublin Core Metadata Element Set (DCMES) was begun. The additional terms were identified, generally in working groups of the Dublin Core Metadata Initiative, and judged by the DCMI Usage Board to be in conformance with principles of good practice for the qualification of Dublin Core metadata elements. Element refinements make the meaning of an element narrower or more specific. A refined element shares the meaning of the unqualified element, but with a more restricted scope. The guiding principle for the qualification of Dublin Core elements, colloquially known as the "Dumb-Down Principle," states that an application that does not understand a specific element refinement term should be able to ignore the qualifier and treat the metadata value as if it were an unqualified (broader) element. While this may result in some loss of specificity, the remaining element value (without the qualifier) should continue to be generally correct and useful for discovery.

Vocabularies proposed	TGN for the spatial metadata element, describing the spatial characteristics of the intellectual content of the resource.
Extra information on application	"Dublin Core, as its name implies, was designed to support the most basic information needs and not those of specialized knowledge institutions. The element/qualifier set will be extended as information needs change. The Dublin Core community will approve new elements or qualifiers that the general information community would find useful and implementers may register elements and qualifiers developed to support special interests. The restrictions on the number and types of elements and qualifiers designated as Dublin Core is to ensure interoperability between the various applications."[76] Dublin Core is often used within the Open Archives Initiative (OAI) framework. The OAI technical infrastructure, specified in the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) defines a mechanism for data providers to expose their metadata. This protocol mandates that individual archives map their metadata to the Dublin Core, a simple and common metadata set for this purpose. Dublin Core is mapped to EAD. [77] More information on using Dublin Core in combination with Resource Description Framework can be found at http://dublincore.org/documents/dcmes-xml and http://dublincore.org/documents/dcq-rdf-xml Viewed 2006-10-02. There is a specific application of Dublin Core for the cultural domain: DC.Culture. DC.Culture enables the searching of item-level metadata through 4 key access points – Who, What, Where and When. DC.Culture is an adaptation of Dublin Core Simple for cultural metadata. It identifies the CIMI/Aquarelle High-Level Access Points "who", "what", "where" and "when" as a relevant framework for culture metadata. See also: Dublin Core – Education, section 2.6.1

[76] http://www.getty.edu/research/conducting_research/standards/intrometadata/path.html

[77] http://www.getty.edu/research/conducting_research/standards/intrometadata/3_crosswalks/index.html

Applied by the following organizations e.g.	Dublin Core has been adapted by a number of communities to suit their own needs and incorporated into several domain specific metadata schemes.[78] Applications and guidelines: The CIMI Guide to Best Practice for Museums using Dublin Core. http://www.cimi.org/public_docs/meta_bestprac_v1_1_210400.pdf Some examples of Dublin Core in use • GEM (Gateway to Educational Materials)[79] • Open Archives Initiative[80] • Western States Dublin Core Metadata Best Practices[81] • AMICO Media Metadata record is based on the Dublin Core[82] • OCLC usage: Connexion, DCPS, ContentDM, Research. "[83] • The European Library • The MICHAEL project See also http://www.ifla.org/II/metadata.htm for a sampling of projects using Dublin Core (in alfabetical order of country name). Examples of metadata element sets based on Dublin Core: • DC.Culture [84]. DC Culture is an adaptation of Dublin Core Simple for cultural Metadata. It identifies the CIMI/Aquarelle High-Level Access Points "who", "what", "where" and "when" as a relevant framework for culture metadata. • PBCore. Public Broadcasting Metadata Dictionary Project Designed to provide a standard way of describing and using metadata for public broadcasters and associated communities. PB Core is built on the foundation of the Dublin Core and has been reviewed by the Dublin Core Metadata Initiative Usage Board. [MIC]
URL(s) documentation	http://dublincore.org/documents/dces/ Viewed 2006-10-19

[78] http://dublincore.org/projects

[79] http://www.thegateway.org/about/documentation/gem-2-element-set-andprofiles

[80] http://www.openarchives.org/

[81] http://content.lib.utah.edu/cdm4/item_viewer.php?CISOROOT=/docs_regional&CISOPTR=1&REC=1 Framework NISO

[82] http://www.amico.org/library.html

[83] Metadata standards / Eric Childress. Presentation for FEDLINK OCLC Users Group Meeting. November 18th 2003.

[84] http://www.minervaeurope.org/DC.Culture.htm Available 2006-06-14 [Minerva Technological Guidelines]

URL guidelines for application	Dublin Core Abstract Model [http://dublincore.org/documents/abstract-model/] Using Dublin Core / Hillman, Diane. 2005-11-07 [http://dublincore.org/documents/usageguide/] Available 2006-06-14. The guidelines for the notation of Dublin Core in XML format can be found at http://dublincore.org/documents/dc-xml-guidelines . Viewed 2006-10-02. NEW METHODS FOR ENHANCING THE EFFECTIVENESS OF THE DUBLIN CORE METADATA STANDARD USING COMPLEX ENCODING SCHEMES / István Szakadát, László Lois, and Gábor Knapp, 2004.[85] The Dublin Core Abstract Model provides a reference model against which particular DC encoding guidelines can be compared, independent of any particular encoding syntax. Such a reference model allows implementers to gain a better understanding of the kinds of descriptions they are trying to encode and facilitates the development of better mappings and translations between different syntaxes. Although the document is primarily aimed at the developers of software applications that support Dublin Core metadata, anyone who is considering implementing Dublin could usefully review the document. Those involved in developing new syntax encoding guidelines for Dublin Core metadata or developing metadata application profiles based on the Dublin Core should also become familiar with the DC Abstract Model.
XML encoding available	Yes

MPEG-7

Name	Multimedia Content Description Interface
Acronym	MPEG-7
Status / version	International standard, September 2001
Type	ISO/IEC 15938
Management	MPEG (Moving Picture Experts Group)
Short description	MPEG-7 is a multimedia description and indexing system that combines XML-based content description with non-textual indexing of physical features (colour, movement, shape, sound etc.) via processing of the media bit stream for multimedia information – audio, video and images. Part 5 of the standard (ISO/IEC 15938) provides descriptive, technical (for video or audio content) and usage metadata. ISO/IEC addresses the publication and rights metadata elements in different parts of the MPEG-21 standard. This system is used to create a hierarchical model. For example, an audio-visual object can be described by its temporal decomposition and by its media source decomposition. The latter is divided into descriptions about the audio segment and the video segment, which is on its turn decomposed into shots, key frames, and objects. MPEG-7 data can describe AV material (especially: still pictures, graphics, 3D models, audio, speech, video) as well as how these elements are combined in a multimedia presentation ('scenarios', composition information). Special cases of these general data types may include facial expressions and personal characteristics. The standard supports descriptions at the segment level (i.e., shots or clips); supports textual and non-textual data. It can reside native on an MPEG-4 stream. Primarily used for born-digital materials.

[85] http://mokk.bme.hu/archive/dc_isd2004/pdf/data/at_download#search=%22%22Istv%C3%A1n%20Szakad%C3%A1t%2C%20L%C3%A1szl%C3%B3%20Lois%2C%20and%20G%C3%A1bor%20Knapp%22%22

	"MPEG-7 labels for searching: The need for universal mechanisms that allow contents to be quickly and easily catalogued and search tools to access increasingly large amounts of multimedia data has become pressing. This is the need that lies behind the creation of MPEG7, which in addition to normal signal coding operations provides a series of support information called "metadata" that permits a description to be added to audio-visual content to allow searches, selections, and time synchronization. This standard allows intelligent searches for information to be performed and permits two independently developed areas of telecommunications to converge: intelligent agents, i.e. artificial intelligence, and signal" [86]
Number of elements	The MPEG-7 Standard consists e.g. of the following parts: • MPEG-7 Description Definition Language - the language for defining the syntax of the MPEG-7 Description Tools and for defining new Description Schemes. • MPEG-7 Visual – the Description Tools dealing with Visual descriptions. • MPEG-7 Audio – the Description Tools dealing with Audio descriptions. • MPEG-7 Multimedia Description Schemes - the Description Tools dealing with generic features and multimedia descriptions.
Extra information on application	MPEG-7 has official liaisons with SMPTE, P_META and TV-Anytime. Application domain: digital libraries, audiovisual archives, image banks, broadcasting (media selection and distribution) and Web applications for teleshopping and educational purposes. MPEG-7 is designed to take into account all the viewpoints under consideration by other leading standards such as, among others, Dublin Core, SMPTE Metadata Dictionary, and EBU P-META.
Applied by the following organizations e.g.	
URL(s) documentation	http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm Viewed 2006-10-19. MPEG-7: the generic Multimedia Content Description Standard, José M. Martínez, Rob Koenen, and Fernando Pereira
URL guidelines for application	MPEG-7: Overview of MPEG-7 Description Tools, José M. Martínez. MPEG-7 Systems – the tools needed to prepare MPEG-7 descriptions for efficient transport and storage and the terminal architecture. MPEG-7 Conformance Testing - guidelines and procedures for testing conformance of MPEG-7 implementations. MPEG-7 Extraction and use of descriptions – informative material (in the form of a Technical Report) about the extraction and use of some of the Description Tools. MPEG-7 Profiles and levels - provides guidelines and standard profiles. MPEG-7 Schema Definition - specifies the schema using the Description Definition Language The Moving Image Collections (MIC) project has published an application profile with user guide, PowerPoint tutorials, a crosswalk to Dublin Core, and a prototype MPEG-7 cataloguing utility in MS-Access. The IBM alphaWorks development team has released a downloadable MPEG-7 Annotation

