ECP-2007-DILI-517005

ATHENA

Recommendations and best practice report regarding the application of standards, including recommendations for a harvesting format and fact sheets for dissemination

Deliverable number	*D3.2*
Dissemination level	*Public*
Delivery date	*31 July 2009*
Status	*Final*
Author(s)	*Gordon McKenna, Collections Trust (UK)* *Chris De Loof, Royal Museums of Art and History (BE)*

e*Content*plus

This project is funded under the eContentplus programme1,
a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable.

[1] OJ L 79, 24.3.2005, p. 1.

Table of Contents

1. Introduction

1.1 The purpose of work package 3

Work package 3 of the ATHENA project (WP3) is tasked with:

1. Reviewing the different standards in use by museums;
2. Facilitating the mapping of those standards to a common metadata standard;
3. Assessing the requirements for the persistent identification of digital objects and collections;
4. Producing tools to support the conversion of museums' data into the common harvesting format for ingestion into the main Europeana service.

WP3 also works together with other work packages in the project. In particular WP3 works closely with WP4 and WP7: feeding information about standards for their work. Also the survey which is the basis of this deliverable was extended to include collecting information on IPR issues for use within WP6.

1.2 Sources for the deliverable

This deliverable is the result of the work of Task 3.2 of the ATHENA project which was concerned with the "*identification of standards and common harvesting formats and the publication of recommendations*". To inform its content we have used:

1. Information from the standards survey of the:
   - ATHENA project (Deliverable D3.1);
   - EuropeanaLocal[2] project (Deliverables D2.1 and D2.2). This has been used to confirm the results of the ATHENA survey.

   These two pieces of work have been largely used to confirm the validity of the recommendations given.
2. The guidance found in the Minerva Project's *Technical Guidelines for Digital Cultural Content Creation Programmes*[3]. This is highly recommended to any organisation already carrying out a digitisation project and especially to those who are considering beginning one.

   Other sources of advice and guidance were consulted and useful material included.
3. The requirements of Europeana[4]. Most guidance was written before the existence of Europeana and does not take into account any special requirements needed for its implementation. Therefore we have included sections describing these in the appropriate places.

[2] EuropeanaLocal was originally called 'EDL Local' (Grant Agreement ECP-2007-DILI-517009) see: http://www.europeanalocal.eu

[3] See: http://www.minervaeurope.org/interoperability/technicalguidelines.htm for links to various versions.

[4] See *Technical documents* at: http://version1.europeana.eu/web/europeana-project/documents

1.3 Overview of the deliverable

Overall the aim of this deliverable is not to 'reinvent the wheel' by adding yet another set of guidelines for an organisation to look at. We do this because we believe that there is an existing set of guidelines available which do a very good job at meeting the needs of cultural organisations in Europe. This is the Minerva *Guidelines* mentioned above. The standards it recommends are those being used by the organisations who took part in the both surveys that we looked at. Their other advantages are:

- ***Multilingual*** – Having a guide in the working language of the organisation is a great advantage for their understanding and adoption. Further translation should be encouraged;
- ***Written for a general cultural sector audience*** – purely technical guidelines are a barrier to the general audience, which most people in the cultural sector are in this area;
- ***Updated*** – Is it very important that technical guidelines are kept up-to-date, especially in the rapidly changing IT area. It may be considered necessary to update the *Guidelines* with information from this document, for example those relating to Europeana.

In terms of content the deliverable is divided into two broad areas:

- Technical standards and guidance – these follow the recommendations of the *Minerva Guidelines* with minor changes and additions where needed;
- Metadata standards and guidance. This section also looks at the creation of a new metadata harvesting XML schema, *LIDO*, together with an outline description of it.

Finally there are the texts for two fact sheets at the end summarising the advice. These are with WP2 who will turn them into publication standard documents for distribution.

2. Technical Standards and Guidance

2.1 Overview

The first and most basic advice on technical standards is:

Use open standard formats when creating and delivering digital content.

Doing this will:

- Maximise access;
- Ensure that content is reusable. It can be created and changed by more than one piece of software;
- Avoid dependency on a single supplier with possible licensing restrictions.

Following this advice is easy for all of the common types of digital content. Therefore there must be a very good reason not to follow it. Even where a proprietary standard has been used it is good practice to have a plan to migrate to an open standard when one becomes available.