[86] http://www.telecomitalia.com/cgi-bin/tiportale/TIPortale/ep/contentView.do?channelId=-12303&LANG=EN&contentId=28721&programId=26849&programPage=%2Fep%2Fprogram%2Feditorial.jsp&tabId=2&pageTypeId=-12157&contentType=EDITORIAL

	Tool (http://www.alphaworks.ibm.com/tech/videoannex)to annotate video sequences with MPEG-7 metadata.
XML encoding available	Yes

MPEG-21

Name	MPEG-21
Acronym	MPEG-21
Status / version	ISO 21000, version 2004 October
Type	International standard
Management	The Moving Picture Experts Group or MPEG (ISO/IEC JTC1/SC29 WG11).
Short description	MPEG describes this standard as a multimedia framework. The goal of this framework, is to enable transparent and augmented use of multimedia resources across a wide range of networks and devices used by different communities. MPEG-21 is based on two essential concepts: the definition of a fundamental unit of distribution and transaction, which is the Digital Item, and the concept of users interacting with them. Digital Items can be considered the kernel of the Multimedia Framework and the users can be considered as who interacts with them inside the Multimedia Framework. At its most basic level, MPEG-21 provides a framework in which one user interacts with another one, and the object of that interaction is a Digital Item. We could thus say that the main objective of the MPEG-21 is to define the technology needed to support users to exchange, access, consume, trade or manipulate Digital Items in an efficient and transparent way. A **Digital Item** comprises the Digital Item Declaration (DID) and the Resources. The **DID** is an XML file describing the Digital Item whereas the Resources are the individually identifiable multimedia Assets of the Digital Item (DI). The DID file may include information such as unique identifiers for the complete DI as well as for Resources, expressions on rights and permissions pertaining to the DI (or parts thereof) and generic metadata describing the Digital Item and its Resources. Finally, the DID contains references to the Resources).Typical examples of Resources include AAC audio files, MPEG-2 video clips, JPEG images, MPEG-4 presentations, HTML pages – but also e.g., video clips or text in proprietary formats. MPEG-21 includes a Rights Expression Language (REL). The MPEG REL data model for a rights expression consists of four basic entities and the relationship among those entities (The principal, to whom the grant is issued; The right that the grant specifies; The resource to which the right in the grant applies; The condition that must be met before the right can be exercised). "MPEG-21: rights in the digital world. At present the MPEG Group is putting the finishing touches to the MPEG21 standard, devised for protecting audio-visual information on various service platforms. This new standard's goal is to define a way to describe, use and exchange usage rights for digital goods, a key technological factor in sustaining any business cycle based on distribution, consultation and

[87] http://www.telecomitalia.com/cgi-bin/tiportale/TIPortale/ep/contentView.do?channelId=-12303&LANG=EN&contentId=28721&programId=26849&programPage=%2Fep%2Fprogram%2Feditorial.jsp&tabId=2&pageTypeId=-12157&contentType=EDITORIAL

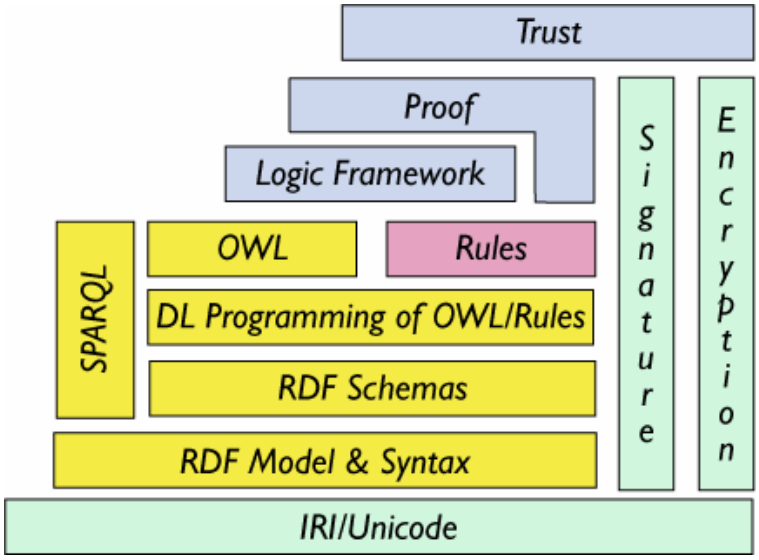
	consumption of content." [87]
Number of elements	Attributes are organised in the following groups: Reference software, File format
Extra information on application	An MPEG-21 Digital Item can be a complex collection of information. Both still and dynamic media (e.g. images and movies) can be included, as well as Digital Item information, meta-data, layout information, and so on. It can include both textual data (e.g. XML) and binary data (e.g. an MPEG-4 presentation or a still picture). For this reason, the MPEG-21 file format will inherit several concepts from MP4, in order to make 'multi-purpose' files possible. A dual-purpose MP4 and MP21 file, for example, would play just the MPEG-4 data on an MP4 player, and would play the MPEG-21 data on an MP21 player.
Applied by the following organizations e.g.	Koninklijke Bibliotheek (National Library of the Netherlands) "In our metadata the encoding scheme for the identifier for compund objects is "mpeg21" indicating that additional parameters may be used for requesting individual components or the mpeg21 record itself. Without additional parameters a default presentation for the specific type of object will be applied."
URL(s) documentation	http://en.wikipedia.org/wiki/MPEG-21 Viewed 2006-09-26. http://mpeg-21.itec.uni-klu.ac.at/cocoon/mpeg21/_mpeg21Parts.html Viewed 2006-09-26. http://xml.coverpages.org/MPEG21-WG-11-N3971-200103.pdf Viewed 2006-12-02. MPEG-21 Digital Item Declaration WD (v2.0)
URL guidelines for application	http://mpeg-21.itec.uni-klu.ac.at/cocoon/mpeg21/_mpeg21UseCase.html Viewed 2006-09-26.
XML encoding available	Yes

4.3 Semantic Web Technologies Within the MultiMatch Project

4.3.1 The Semantic Web

The Semantic Web (SW) aims to enable documents to contain computer-readable meaning (semantics) with the goal that documents should become computer-understandable. This is achieved via the interaction of a number of complimentary markup languages and processing tools.

Currently documents generally contain markup which facilitates the presentation of the contents in a human-readable form, i.e. font-types, positional information, etc. It is possible to infer some of the meaning behind the content, for example the summary of a document's contents can be assumed to be given by its title or in a section headed summary or abstract, given a table with headers labelled item and price, the rows can be assumed to provide the relative price for each specified item. Web scraping attempts to make use of such regularities in HTML documents to extract such information; with varying degrees of success. However the SW aims to make the meaning behind the contents of a page explicit. Thus an item would be represented in a standardised form (i.e. an item number) which might be linked to a repository given further description and specification and each item would be linked to a given price (i.e. a numerical representation in Euro).



Within the SW, the contents of a document are marked-up using the Extensible Markup Language (XML) which provides the syntax for structuring the contents, with XML Schemas providing restrictions for the structuring of this syntax. This identifies the entities (resources) within a document, but does not impose any semantics on those entities. A Resource Description Framework (RDF) model is used to provide a data model describing how the entities (resources) are related; in RDF each resource is described by a URI. RDF Schema (RDFS) is a vocabulary for describing groups of related resources and the relationships between them. RDFS uses resources to determine characteristics of other resources, such as the domains and ranges of properties. RDFS (or the more expressive Web Ontology Language (OWL)) is used to define an ontology which provides a conceptualisation of a given domain.[88]

4.3.2 Application of the Semantic Web in MultiMatch

As part of the MultiMatch, documents, within the Cultural Heritage domain, will be marked-up with semantic information (or metadata) from a common vocabulary. One criticism levelled at the SW is the cost associated with providing this markup; the project will examine the use of classification and information extraction techniques to alleviate this problem. The SW is also concerned with the interoperability between different vocabularies (and ontologies); an issue which will also have to be addressed within MultiMatch. There are also other issues which relate to the SW, such as "trust" and the provenance of information, privacy and censorship and the provision of Web services which, whilst not central, will be examined in the project.

[88] Image taken from: http://www.w3.org/Consortium/Offices/Presentations/RDFTutorial/figures/TwoTowers.png

Whilst there is no specific aim to direct the results of the MultiMatch project towards providing material for the SW, there is an obvious relationship between the goals of MultiMatch and the SW. Much of the technology examined in MultiMatch will consider issues relevant to the development of the SW. Thus the project should both add to and benefit from SW technologies and research, and provide tools and materials which will be exploitable in the context of the SW.

5 Summary and Further Research

Chapter 2 lists forty knowledge representation standards and documents their usage. In Chapter 4, a further nine generically applicable standards were added. This final chapter summarises the most relevant standard(s) for each sub-sector, taking into consideration both their current use and the aims of the MultiMatch project. The generic schemas (Dublin Core, MPEG-7) and reference models (FRBR, CIDOC-CRM) are also analysed.

In order to provide an overview of the metadata schemas, the analysis methodology from De Sutter (et. al.) is used. This methodology distinguishes four criteria to categorize metadata schema:

1. Internal vs. Exchange Metadata Model
2. Flat vs. Hierarchical Metadata Model Criterion
3. Supported Types of Metadata
4. Syntax and Semantics

The objective of this deliverable has been to perform an in-depth investigation of the metadata schemas and semantic mark-up formalisms adopted in the cultural heritage domain in order to have a clear view of the current state-of-the-art in this sector and to provide informative documentation on which to base the decisions that will be taken with respect to the approach that will be adopted in MultiMatch for knowledge representation. The final decision will be motivated and described in detail in D2.2. Here below we provide an initial analysis.

5.1 Metadata in the Cultural Heritage domain and MultiMatch

This paragraph summarises the most relevant standard(s) for each sub-domain of the cultural heritage domain, together with the most relevant generic schemas (Dublin Core, MPEG-7, MPEG-21) and a first analysis of the reference models FRBR and CIDOC-CRM. In advance of the user and other requirements for MultiMatch, yet to be published, a preliminary indication is given of the possible usability of the most important metadata schemes and controlled vocabularies within MultiMatch. As multilinguality is a major issue of the MultiMatch project, this factor is considered in the preliminary indication of the usability of the controlled vocabularies.

5.1.1 Archives

The archive community has developed the Encoded Archival Description (EAD) and General International Standard Archival Description (ISAD(G)) standards to provide for the administration and discovery of archival records. Both **EAD** and **ISAD(G**) are broadly accepted metadata schemas: EAD for describing finding aids to search archives and ISAD(G) for describing documents in archives. ISAD(G) contains useful metadata elements for the MultiMatch metadata schema.

Of the controlled vocabularies used by archives and selected for this deliverable, the **IPTC thesaurus** is the most interesting for subject indexing in general, as it is multilingual. Although the IPTC thesaurus is constructed to index news douments, it could be applied for the rough classification or subject indexing of cultural heritage objects. The International Standard Archival Authority Record for Corporate Bodies, Persons and Families, **ISAAR (CPF),** could also play a role in the semantic background information for the MultiMatch database, especially if one of the partners applies this standard. The adoption of the ISAAR (CPF) standards with its four information areas on Corporate Bodies, Persons and Families seem rather elaborate for application within the MultiMatch metadata schema/database.

The **Thésaurus architecture et patrimoine** is monolingual, and covers a restricted subject area within cultural heritage. When or if this subject area – architecture, historical buildings and furniture – proves important in the MultiMatch database, this thesaurus will be relevant for semantic background information.

The same can be said of the **UK Archival Thesaurus**. If archival cultural heritage objects form an important part of the cultural heritage objects indexed by the MultiMatch database, then this monolingual thesaurus could be helpful to provide semantic background information. However, our first impression is that the UNESCO thesaurus itself might be more helpful.

5.1.2 Libraries

Both the Functional Requirements for Bibliographic Records (FRBR) and Machine Readable Cataloguing (MARC) are widely accepted throughout the world.

The **FRBR** conceptual model is important for the MultiMatch application, as it provides a useful view on relationships between the entities concerning the cultural heritage objects described. This view can help to construct the hierarchical structure that is probably needed in the MultiMatch metadata schema. A structured data model, with clear relationships between and among entities, provides the user with means to navigate through related cultural heritage objects.

Although the **MARC** standard is widely adopted by library communities, and is a well-maintained and mature standard, we do not feel that it is a suitable candidate as the metadata schema for MultiMatch.

This is not only because the standard is virtually unused outside the library domain and the future of the MARC formats is a matter of some debate in the worldwide library science community, but also due to:

- the bibliographical focus of this standard,
- the flatness of the metadata model and
- the limited ability to convey complex relationships, hierarchy, attributes at tag/subfield level [89].

However, the associated Metadata Object Description Schema (**MODS)** recommendation could prove useful for the technical partners of MultiMatch, because of the focus on metadata exchange (the schema is well-suited for OAI-harvesting).

The Metadata Encoding and Transmission Language **(METS)** is a standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library, and could prove useful for application in MultiMatch depending on the technical solution that will be chosen for the exchange and transmission of metadata describing cultural heritage objects. Further study is needed to establish whether the essential metadata included in this encoding standard represent a sufficiently robust metadata schema for MultiMatch.

The **controlled vocabularies** most widely used in the library domain are two multilingual classification systems, DDC and UDC. As both provide extensive classification systems covering all subject areas, they are likely to prove useful as navigation tool for MultiMatch and in providing semantic background information.

The term list Library of Congress Subject Headings (LCSH) is also widely used and multilingual, the related French controlled vocabulary is RAMEAU. The usability of these knowledge organization systems for MultiMatch remains to be seen. It is largely dependent on the decisions on the design of the semantic background information.

[89] Metadata standards / Eric Childress. Presentation for FEDLINK OCLC Users Group Meeting. November 18th 2003.

It is likely that the monolingual controlled vocabularies, LCAF and LCC, will not be of great use for MultiMatch. Similarly, the conceptual model Functional Requirements on Authority Records (FRAR), with guidelines and rule for setting up authority files, does not seem very useful for MultiMatch.

5.1.3 Museums

The three most used standards in the museum domain are: Categories for the Description of Works of Art (CDWA), Visual Resources Association Core Categories (VRA) and Object ID.

- **CDWA**, like the British SPECTRUM standard, support the management of museum objects. The CDWA also includes data elements for visual surrogates; while VRA focuses on the surrogate, CDWA provides much richer, more detailed information for the original work.
- **VRA** Core is comprised of elements that are designed to facilitate the sharing of information among visual resources collections about works and their visual representations. CIMI is also a detailed metadata schema in use in the museum domain.
- **Object ID** is an international standard, developed from a subset of the CDWA that codifies the minimum set of data elements needed to protect or recover an object from theft.

All of these three standards can influence in the choice or the development of the metadata schema for MultiMatch.

The controlled vocabularies widely in use in this domain, especially the Art & Architecture Thesaurus (**AAT**), the Union List of Artist Names (**ULAN**) and the Getty Thesaurus of Geographic Names (**TGN**), all Getty vocabularies, are likely to play a useful role in the metadata schema of MultiMatch as well as in the semantic background information. As the TGN includes coordinates, this vocabulary is particularly likely to be useful in offering geographical search functionality in MultiMatch.

Many controlled vocabularies in this domain are monolingual, thus the two multilingual vocabularies; the UNESCO thesaurus (English, French, Russian and Spanish) and the AAT (available in English and Dutch), are likely to be the most interesting to MultiMatch.

5.1.4 Educational sector

This sector seems to be using mainly: Learning Object Metadata (IEEE LOM), IMS and qualified DC metadata. Of those, the **IEEE LOM** standard is the most relevant metadata commonly in use in the Educational sector.

Learning Object Metadata is used to describe educational resources in course management systems and learning management systems. The LOM has been incorporated into a number of other standards, including the IMS Global Learning Consortium's Meta-Data Specification. LOM has also been mapped to Dublin Core. Thus, for the sake of interoperability, the MultiMatch metadata schema can take this standard into account. For the sake of interoperability the MultiMatch metadata schema should take this standard into account.

Apart from the ERIC thesaurus, there is not much agreement on the use of controlled vocabularies in this domain. The usability for MultiMatch, of the controlled vocabularies selected for this deliverable depends mainly on the subject area of the content providers of this project.

5.1.5 Audiovisual sector

The audiovisual sector employs generic schemas such as FRBR and MPEG-7, as well as more domain specific schemas, namely **P_META** and **SMEF-DM**. The focus of each of these metadata schemas will be compared with the goals of MultiMatch in order to decide on their usability.

In this domain no national or international standards appear to exist. The institutions mainly use proprietary controlled vocabularies, generally monolingual. In the first instance, the controlled vocabularies in use by the content providers of this project will be taken into account. See also paragraph 5.3 on the possible role of controlled vocabularies within MultiMatch.

5.1.6 Geospatial sector

The standardization efforts in the geospatial sector include two important metadata schemas, namely the Standard for Digital Geospatial Metadata (**CSDGM**) and the **ISO 19115:2003** standard.

At this moment it is not yet clear to what extent geospatial information on the cultural heritage objects concerned will be applied in MultiMatch. The selected standards present an overview of the possibilities in this field. It is likely that, to start with, only longitude and latitude will play a role in the MultiMatch database. Therefore the TGN structured vocabulary with geographic names, widely used by museums, is likely to be very useful for MultiMatch as it includes coordinates.

Geospatial information in MultiMatch can take the form of controlled vocabularies for names of places; where the cultural heritage objects concerned have been found or where the physical objects concerned actually are stored. When the cultural heritage object itself is a site of natural heritage, archaeology or open cultural heritage, the metadata schema of MultiMatch can be enriched with specific geospatial metadata. This metadata can be applied to provide geographical functionality in the user interface, such as the use of a map to form a query or display results. The standards mentioned in section 2.8.1 can be used as guidance. An example of a globally distributed georeferenced digital library is Alexandria Digital Library (ADL) http://www.alexandria.ucsb.edu/research/index.htm .

5.1.7 Generic metadata schemas

The three generic metadata schemas described in this deliverable, Dublin Core, MPEG-7 and MPEG-21, are all very pertinent to MultiMatch. Dublin Core provides a true core schema (Dublin Core Simple) together with the opportunity to expand the schema with extra qualifiers. MPEG-7 and MPEG-21 provide far more comprehensive (and complex) standards, and are particularly suited to digital media. They are all fairly widely used inside, as well as outside, the cultural heritage domain. Which of them matches the focus of MultiMatch best is object of further research. As stated in the Introduction; this research will be an organised effort with other WP's and will be documented in D2.2.