Detailed advice on which technical standards to use is based on the environment the material is being used in. In broad terms there are three '**use environments**':

Master
This is where the digital surrogate is created from an analogue original. Sometimes this is described as creating an archival master. This can be done by a number of techniques including: photography; scanning; sampling; OCR (optical character recognition), 3-D modelling, and so on. Born digital content will be archival by default.

The activity usually takes place at the collection holding organisation and in their systems.

Key concepts for digital content in this environment are:

- Maximum quality (e.g. no compression);
- Preservation;
- Open source.

Service
This is where users of the material are given **meaningful access** to a **single** piece of digital content. Delivery usually includes relevant metadata describing the significance of the material being accessed.

Keywords for digital content in this environment are:

- Usable quality (for service being offered);
- Reasonable speed of delivery;
- Rights protection (either by size restriction or other means).

Discovery
This is where users are given access to a set of thumbnails of digital content. The aim here is to review the results and move on to more detailed information. Delivery is usually part of the result set of a search and includes discovery metadata.

Keywords for digital content in this environment are:

- Maximum speed of delivery;
- Minimal size
- Recognisability.

The service and discovery often appear together on the part of an organisation's website where they display their collections online. Portals do the same when they aggregate and deliver digital content.

Portals can also only implement the discovery environment. They point to the service environment on the collections holder's website.

Europeana is unusual in that it really only implements the discovery environment but uses a link, if available, to **appear** to give access to content in the service environment.

The sections below give recommendations (file formats and quality) for the various common media types in the three use environments.

2.2 Text recommendations

Parameter	**Use Environment**	
	Master	***Service***
File Format	XML [preferred] PDF; DjVu [alternative]	XHTML; HTML [preferred] PDF; DjVu [alternative] ODF; RTF; Microsoft Word [supplementary]

The discovery use environment can be provided for by using an image of the text (see below)

2.3 Images recommendations

Parameter	**Use Environment**		
	Master	***Service***	***Discovery***
File Format	TIFF	JPEG; PNG	JPEG; PNG
Colour Quality	8 bit greyscale 24 bit colour	8 bit greyscale 24 bit colour	8 bit greyscale 24 bit colour
Resolution (dpi)	600 (photographs) 2400 (slides)	150-200	72
Maximum dimension (pixels)	[not applicable]	600	100-200

2.4 Audio recommendations

Parameter	Master Use Environment
File Format	Uncompressed [preferred]: WAV; AIFF Compressed [alternative]: MP3; WMA; RealAudio; AU
Creation quality	24-bit stereo and 48/96 KHz sample rate

Parameter	Service Use Environment
File Format	Compressed [preferred]: MP3; RealAudio; WMA Uncompressed [alternative]: WAV; AIFF; AU
Delivery quality	256 Kbps (near CD quality); 160 Kbps (good quality)

The discovery use environment may be provided by a relevant image (see below)

2.5 Video recommendations

Parameter	Master Use Environment
File Format	Uncompressed [preferred]: RAW AVI Compressed [alternative]: MPEG (MPEG-1, MPEG-2 or MPEG-4); WMF; ASF; Quicktime.
Quality	Frame size of 720x576 pixels; Frame rate of 25 frames per second; 24-bit colour; PAL colour encoding

Parameter	Service Use Environment
File Format – for downloading	MPEG-1; AVI; WMV; Quicktime
File Format – for streaming	ASF; WMV; Quicktime

The discovery use environment may be provided by a relevant still image from the video (see above)

2.6 Vector graphic recommendations

Parameter	Use Environment	
	Master	***Service***
File Format	SVG [preferred] SWF [alternative]	SVG [preferred]

The discovery use environment may be provided by a relevant image (see above)

2.7 Virtual reality recommendations

Parameter	Use Environment	
	Master	***Service***
File Format	X3D [preferred] QuickTime VR [alternative]	X3D [preferred] QuickTime VR [alternative]

The discovery use environment may be provided by a relevant image (see above)

A summary of these guidelines and recommendations can be found in the fact sheet at the end of the deliverable.

2.8 Note on guidelines for geographic location description and GIS

This area is dealt with in WP7 and is the subject of deliverable D7.2.

2.9 Examples of best practice

There are many sources for best practice for digitisation[5]. Here we give a selection of the most useful. They include websites, web pages, and PDFs. All are available on the Web:

Best Practice	Type of Materials Covered
BCR's CDP Digital Imaging Best Practices Working Group. [US] *BCR's CDP Digital Imaging Best Practices. Version 2.0.* Bibliographical Center for Research. June 2008. http://www.bcr.org/dps/cdp/best/digital-imaging-bp.pdf	***Image***: • Artwork/3-Dimensional Objects; • Film (into still images); • Graphics; • Maps; • Photos; • Text.
Canadian Heritage Information Network (CHIN). [CA] *Capture Your Collections: Planning and Implementing Digitization Projects*. (web pages) http://www.chin.gc.ca/English/Digital_Content/Capture_Collections/course_contents.html **Le Réseau canadien d'information sur le patrimoine (RCIP)**. [CA] Numérisez vos collections: Planification et mise en œuvre de projets de numérisation. (web pages) http://www.rcip.gc.ca/Francais/Contenu_Numerique/Numerisez_Collections/contenu_du_cours.html	General advice particularly for: ***Image***.
CDP Digital Audio Working Group. [US] *Digital Audio Best Practices. Version 2.1*. Colorado Digitization Program. October 2006. http://www.bcr.org/dps/cdp/best/digital-audio-bp.pdf	***Audio***: • Spoken language; • Oral history recordings; • Musical recordings.
Thomas C. Christensen and **Julia Welter**. [EU]	Digitization of film as:

[5] See http://www.minervaeurope.org/bestpractices/listgoodpract.htm

Guidelines for digitization, digital storage and retrieval. EFG – The European Film Gateway, 2009 http://www.europeanfilmgateway.eu/downloads/D4-3_Guidelines%20for%20digitization_20091030.pdf	***Image***; ***Video***.
Best Practice	**Type of Materials Covered**
DEN - Digitaal Erfgoed Nederland. [NL] *ICT-register voor het cultureel erfgoedl*. [ICT for cultural heritage register] (web pages) http://matrix.den.nl/matrix.aspx?matrixid=register&view=Digitaal_Erfgoed&start=&zoekterm=&f_status=DE+BASIS *Leidraad erfgoed digital*. [Digital heritage guide] (web pages) http://www.den.nl/publicaties/leidraad/inhoudsopgave	File formats for: ***Image***; **Video**; ***Audio***.
Kate Fernie, **Giuliana De Francesco** and **David Dawson** (Eds). [EU] *Technical Guidelines for Digital Cultural Content Creation Programmes: Version 2.0*. MINERVA eC Project, 2008. http://www.minervaeurope.org/publications/MINERVA%20TG%202.0.pdf (*Version1* is available in German, Italian, Dutch and French. *Version 2.0* will be available soon).	Guidelines for: ***Text;*** ***Image***; **Video**; ***Audio***.
JISC Digital Media. [UK] *Still images, moving images and sound advice* (website) http://www.jiscdigitalmedia.ac.uk: • http://www.jiscdigitalmedia.ac.uk/stillimages/ • http://www.jiscdigitalmedia.ac.uk/movingimages/ • http://www.jiscdigitalmedia.ac.uk/audio/ • http://www.jiscdigitalmedia.ac.uk/crossmedia/	Advice on digitising analogue material and creating new digital material for: ***Image***; **Video**; ***Audio***; ***Cross-media***.
Ministère de la culture et de la communication. [FR]	Includes the MINERVA guidelines

La numérisation des fonds patrimoniaux: informations techniques. [Digitization of cultural heritage: technical recommendations] http://www.culture.gouv.fr/culture/mrt/numerisation/index.html	so: ***Text;*** ***Image***; **Video**; ***Audio***.
Franz Pavuza. [EU] *Short guidelines for video digitisation*. TAPE (Training for Audiovisual Preservation in Europe), 2008. http://www.tape-online.net/Short_Guidelines_Video_Digitisation.pdf	Guidelines for: **Video** (tape).

Best Practice	Type of Materials Covered
Vassilios Tsioukas and **Miltiades Daniil**. [GR] '3D digitization of historical maps' in *e-Perimetron*, Vol. 4, No. 1, 2009. pp45-52 http://www.e-perimetron.org/Vol_4_1/Tsioukas_Daniil.pdf	Guidelines for: **Image** - maps

3. Metadata Standards and Guidance

3.1 Introduction

The first and most basic advice on metadata standards is:

Use standards for creating and delivering metadata.

Doing this will:

- Maximise interoperability between systems;
- Ensure that metadata is reusable. It can be created and used in more than one system;
- Avoid dependency on a single system supplier or a limited set of staff familiar with your system.