5.1.8 Reference models

As discussed above, the "CIDOC object-oriented Conceptual Reference Model" (**CRM**) represents an ontology for cultural heritage information – i.e. it describes, in a formal language, the explicit and implicit concepts and relations underlying the documentation structures used for cultural heritage. CRM covers any kind of data (either "descriptive" data or "authority" data) created by museums in the fields of fine arts, archaeology, natural history, etc.

The development of CRM has been going on for more than a decade, and since September 2000 it has been progressing as an ISO standard (ISO/AWI 21127) in a joint effort of the CIDOC CRM SIG and ISO/TC46/SC4. It was approved as a standard in September 2006.[90] The primary role of the CRM is to serve as the semantic 'glue' needed to transform disparate, localized information sources into a coherent and valuable global resource.

The central notion in CIDOC CRM is the notion of Event: something that happens in space and time and brings about some change in the world.
An **event** (i.e., an instance of the Event class) can involve:

- instances of the Actor class (persons, groups…), who can play a decisive role in provoking the event or just witness it or undergo it, and who are referred to through instances of Actor Appellation;
- bits of the physical world and/or creations of the mind (i.e., instances of the class named Physical Thing and/or the class named Conceptual Object; e.g., canvass and paint on the one hand, and the image formed by the paint on the canvass, on the other hand), which are referred to through instances of Appellation (names, titles, codes, whatever).

In addition, an event:

- occurs in **time**, and has therefore a duration, i.e., an instance of the class named Time-Span, which is referred to through instances of Time Appellation (e.g., instances of Date);
- and occurs in **space**, and can as such be related to an instance of Place, which is referred to through an instance of Place Appellation.

The CIDOC CRM is not a metadata standard, but can be used to express metadata standards. It might play a useful role in the integration or mapping of the diverse metadata structures of the content providers to a central metadata schema for MultiMatch. The CIDOC CRM model is highly likely to have a significant impact within MultiMatch, especially considering:

- the positive development of this reference model into a standard;
- its focus on documenting cultural heritage;
- its scope covering rich information exchange between museums, libraries and archives;
- the applications of CIDOC CRM at this point;
- ongoing academic interest in the model, ongoing research.

90 (http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=34424&scopelist=PROGRAMME

There has also been interesting and potentially relevant research into combining the use of CIDOC CRM and MPEG-7, as CRM does not seem to focus enough on describing multimedia content, including the technical features of segments and shots.[91]

The **FRBR** reference model has been developed in the library domain with the aim to provide the user with navigation tools across related works in an otherwise flat structured, bibliographic catalogue. The audiovisual sector has also applied this reference model to design a data model[92], with all the required entities, and also to provide a hierarchical metadata schema for digital libraries with various, cultural AV material.

The application of the FRBR model as a common ground for interchange and integration between libraries fits well with the current focus on cross-domain semantic interoperability in digital libraries. Norwegian University of Science and Technology is participating in the DELOS NoE activity on development of the FRBROO ontology, and future activities include adapting the conversion system to produce bibliographic information encoded as RDF using the FRBROO ontology, for the purpose of cross-domain integration and interoperability using semantic Web technology.[93]

The idea that both the library and museum communities might benefit from harmonising the two models was first expressed in 2000, on the occasion of the ELAG 2000 conference. This idea led eventually to the formation in 2003 of the International Working Group on FRBR/CIDOC CRM Harmonisation.

The common goals are to express the IFLA FRBR model with the concepts, ontological methodology and notation conventions provided by the CIDOC CRM, and to merge the two object-oriented models thus obtained. This Working Group is now being supported by the DELOS NoE. The first draft definition of the **FRBROO** model, i.e. the object-oriented version of FRBR harmonized with CIDOC CRM (version 0.6.7)[94], is available for public discussion since August 2006.

This formal ontology is intended to capture and represent the underlying semantics of bibliographic information and to facilitate the integration, mediation and interchange of bibliographic and museum information. Its major innovation is a realistic, explicit model of the intellectual creation process (see Figure below). Work will continue with modelling information about authority records and performing arts.[95]

---

[91] Combining the CIDOC CRM and MPEG-7 to describe multimedia in museums / Jane Hunter, 2002. In Proceedings of the International Conference about Museums and the Web. Boston, Massachusetts (2002)

The use of CRM Core in Multimedia Annotation / Patrick Sinclair [et al.], 2006. In: Proceedings of First International Workshop on Semantic Web Annotations for Multimedia (SWAMM), Edinburgh, Scotland.

Cultural Heritage on the Semantic Web – the Museum24 project / Barnabas Szasz [et al.], 2006. Presented at Symposium on Digital Semantic Content across Cultures. Paris, 4-5 May, 2006

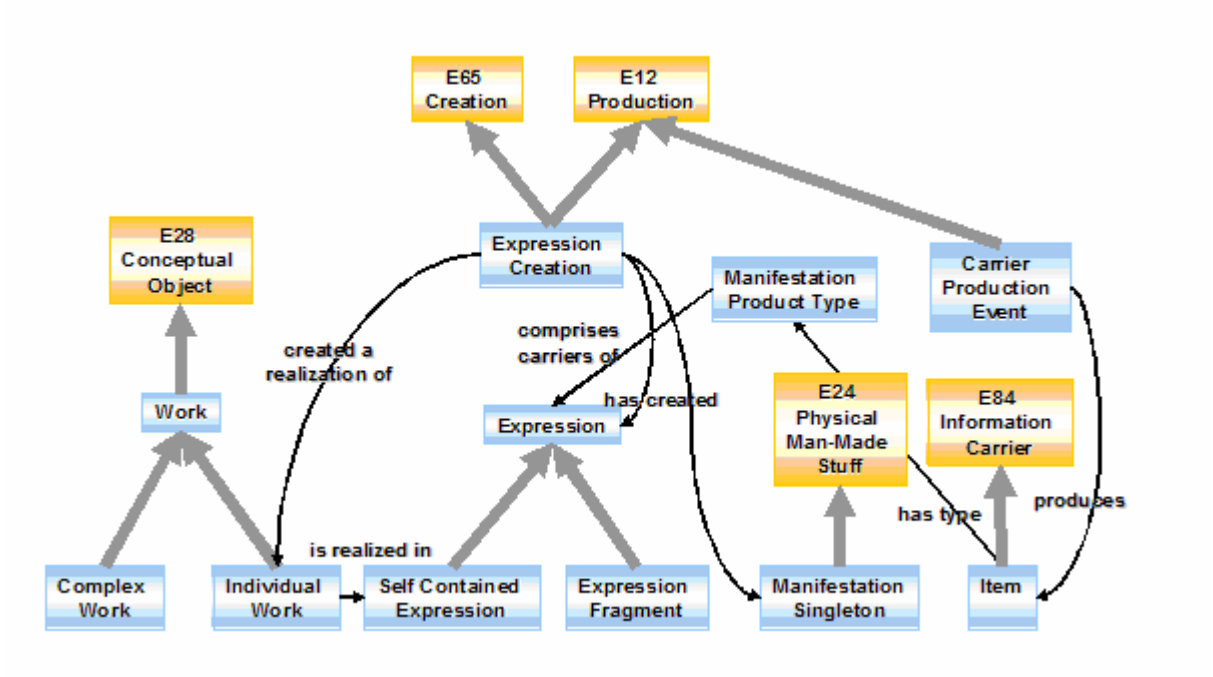
[92] A technical model for a logical database structure of an information system.

93 A Tool for Converting Bibliographic Records / Trond Aalberg, July 2006 http://www.ercim.org/publication/Ercim_News/enw66/aalberg.html

[94] http://cidoc.ics.forth.gr/docs/frbr_oo/FRBR_oo_V0.6.7_no_mapping.doc

The comprehensive description of the FRBRoo Model, comprising the draft definition (version 0.6.7), the mapping from FRBRer to FRBRoo and all CRM constructs directly referred to in the definition and the mapping is available at http://cidoc.ics.forth.gr/docs/frbr_oo/FRBR_oo_V0.6.7.doc

[95] Increasing the Power of Semantic Interoperability for the European Library / Martin Doerr, July 2006 http://www.ercim.org/publication/Ercim_News/enw66/doerr.html



Partial model of the intellectual creation process.

A recent practical application of these models is the derivation of the CRM Core Metadata schema[96], which is compatible and similar in coverage and complexity to Dublin Core, but much more powerful. It allows for a minimal description of complex processes, scientific and archaeological data, and is widely extensible in a consistent way by the CRM-FRBR concepts. According to Doerr, CRM Core can be easily used by Digital Libraries. [Doerr 2005, 2006]

It is expected that FRBROO will be regarded as a new, "official" release of the IFLA FRBR model. FRBROO will be used for implementation purposes, most notably in the context of integrated information system design and Semantic Web activities. (Zoemer & Le Boeuf)

Detailed analyses of CIDOC CRM, FRBR, CRM Core and of the harmonization version, in relation to MultiMatch, will be included in D2.2.

5.2 Overview of the Most Important Metadata Schemas

5.2.1 Methodology

Until now, this deliverable has described the dimensions of each schema using tables outlining for example the descriptions, number of elements, URL with further information etc. In order to distinguish one schema from the other and to select the appropriate standard, we needed an objective framework that focuses on the differences and applicability from one schema over the other. The methodology from De Sutter (et. al.) in their paper "Evaluation of Metadata Standards in the Context of Digital Audio-Visual Libraries"[97] [Sutter, 2006] will be used. In the paper, the authors describe criteria that can be used to select the metadata standard that is best suited for the application in mind.

[96] Definition of the CIDOC CRM: http://cidoc.ics.forth.gr

97 Sutter, R de. [et. al.] Evaluation of Metadata Standards in the Context of Digital Audio-Visual Libraries. Published in: Julio Gonzalo, Costantino Thanos, M. Felisa Verdejo, Rafael C. Carrasco (Eds.): Research and Advanced Technology for Digital Libraries, 10th European Conference, ECDL 2006, Alicante, Spain, September 17-22, 2006, Proceedings. Lecture Notes in Computer Science 4172 Springer 2006.

The criteria listed by the authors are composed in such a way that all aspects ranging from content organization to the different types of metadata are taken into account, but independently to any restriction imposed by a particular media asset management system.

Four criteria are used:

1. Internal vs. Exchange Metadata Model
2. Flat vs. Hierarchical Metadata Model Criterion
3. Supported Types of Metadata
4. Syntax and Semantics

These are explained below (excerpts from the paper, amended for the scope of this deliverable).

- **Criterion 1: Internal vs. Exchange Metadata Model**. On the one hand, particular metadata models are specifically developed for managing the metadata in the interior of a system. These metadata models are further referred to as internal metadata models. Usually, these metadata models are represented as Entity Relationship Diagrams which describe the architecture of the database that stores the metadata of the audio-visual material.
  On the other hand, other metadata models are used to describe the way the information is to be transmitted from source to destination. Here, the metadata models are called exchange metadata models. These models are used to exchange information about the audio-visual material and are specifically intended for the transmission of metadata between different systems. Here, exchange must be seen as broad as possible, namely between any combination of content creator, content distributor, archive, and consumers.

- **Criterion 2: Flat vs. Hierarchical Metadata Model.** The structural organization of the description of the essence is a second criterion. In general, the heritage organizations decide how detailed the metadata needs to be. Two extreme visions can be identified. On the one hand the essence is considered as an elementary and indivisible unit, resulting in a coarse description, and on the other hand, the essence is divided in small sub-pieces each annotated separately, resulting in a fine-detailed description.
  - If the essence is considered as an elementary and indivisible unit, the cultural heritage organisations can associate this elementary unit with, for example, a program. The metadata describes the essence as a whole and does not describe the individual parts therein. This model is mostly referred to as a flat metadata model.
  - Sub-parts of the essence can be annotated with much more detail. The additional metadata belongs to the individual parts and permits the users of the archive to perform more detailed searches on the content. Every editorial object can be annotated with additional descriptive metadata, so it is possible to search on the editorial object itself. In turn, editorial objects can be broken down into different media objects. Hence, it is not necessary to repeat the same information for every object, but the program inherits information from other levels. The underlying idea is that information has to be added to the objects at the right location. This concept is referred to as a hierarchical metadata model.

- **Criterion 3: Supported Types of Metadata.** Metadata describes the essence. The requirements of the users determine the needed types of metadata. There are two rules that must be observed as explained in the introduction of this paper: 1) essence is unusable without metadata, and 2) the content is valueless without rights information. Hereafter different types of metadata for the preservation of audio-visual material are discussed.
  - Identification metadata. The identification metadata is primarily about the information to singularly identify the essence. This can be done by human interpretable fields, like a title or an index, or by machine understandable identifiers, like a Unique Material Identifier (UMID) or a Uniform Resource Identifier (URI). Besides the identification metadata related

to the essence, other identifying information is necessary to locate related documents that are potentially stored in another system.

- Description & classification metadata. The descriptive metadata describes what the essence expresses. This could be done by creating a textual summary of the contents of (for example) a television programme or book. In some cases, the keywords are selected from an organized dictionary of terms, i.e., the thesaurus. A classification is a system of coding and organizing materials (books, serials, audiovisual materials, computer files, maps, manuscripts, realia[98]) according to their subject. A classification consists of tables of subject headings and classification schedules used to assign a class number to each item being classified, based on that item's subject.
- Technical metadata. The technical metadata describes the technological characteristics of the related essence. For example technical metadata relating to the digitization process (i.e., scanner model, scanner resolution, color schemes, file size of the master file, etc.)[99]
- Security & rights metadata. The security metadata handles all aspects from secure transmission (i.e., the encryption method) to access rights. The latter augments the content into an asset. The access rights metadata can be split up in information about the rights holder and information about contracts. The rights holder is the organization who owns the rights of the audio-visual material.

- **Criterion 4: Syntax and Semantics.** Some standards define only syntax, others only semantics, and some define both. The syntax defines how the representation of the metadata must be done. One of the most important questions about the syntax is the choice between a textual and a binary representation. The textual representation has the advantage that the metadata is human readable, but at the same time it is very verbose. The binary representation is dense, but it has the disadvantage that it can only be handled by machines. In case of plain text notation, the XML is mostly used. If so, the metadata standard provides, besides the standard itself, an XML Schema that punctiliously determines the syntax of the metadata.
  Using the XML Schema makes it possible to check the correctness (i.e., validity) of the metadata. This characteristic enables interoperability.
  The semantics of the metadata standard determine the meaning of the metadata elements. Without any semantic description, one is free to assume the denotation of the different metadata elements, presumably resulting in different interpretations thereof between users. Only if the description of the metadata elements is closed (i.e., every metadata element is semantically described), all users must agree on the sense of the metadata elements improving the interoperability.

5.2.2 The schema

As multimedia is a factor to consider in MultiMatch, an extra column is added to the schema here below to indicate if the metadata schema is suitable to describe multimedia content and contains metadata elements for technical metadata to identify shots and segments and to contain technical features (i.e. color histograms). This is an important criterion for the MultiMatch metadata schema, when one considers the fact that the metadata associated with multimedia objects are infinitely more complex than simple metadata for resource discovery of simple atomic textual documents.

In the schema below the most striking aspects are highlighted. See further in section 5.3.2.

98 In library classification systems, "realia" are objects (such as coins, tools, games, toys, or other physical objects) that do not easily fit into the neat categories of books, periodicals, sound recordings, etc.

99 http://www.cdlib.org/inside/diglib/guidelines/basicreqs.html#techmd

Metadata schema[100]	Criterion 1	Criterion 2	Criterion 3	Criterion 4	(primary) sub-domain	Multimedia
CDWA	Internal	hierarchical	identification, description, administrative (including conservation and treatment history), rights	no, closed semantics	Museums	No
CDWA Lite	exchange	hierarchical	identification, description, administrative	XML, closed semantics	Museums	No
Dublin Core	Exchange	flat	identification, description, technical (limited)	[101], open semantics	**All**	No
CSDGM	internal + exchange	flat	all (extensive spatial metadata)	XML, closed semantics	Geospatial sector	No
IEEE LOM	Internal	flat	identification, description (including pedagogical metadata, administration	XML, closed semantics	Educational sector	No
ISAD(G)	Internal	hierarchical	identification, description, administrative, rights	no, closed semantics	Archives	No
ISO 19115:2003	Exchange	flat	identification, description (extensive spatial metadata)	XML, closed semantics	Geospatial sector	No
MARC	Internal	hierarchical	All	XML, closed semantics	Libraries	No
MODS	Exchange	hierarchical	identification, description	XML, closed semantics	Libraries	No
MPEG-7	Exchange	hierarchical	identification, description, technical	XML, closed semantics	Audio visual sector	**Yes**
Object ID	Exchange	flat	identification (extensive), description	[102], closed semantics	Museums	No
P_Meta	Exchange	hierarchical	All	XML, closed semantics	Audio visual sector	Yes
SMEF-DM	Internal	hierarchical	all (extensive rights metadata)	ERD, open semantics	Audio visual sector	Yes
VRA Core	Exchange	hierarchical	identification, description, administrative, rights	no, closed semantics	Museums	No

100 The Encoded Archival Description is not included in this schema, while the purpose of this metadata schema is to descibe finding aids, and not to archive objects themselves.

101 Dublin Core can be mapped to XML and RDF.

102 The schema is only available as text with guidelines included, but can easily be put into XML, as the number of metadata elements is less than 15.

5.3 Further Research

We conclude this deliverable outlining the further research to be done in order to make an informed choice regarding the use of metadata schema and controlled vocabularies for MultiMatch. This process and its outcome will be presented in D2.2 (PM10). The user requirements that are now nearly defined by WP1 will provide pivotal input for D2.2, as will the detailed specifications of the first prototype defined by WP3 and decisions on the technical infrastructure.

5.3.1 Interoperability and MultiMatch

Interoperability means enabling information that originates in one context (i.e. system, department, process, organization) to be used in another, in ways that are as automated as possible.

The DELOS Network of Excellence has finished a comprehensive report on Semantic Interoperability in Digital Library Systems[103]. The report (written as part of the WP5 cluster "Knowledge Extraction and Semantic Interoperability") defines interoperability very broadly as enabling any form of inter-system communication, or the ability of a system to make use of data from a previously unforeseen source. Interoperability in general is concerned with the capability of different information systems to communicate. This communication may take various forms such as the transfer, exchange, transformation, mediation, migration or integration of information.