This is very similar to the advice given earlier in this document on the use technical standards. However when one examines the results of the survey undertaken by the ATHENA project the challenge that needs to be met to implement this advice becomes apparent. The survey highlighted the complex landscape for cultural metadata:

- Domain specific metadata, e.g. *SPECTRUM* (museums), *ISAD(G)* and *EAD* (archives), and *MARC* (libraries);
- Country specific metadata, e.g. in Italy (ICCD[6]);
- Organisational specific metadata created in-house, perhaps adapting standards;
- Application profiles of Dublin Core being used to deliver content on the Web in a cross domain portal, e.g. in Italy (PICO AP[7]) and UK (PNDS DCAP[8]).

This landscape hampers metadata interoperability. To simplify this situation perhaps we should look at the use environments for metadata.

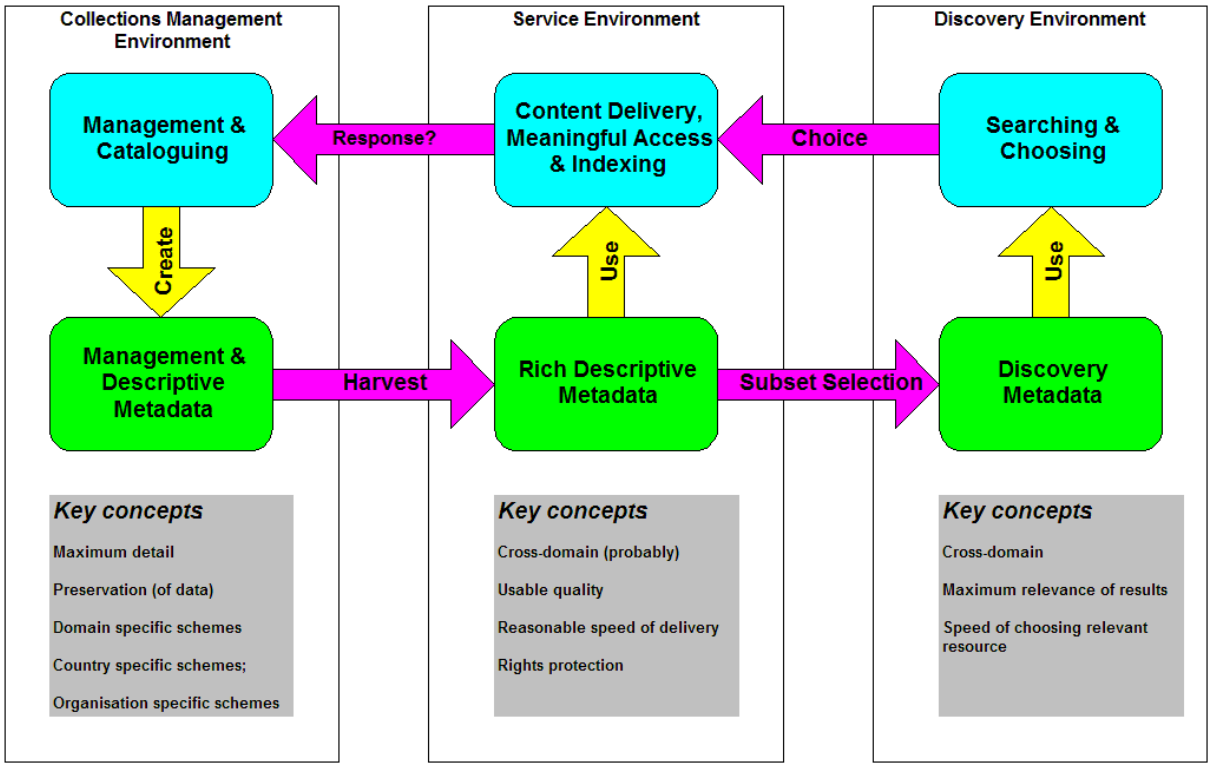
3.2 Metadata use environments

As with technical standards there are three use environments which can be summarised in the following diagram:

[6] See http://www.iccd.beniculturali.it/Catalogazione/standard-catalografici/metadati/metadati?set_language=it

[7] See http://purl.org/pico/picoap1.0.xml

[8] See http://www.ukoln.ac.uk/metadata/pns/pndsdcap



Collections management

This is where metadata is **created**. The information recorded comes from a number of sources:

- Collections management activities of the organisation (for example: acquisition; loans; conservation, rights management and use);
- Descriptions of the object itself (for example: type; title; material; dimensions; subject of intellectual and visual content);
- Connections to events during its existence (for example: creation; field collection; use and association);
- Connections to persons, organisations, and places during its existence (these are often intimately connected to the events mentioned above).