Semantic interoperability ("SI") is characterised by the capability of different information systems to communicate information consistent with the intended meaning of the encoded information (as intended by the creators or maintainers of the information system). It involves:

- processing of the shared information so that it is consistent with the intended meaning
- encoding of queries and presentation of information so that it conforms with the intended meaning regardless of the source of information.

It is to be expected that the MultiMatch infrastructure will be set up in such a way, that the ability of MultiMatch to process heterogeneous data from various sources will be maximised. Interoperability is one of the pivotal research questions MultiMatch focuses on.

The metadata schema of MultiMatch will have to be presented in XML format, the almost uniformly adopted standard, which will facilitate the interaction with Semantic Web technologies. A metadata standard with closed semantics improves interoperability, if the usage of that standard is widely spreadIf the metadata schema of MultiMatch builds on a standard with closed semantics, it is likely that those semantics will be respected. However, it cannot be excluded beforehand that some of the semantics will be changed in order to fulfill other requirements of this project.

Semantic interoperability, for example via automated mapping of metadata structures in the provided resources to the MultiMatch metadata schema, will be reported in D2.2.

103 DELOS - D5.3.1: Semantic Interoperability in Digital Library Systems, 29th June 2005.

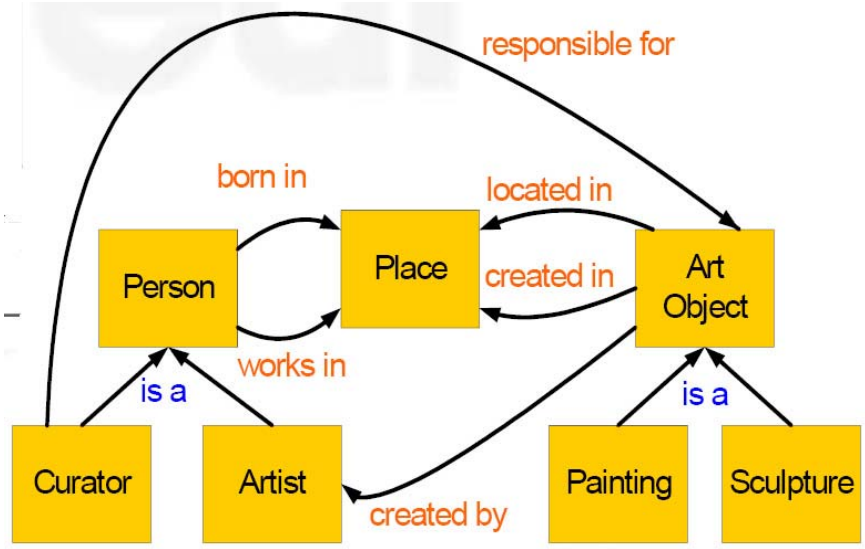
5.3.2 Data Model and Metadata Schema for MultiMatch

The current line of thinking can at this point be outlined briefly. This is the starting point for the next activity in WP2. Both the data model and the metadata schema for MultiMatch will have to fulfil the user requirements and the technical and administrative project requirements. These requirements (both technical and user requirements) will influence the approach regarding the metadata schema and the data model for MultiMatch, as well as the concepts by which searching is provided.

For instance, if the user should be able to find images of paintings with specific physical features, then the concepts in the data model and the elements of the metadata schema of MultiMatch should be able to facilitate these queries. Another example of a concept search that introduces specific wishes for the schema is: "for which countries the MultiMatch database contains 19th century photographs?"

Relationships between objects are important in the domain MultiMatch is covering. Typical library objects, such as books, can be about museum objects, and museum object can represent events or characters found in books (e.g. an artistic criticism of Millais's painting of "Ophelia's death"). It seems possible that the data model of MultiMatch will have to deal with this fact; the user requirements will make clear if this assumption is correct. If so, the further research for D2.2 will include the question whether such interrelationships should be integrated in common information storage, or at least virtually integrated through mediation devices that allow a query to be simultaneously launched on distinct information depositories, which requires common semantic tools such as FRBROO 'plugged into' CIDOC CRM.[104]

The figure below is taken from a presentation on the SCULPTEUR project[105] [106] to illustrate here the role of concepts and their relationships in a data model.



MultiMatch acknowledges the fact that current and future content providers will typically not apply the same data model and metadata schema. Core MultiMatch metadata will be extracted from the metadata provided describing the selected cultural heritage objects, and converted into the central metadata schema. The rest of the metadata, contained in the possibly rich descriptions provided, will be admitted to the semantic background information of MultiMatch. Thus making it possible:

104 Conceptual models: museums & libraries: towards an object-oriented formulation of FRBR aligned on the CIDOC CRM ontology / Maja Žumer (University of Ljubljana) & Patrick Le Bœuf (National Library of France). - ELAG 2006 "New tools and new library practices", Bucharest, 26 April 2006

105 http://www.sculpteurweb.org/

106 Methods for search and retrieval of large multimedia collections : ECDL2004 Tutorial A / Matthew Addis, IT Innovation Centre.

- For the user to read the content of these metadata, when viewing the search results and
- For the metadata provided to play a useful role in associative searching.

The metadata schema for MultiMatch will have to contain all the elements needed to describe the cultural heritage objects within the domain or scope of this project.

Further research in MultiMatch will be aimed at answering various questions. For example:

- Will a metadata schema for the management of digital objects suffice, or does MultiMatch also have to deal with physical cultural heritage objects and their descriptions?
- Will MultiMatch deal with both primary objects (e.g. cooking pots, paintings, monumental buildings, artistic photographs, literature) and secondary objects (e.g. photographs of, journal articles on, video recordings of cooking pots, paintings, monumental buildings, artistic photographs, literature)?

To start with, we do not want to exclude a specific sub-domain of the cultural heritage domain a priori. Therefore it is to be expected that the MultiMatch metadata schema will be as generally applicable as possible. Further research will make clear whether one of the standard metadata schemas described in this deliverable can fulfil all the requirements of MultiMatch. These requirements will also dictate the desired degree of *granularity* of the metadata schema needed. In other words, the amount of detail to be captured and represented in the metadata record. Will a "core record", such as the Dublin Core with its fifteen element set (any of which are optional, repeatable, and extensible) do? Does MultiMatch need a rich, detailed metadata to adequately represent the resources and the particular purposes of the project? Or will we define a proprietary MultiMatch core schema based on Dublin Core for the sake of interoperability or other issues?

The various scenarios will be studied in the next phase of the project. In any case, the particular intent of the services provided and the types of metadata to be supported will influence the metadata schema for MultiMatch. Taking into account the four criteria to evaluate metadata schemas, as presented in section 5.2, the metadata schema for MultiMatch will most likely have the following properties:

1. The metadata schema for MultiMatch will not be focused on either the internal system or on exchange. The infrastructure of MultiMatch will make the transmission of metadata between heterogeneous collections possible. The metadata schemas in use with the content providers will have to be mapped (semi)automatically to the MultiMatch schema. There are several so-called crosswalks available that map the structural components of the metadata standards described in this deliverable. The metadata schema of MultiMatch will be used to present information about the cultural heritage material and (ontology based) navigate through this material.
2. The metadata schema will have to describe the cultural heritage object as a whole, as well as relevant sub-parts of the essence and digital surrogates; the schema needs thus to be hierarchical of nature.
3. It is to be expected, that MultiMatch will handle various types of metadata:
    a. descriptive metadata – both metadata that formally describe the object (for example title, creator, creation date) as well as some semantic elements (for example subject keywords, geographic places);
    b. technical metadata – probably mainly concerning the surrogate or the image of the cultural heritage object, and less concerning the physical cultural heritage object itself;
    c. administrative metadata – some metadata to administer the objects concerned (e.g. content provider name, location information, language of record, record number), possibly also some metadata for the rights management. For instance, the extent to which metadata on copyrights are needed within the central metadata schema is at this point not clear.
    d. Examples of typical metadata that will probably not be the focus of the MultiMatch metadata schema include administrative and technical data on museum objects that are needed for the internal management of the museum collection (typically gallery and museum information systems)

Considering the first analysis in section 5.2.2 together with this section in the context of MultiMatch, it is possible to conclude this deliverable indicating that the following standards will be taken into account in D2.2:

- Dublin Core: because it is in use through the whole of the cultural heritage domain.
- MPEG-7: because it can handle multimedia in a way appropriate for MultiMatch.
- FRBR: because it provides a data model with relationships and a hierarchy that are probably useful for MultiMatch. (Annex 3 includes the graphical representation of the FRBR entity-relationship model).
- CIDOC CRM: because it provides a reference model for the cultural heritage domain. (Annex 4 includes the graphical representation CIDOC class hierarchy).

5.3.3 Controlled vocabularies and MultiMatch

In section 5.1 the possible role of several controlled vocabularies in the semantic background information is mentioned. In the deliverable to follow, D2.2, the possible roles of controlled vocabularies will be studied in more detail, in the context of the then known user requirements.

It is likely that the following roles will be reviewed in D2.2:

1. input control and search assistance via closed lists of preferred terms;
2. the browse functionality in the MultiMatch user interface. Dalmau affirms that structured forms of browse and search can be successfully integrated into digital collections to significantly improve the user's discovery experience.[107]
3. supporting multilingual searching via multilingual term lists or thesauri;
4. the reinforcement of the semantic background information via the associations and background information the controlled vocabularies can provide in the semantic web on the cultural heritage objects (see also: section 4.3). Classifications and thesauri can be seen as ontologies with a limited number of relationships between concepts.