Usually takes place at the collection holding organisation, within their own systems, and with a lot of human effort.

Key concepts for metadata in this environment are:

- Maximum detail (all the relevant data);
- Preservation (of data);
- Domain specific schemes (museums, libraries and archives use different metadata schemes);
- Country specific schemes;
- Organisation specific schemes (these might be in-house or adaptations of standards).

Service

This is where users are given **meaningful access** to a **single** piece of metadata describing an object or other piece of cultural material. Delivery usually includes a digital surrogate for the material.

Key concepts for metadata in this environment are:

- Cross-domain (probably contains material from more than one)
- Usable quality (for service being offered – often audience specific);

- Reasonable speed of delivery;
- Rights protection (copyright statement or technical means).

Metadata here is a subset of the metadata in the collections management environment and should ideally be **harvested** from there.

This environment should also provide a means for collecting a user's response to the object which could feed back information into the collections management environment. For example additional information about the content of a photograph might be provided by the user of the service which was unknown to its owning organisation.

Discovery
This is where users are given access to a set of metadata from many objects. Delivery is usually part of the result set of a search together with a thumbnail of some kind. Users **choose** a content they want to look at in the service environment.

Key concepts for metadata in this environment are:

- Cross-domain;
- Maximum relevance of results;
- Speed of choosing relevant resource (limited set of metadata elements).

Metadata here is a subset of the metadata in the service or collections management environments.

The appearance of the service and discovery environments in an organisation's website, portals, aggregators and Europeana is the same as that for the digital content.

The major issue in this picture of use environments is – What is the metadata for the service environment? This issue led to a major element in the work of WP3 - The creation of a metadata XML schema for use in the service environment.

3.3 *LIDO* (Light Information Describing Objects)

3.3.1 The need for a rich harvesting schema

From the above it can be seen that the potentially rich metadata that is harvested from the collection management environment has the key role in providing a good service for users. The question that needs to be answered as a result of this analysis is: *Which metadata scheme should be used?*

Dublin Core (DC), in some version or other, is the commonly used metadata schema in both the service and discovery environment. However the work of the ATHENA project questions its use for museum content in particular.

There is a common view within the museum community that a DC derived metadata schemas do not deliver a rich enough view of museum content. The importance of a museum object, especially outside the area of fine art, is often not covered adequately. DC-based systems 'flatten out' museum metadata, with most of the data going into limited subset of elements. Taking as an example from *SPECTRUM* the date-related elements:

- Object production date;
- Field collection date;
- Content - date;
- Associated date;
- Associated event date.

The data from all these will end up in the same *date* element in a simple DC-based. There are similar effects for the 'who' 'what', 'where' classes of museum metadata elements. Also there is a loss of the relationships between the different classes and the events they relate to. So it becomes difficult to query the data in complex ways. Finally the ATHENA survey revealed the lack of standard DC-based metadata scheme, built into its design which allows for extensions. This is a barrier to interoperability.

All the above issues led the ATHENA project to seek to use, or develop, a metadata schema suitable for harvesting museum data into the service environment. The end result of the process[9] was the decision to create a schema based on the existing *museumdat* schema – *LIDO* (Light Information Describing Objects).

Such a metadata schema is especially useful when metadata is created from the cataloguing information. When the metadata is already in a DC-based schema (e.g. ESE) ATHENA might harvest such data directly.

3.3.2 Sources for LIDO

LIDO is not really a new schema. It builds on existing standards and best practise from a number of different countries in Europe and the rest of the world. Here we do not give full descriptions of these sources (see the footnotes for full information) but discuss them in the context of the development of *LIDO*.

CDWA Lite

CDWA Lite[10] was created by the J. Paul Getty Trust as an XML schema for harvesting data using the *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH). *CDWA Lite* is based on the *Categories for the Description of Works of Art* (CDWA)[11] which was designed to describe art, and other material culture related, collection databases in terms of 532 categories and subcategories. *CDWA Lite* is a 'core' of these.

One of the important features of *CDWA Lite* is the use of two types of element for display and indexing:

- ***Indexing elements*** containing data that is designed to facilitate machine retrieval from the portal's database;
- ***Display elements*** containing data in a form that is designed for a human to be able to use easily.

An example of using these is where the indexing element's data for a creator's name appears 'reversed': *Family Name, Forename(s)*. The corresponding display element data will have the 'normal' order and may also have additional biographical information such as birth and death dates.

CIDOC CRM

[9] The process of creating *LIDO* is described in D3.3 together with a technical description and the XML schema as annexes.

[10] See: http://www.getty.edu/research/conducting_research/standards/cdwa/cdwalite.html

[11] See: http://www.getty.edu/research/conducting_research/standards/cdwa/introduction.html

The *CIDOC Conceptual Reference Model* (CIDOC CRM)[12] is the result of over 10 years work by CIDOC Documentation Standards Working Group and CIDOC CRM SIG. It is also an ISO standard.

The CIDOC CRM is a formal standard that defines cultural heritage documentation concepts and the relationships between those concepts. Its purpose is to give those working in cultural heritage a flexible standard framework that any of their data can be mapped to. This means domain experts and those who are implementing information systems are able to create sensible requirements for those systems. It also gives a way to link together different sources of information in a meaningful way.

museumdat

museumdat[13] is the foundation for the new *LIDO* XML harvesting schema that was developed within the work of WP3 of the ATHENA project. It is an XML harvesting format that is optimised for the search, retrieval and publication of potentially rich museum metadata. The data is extracted is some way from an organisation's collections management system and published on a portal site. It was created by using both *CDWA Lite* and the *CIDOC CRM*[14].

CDWA Lite met most of the needs of a schema for use with German cultural portals. However it was decided that it needed to be reconfigured to allow the harvesting data about objects from wider areas than *CDWA Lite* supports; for example from social history and natural sciences. This led to the reduction of mandatory elements to:

- Object / Work Type;
- Title or Object Name;
- Record Element with ID and Source.

Also added were element attributes for:

- Multilingual support;
- Control of data conversion;
- Handling of controlled vocabularies.

museumdat was also designed so that it complied with the *CIDOC CRM*. The important aspect of the *CIDOC CRM* is its event-oriented approach. Using this led to the:

Descriptive elements to be grouped by:

- Classification;
- Identification;
- Description;
- Event;
- Relation.

Introduction of ***Event Set*** with groups of information for:

- Actor,
- Date;