As no parts of the cultural heritage domain will be excluded beforehand, the metadata schema of MultiMatch will probably require an integrated, shared ontology for the information accumulated by archives, libraries, museums as well as by the other identified sub-domains. This shared ontology will make it possible for all the collections that the participants in this domain hold, and attribute to the vision of a 'digital continuum' with unrestricted, sustainable and reliable digital access to Europe's cultural heritage.

Acknowledgments

The authors gratefully acknowledge the internal reviewing of a near final version of this deliverable by project partners from ISTI-CNR, UniGE and USFD and external expert Dr. Véronique Malaisé from the Vrije Universiteit Amsterdam. Their comments and suggestions offered proved very useful in preparing the final version.

[107] Integrating thesaurus relationships into search and browse in an online photograph collection / Michelle Dalmau et al. - Library Hi Tech. - Vol.23 no.3, 2005; p. 425-452.

Annex 1. Abbreviations of the standards mentioned

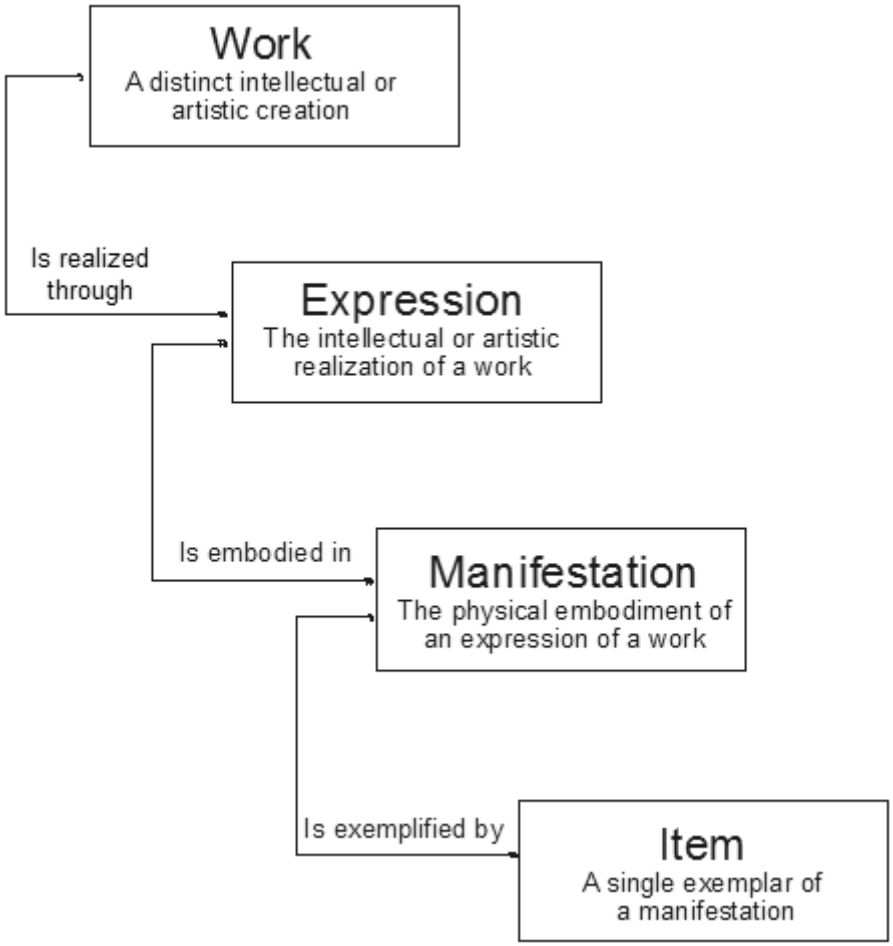
AAT	Art and Architecture Thesaurus
CDWA	Categories for the Description of Works of Art
CIDOC CRM	CIDOC Conceptual Reference Model
CSDGM	Standard for Digital Geospatial Metadata
DCMI	Dublin Core Metadata Initiative
DDC	Dewey Decimal Classification code
EAD	Encoded Archival Description
FIAF	International Federation of Film Archives Cataloguing Rules
FRAR	Functional Requirements on Authority Records
FRBR	Functional Requirements for Bibliographic Records
GEM	Gateway to Educational Material
ISAD(G)	General International Standard Archival Description
IPTC thesaurus	International Press Telecommunications Council thesaurus
ISAAR (CPF)	International Standard Archival Authority Record for Corporate Bodies, Persons and Families
LCC	Library of Congress Classification
LCSH	Library of Congress Subject Headings
LOM	IEEE Standard for Learning Object Metadata
MARC	Machine Readable Cataloguing
METS	Metadata Encoding and Transmission Language
MPEG-7	Multimedia Content Description Interface
MPEG-21	Moving Picture Experts Group, MPEG-21 standard
MODS	Metadata Object Description Schema
MXF	Material Exchange Format
OGC Specifications	Open Geospatial Consortium Specifications
P_META	P_META Exchange scheme
RDF	Resource Description Framework
SKOS Core	Simple Knowledge Organisation System
SMEF-DM	Standard Media Exchange Framework Data Model
SMPTE MD	SMPTE Metadata Dictionary
TGN	Thesaurus of Geographic Names
UDC	Universal Decimal Classification code
ULAN	Union List of Artists Names
VRA Core	Visual Resources Association Core Categories

Annex 2. Selected Biography

Aalberg, Trond (2006) A Tool for Converting Bibliographic Records. NTNU.
Bœuf, Patrick Le. Using an ontology-driven system to integrate museum information and library information. Paper presented on the occasion of the Symposium on Digital Semantic Content across Cultures, Paris, the Louvre, 4-5 May 2006.
Childress, Eric. Metadata standards. Presentation for FEDLINK OCLC Users Group Meeting. November 18th 2003.
Dalmau, Michelle. Integrating thesaurus relationships into search and browse in an online photograph collection. Library Hi Tech. - Vol.23 no.3, 2005; p. 425-452.
Doerr, Martin. Increasing the Power of Semantic Interoperability for the European Library. Published in: Julio Gonzalo, Costantino Thanos, M. Felisa Verdejo, Rafael C. Carrasco (Eds.): Research and Advanced Technology for Digital Libraries, Proceedings: 10th European Conference, ECDL 2006, Alicante, Spain, September 17-22, 2006.
Doerr, Martin. The CIDOC CRM, a Standard for the Integration of Cultural Information. Presentation. Nurnberg, 14-15 November 2005.
Szakadát István; Lois, László; Knapp, Gábor (2005) New Methods for Enhancing the Effectiveness of the Dublin Core Metadata Standard Using Complex Encoding Schemes Document Actions. In: Information Systems Development. Advances in Theory, Practice, and Education, edited by Vasilecas, O., Caplinskas, A., Wojtkowski, G., Wojtkowski, W., Zupancic, J., Wrycza, S. . Springer, pages 365-375.
Hunter, Jane. Combining the CIDOC CRM and MPEG-7 to describe multimedia in museums. In Proceedings of the International Conference about Museums and the Web. Boston, Massachusetts , 2002.
Jong, A. de (2002) Metadata in the audiovisual production environment : an introduction, Hilversum, Beeld en Geluid.
Sinclair, Patrick [et al.]. The use of CRM Core in Multimedia Annotation. In: Proceedings of First International Workshop on Semantic Web Annotations for Multimedia (SWAMM), Edinburgh, Scotland, 2006.
Smeulders, Arnold W.M., Lynda Hardman, Guus Schreiber, and Jan-Mark Geuzebroek (2003) An integrated multimedia approach to cultural heritage e-documents.
Sutter, R de [et. al.]. Evaluation of Metadata Standards in the Context of Digital Audio-Visual Libraries. Published in: Julio Gonzalo, Costantino Thanos, M. Felisa Verdejo, Rafael C. Carrasco (Eds.): Research and Advanced Technology for Digital Libraries, 10th European Conference, ECDL 2006, Alicante, Spain, September 17-22, 2006, Proceedings. Lecture Notes in Computer Science 4172 Springer 2006.
Szasz. Barnabas [et al.]Cultural Heritage on the Semantic Web – the Museum24 project. Presented at Symposium on Digital Semantic Content across Cultures. Paris, 4-5 May, 2006.
Žumer, Maja and Patrick Le Bœuf. Conceptual models: museums & libraries: towards an object-oriented formulation of FRBR aligned on the CIDOC CRM. ELAG 2006 “New tools and new library practices”,Bucharest, 2006.