---

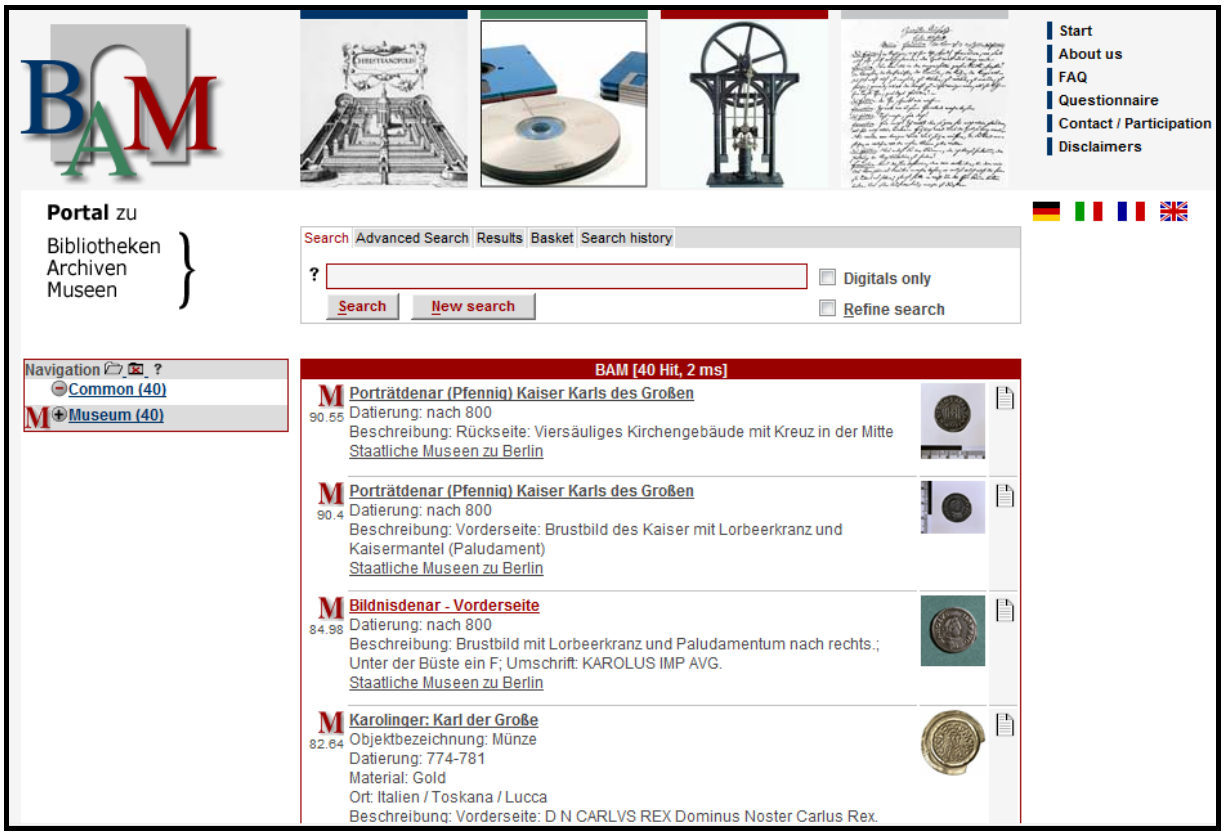
[12] See: http://cidoc.ics.forth.gr

[13] *museumdat* was created by the work of Fachgruppe Dokumentation des Deutschen Museumsbundes (Documentation Special Interest Group of the German Museums Association). Full details can be found at: http://www.museumdat.org/index.php?ln=en

[14] For a full description of this process see: http://museum.zib.de/museumdat/cdwalite_and_museumdat.pdf

- Location.

museumdat is successfully being used in the cross domain BAM Portal[15] in Germany. The next page has a screenshot:



SPECTRUM

SPECTRUM[16] is the final standard that was used in the creation of *LIDO*. *SPECTRUM* is the UK and international standard for Collections Management. It consists of two main sections:

- ***Procedures***;
- ***Information requirements***.

Procedures define best practice for collections management in terms of 21 different activities that commonly take place in collections holding organisations, for example: Object entry; Loans in; Acquisition; Location movement control; Cataloguing; Conservation and collections care; Rights management; Use of collections; Object exit; Loans out; and Deaccession and disposal.

Information requirements are those *Units of information* that need to be recorded and maintained in order to properly document the procedures and the collections themselves. Sets of related Units are brought together into *Information groups*. These enable the recording of: an object (both physically and with regard to events in its history); events that take place in the organisation (e.g. an audit); persons, organisations, peoples and places, etc. associated with objects and events. The Units are available as an XML schema[17].

[15] See: http://www.bam-portal.de

[16] See: http://www.collectionstrust.org.uk/specfaq. Download from: http://www.collectionstrust.org.uk/spectrum

[17] See: http://www.collectionstrust.org.uk/schema

In relationship to the creation of *LIDO SPECTRUM*'s role was to 'inform' *museumdat* with concepts from its Units of information. The result was significant change to *museumdat* and also was a factor to the change of name to LIDO. SPECTRUM's contribution was to allow:

- The information on all relevant entities to be recorded together with their relationships to each other - events, people, persons, organisations, places, locations, objects and subjects;
- The rights associated with content to be more fully described;
- Different language versions of the same data to be supported in a flexible way.

3.3.3 Outline of LIDO

Full details of *LIDO* can be found in the technical description and XML schema documentation found as annexes to the deliverable D3.3. Here we give an outline of *LIDO*.

The first point to make about *LIDO* is that it is a **harvesting schema**. It should not be used as a basis for a collection management system. It is for delivering metadata for use in the service environment of an organisation's online collections database, portals, and aggregations, including Europeana itself. In particular it does not support such activities as loans and acquisition. Its strength lies with its ability to support the full range of descriptive information about museum objects.

LIDO is made up of a nested set of 'wrapper' and 'set' elements which structure records in culturally significant ways. An important part its design is concept of events, taken from the CIDOC CRM. So, for example, the creation, collection, and use of an object are defined as events which have associated entities such as date, places and actors. These can then all be represented in a consistent way.

The structural elements of *LIDO* contain 'data elements' which hold the information that is being harvested and ultimately delivered to the user of the service environment.

LIDO also allows for the recording of information about the sources for data (e.g. in a book) and controlled terminology (e.g. the identification code for a term in a thesaurus).