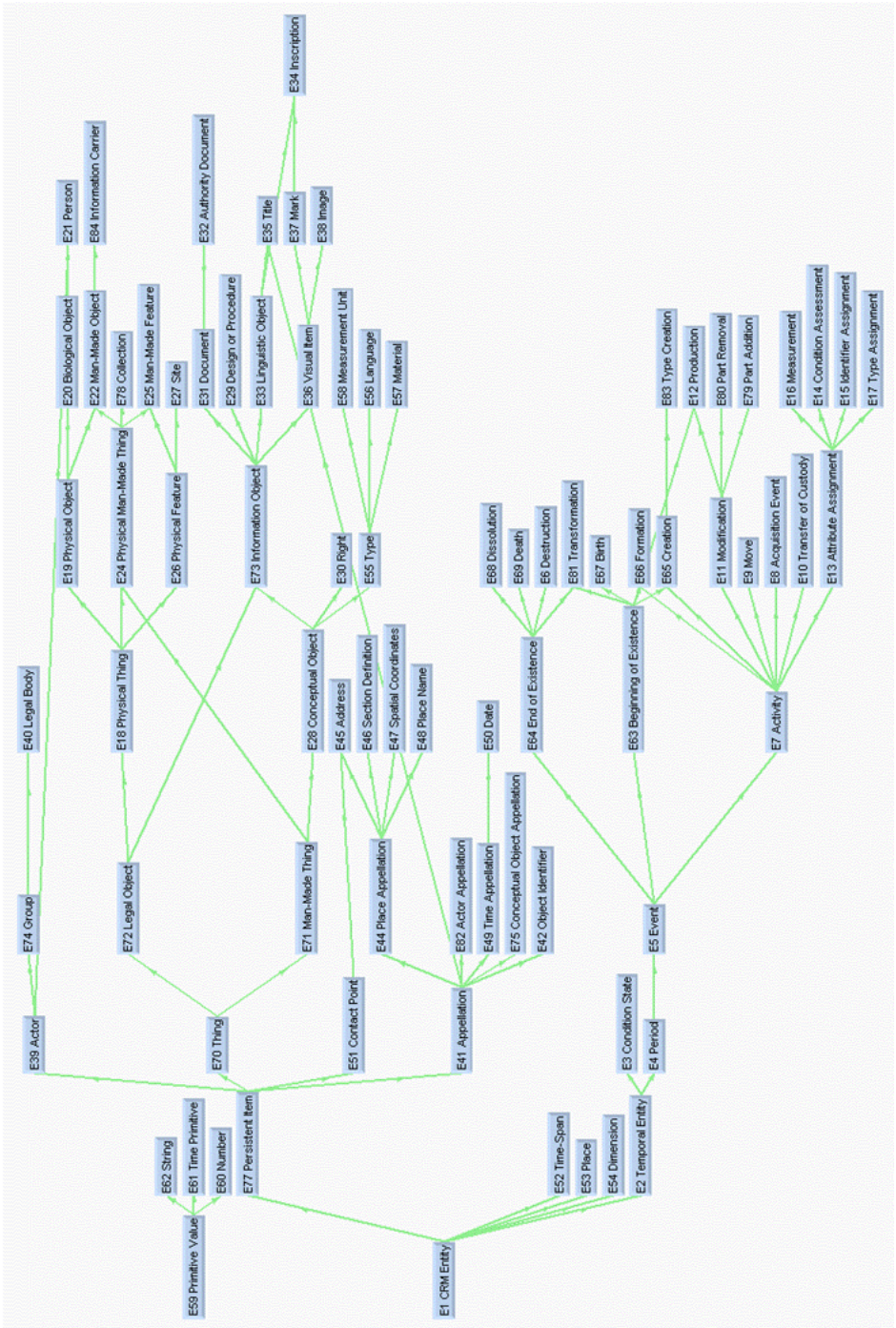
Annex 3. FRBR entity-relationship model

Source: http://www.dlib.org/dlib/september02/hickey/hickey-fig1.gif



Annex 4. CIDOC class hierarchy

Source:
http://cidoc.ics.forth.gr/cidoc_graphical_representation/crm_class_hierarchy_files/class_hierarchy.gif



Annex 5. Alinari Dublin Core element set

Dublin Core	Element Name: Title	
	Label:	Title
	Definition:	A name given to the resource.
	Comment:	Typically, Title will be a name by which the resource is formally known.

Dublin Core	Element Name: Creator	
	Label:	Creator
	Definition:	An entity primarily responsible for making the content of the resource.
	Comment:	Examples of Creator include a person, an organization, or a service. Typically, the name of a Creator should be used to indicate the entity.

Dublin Core	Element Name: Subject	
	Label:	Subject and Keywords
	Definition:	A topic of the content of the resource.
	Comment:	Typically, Subject will be expressed as keywords, key phrases or classification codes that describe a topic of the resource. Recommended best practice is to select a value from a controlled vocabulary or formal classification scheme.

Dublin Core	Element Name: Description	
	Label:	Description
	Definition:	An account of the content of the resource.
	Comment:	Examples of Description include, but is not limited to: an abstract, table of contents, reference to a graphical representation of content or a free-text account of the content.

Dublin Core	Element Name: Publisher	
	Label:	Publisher
	Definition:	An entity responsible for making the resource available
	Comment:	Examples of Publisher include a person, an organization, or a service. Typically, the name of a Publisher should be used to indicate the entity.

Dublin Core	Element Name: Contributor	
	Label:	Contributor
	Definition:	An entity responsible for making contributions to the content of the resource.

Comment:	Examples of Contributor include a person, an organization, or a service. Typically, the name of a Contributor should be used to indicate the entity.

Dublin Core	Element Name: START-Date	
	Label:	START-Date
	Definition:	A date of an event in the lifecycle of the resource.
	Comment:	Typically, Date will be associated with the creation or availability of the resource. Recommended best practice for encoding the date value is defined in a profile of ISO 8601 [W3CDTF] and includes (among others) dates of the form YYYY-MM-DD. Look at the END-Date for range conventions.

ALINARI	Element Name: END-Date	
	Label:	END-Date
	Definition:	A date of an event in the lifecycle of the resource.
	Comment:	This date is needed for a range dating method (i.e. 1700 - 1888). If we use a range dating convention then START-Date is the beginning (1700) and END-Date is the closing period (1888). If the date is exact then START-Date=END-Date. Typically, Date will be associated with the creation or availability of the resource. Recommended best practice for encoding the date value is defined in a profile of ISO 8601 [W3CDTF] and includes (among others) dates of the form YYYY-MM-DD.

Dublin Core	Element Name: Type	
	Label:	Resource Type
	Definition:	The nature or genre of the content of the resource.
	Comment:	Type includes terms describing general categories, functions, genres, or aggregation levels for content. Recommended best practice is to select a value from a controlled vocabulary (for example, the DCMI Type Vocabulary [DCT1]). To describe the physical or digital manifestation of the resource, use the FORMAT element.

Dublin Core	Element Name: Format	
	Label:	Format
	Definition:	The physical or digital manifestation of the resource.
	Comment:	Typically, Format may include the media-type or dimensions of the resource. Format may be used to identify the software, hardware, or other equipment needed to display or operate the resource. Examples of dimensions include size and duration. Recommended best practice is to select a value from a controlled vocabulary (for example, the list of Internet Media Types [MIME] defining computer media formats).

Dublin Core	Element Name: Identifier	
	Label:	Resource Identifier
	Definition:	An unambiguous reference to the resource within a given context.
	Comment:	Recommended best practice is to identify the resource by means of a string or number conforming to a formal identification system. Formal identification systems include but are not limited to the Uniform Resource Identifier (URI) (including the Uniform Resource Locator (URL)), the Digital Object Identifier (DOI) and the International Standard Book Number (ISBN).

ALINARI	Element Name: Resource Name	
	Label:	File Name
	Definition:	An unique name to the resource without extention
	Comment:	Tipically the inventory code of the resource. File name [a-z A-Z 0-9 _ -] with no special carachters, nor spaces es: ACA-F-000000-0001

Dublin Core	Element Name: Source	
	Label:	Source
	Definition:	A Reference to a resource from which the present resource is derived.
	Comment:	The present resource may be derived from the Source resource in whole or in part. Recommended best practice is to identify the referenced resource by means of a string or number conforming to a formal identification system.

ALINARI	Element Name: URL Preview	
	Label:	URL Preview
	Definition:	A Reference to a resource from which the present resource is derived.
	Comment:	Needed to preview the content item. It represents the content (128x128 pixel max)

Dublin Core	Element Name: Language	
	Label:	Language
	Definition:	A language of the intellectual content of the resource.
	Comment:	Recommended best practice is to use RFC 3066 [RFC3066] which, in conjunction with ISO639 [ISO639]), defines two- and three-letter primary language tags with optional subtags. Examples include "en" or "eng" for English, "akk" for Akkadian", and "en-GB" for English used in the United Kingdom.

Dublin Core	Element Name: Relation	
	Label:	Relation
	Definition:	A reference to a related resource.
	Comment:	Recommended best practice is to identify the referenced resource by means of a string or number conforming to a formal identification system.

Dublin Core	Element Name: Coverage	
	Label:	Coverage
	Definition:	The extent or scope of the content of the resource.
	Comment:	Typically, Coverage will include spatial location (a place name or geographic coordinates), temporal period (a period label, date, or date range) or jurisdiction (such as a named administrative entity). Recommended best practice is to select a value from a controlled vocabulary (for example, the Thesaurus of Geographic Names [TGN]) and to use, where appropriate, named places or time periods in preference to numeric identifiers such as sets of coordinates or date ranges.

Dublin Core	Element Name: Rights	
	Label:	Rights Management
	Definition:	Information about rights held in and over the resource.
	Comment:	Typically, Rights will contain a rights management statement for the resource, or reference a service providing such information. Rights information often encompasses Intellectual Property Rights (IPR), Copyright, and various Property Rights. If the Rights element is absent, no assumptions may be made about any rights held in or over the resource.