Conceptually there are 7 areas in a *LIDO* record for an object:

LIDO Area	**Information recorded**
Object Identification	Basic information about the object: • **Title (or object name if no title) [mandatory]**; • Inscriptions – transcript and/or description; • Repository – the organisation that holds the physical object and its identifier; • Display and edition information – especially for prints; • Description – descriptive text; • Measurements.
Object Classification	Information about the type of object: • **Object name [mandatory]**; • Other classification terms for the object – e.g. form, age, sex, status and phase.
Relation	Relations of the object to:

	• Its subject (content or visual) – concepts, actors, events, dates, places, events, and objects; • Other objects.
Event	Events that the object has taken part in, such as: • Creation; • Field collection; • Acquisition; • Exhibition; • Use, etc. For each event, information, if relevant, about: • Event ID; • Event type; • Object's role in the event; • Event name; • Actors (persons and organisations); • Cultures involved; • Dates (or periods); • Places; • Event method; • Materials and techniques used; • Other objects present at the event; • Related events; • Description of the event.
Rights Work	Information about the rights associated with the object, metadata and the digital surrogate being harvested into the service environment (especially copyright): • Rights type; • Rights holder; • Rights dates; • Credit line.
Record	Basic information about the record: • **ID [mandatory]**; • Type; • Source; • Metadata about the record.
Resource	Information about digital resource being supplied to the service environment (e.g. Europeana): • Link – URL of the resource; • Resource ID; • Relationship type – e.g. conservation, historical, reconstruction; • Resource type – its medium (e.g. x-ray); • Resource rights – of the resource where different from the

	object; • View description; • View type – vantage point of the resource; • View date; • Resource source – if not from the holding organisation; • Related resources; • Resource metadata location – pointer to other information about the resource.

3.4 Metadata standard recommendations for use environments

This section gives recommendations for metadata standards for each use environment. It also notes any issues relevant to the ATHENA system and Europeana in general. The recommendations are:

3.4.1 Collections management

The choice of a collections management standard depends on the domain that your collection belongs to:

Domain	Recommended standard
Museums	*SPECTRUM*
Libraries	*MARC*
Archives	*ISAD(G)*; *EAD*

If you have an in-house system you should be able to map your metadata elements to the metadata elements of these standards.

You should be able to export from the rich metadata in your collections management system into the schemas given in the next section.

If your system is not as comprehensive as one of the recommended standards you might not be able to submit rich metadata to the ATHENA aggregation.

3.4.2 Service

This use environment is where the ATHENA system resides. Its aim is to give, potentially, enhanced functionality to Europeana by collecting richer metadata. The choice of which standard depends on richness of the metadata that is in the system supplying data to the ATHENA aggregator:

Richness of metadata	Recommended standard
High	*LIDO*
Low	*ESE*

Partners supplying content will have make a decision on which of the above they will be able export their data into. For example museums with simple data, perhaps as a spreadsheet, may

choose to go directly to *ESE*. However if museums do not use *LIDO* then they will not be able to offer a richer service and discovery environment to their users.

3.4.3 Discovery

The metadata here will either be a subset of *LIDO* or *ESE*. In the former case the ATHENA system will populate the subset automatically for ingestion by Europeana.

3.4.4 Europeana requirements

The proper functioning of the Europeana portal depends on individual content suppliers and aggregators serving up their data in the *ESE* format. *ESE* is based on the Dublin Core format. This is relatively easy to map to from a collections holder's system via *LIDO*. However *ESE* has some elements that refine and extend simple DC which it is important to populate with data. Therefore individual content suppliers must supply this data in order to successfully submit their content for ingestion by Europeana.

The relevant *ESE* elements are:

Element	***Definition and notes***	***Data requirements***
isShownBy	An unambiguous URL reference to the digital object on the content provider's web site in the best available resolution/quality. (i.e. a link to the content as a text, image, sound, or video file **not** to the webpage with it on) Data here will allow the full functionality of Europeana and the automatic generation of a thumbnail by them. If this cannot be given then you **must** provide data for **isShownAt**.	Must be valid URI (e.g. URL)
isShownAt	An unambiguous URL reference to the digital object on the content provider's website in its full information context. If this cannot be given then you **must** provide data for **isShownBy**.	Must be valid URI (e.g. URL)
object	For image thumbnails, if you can give a URL to a thumbnail on your website then give that URL here. However these thumbnails smaller than 110 pixels high then it will be scaled up to that size by Europeana. If you do not have a thumbnail then you may give the same data as in **isShownby** element.	Must be valid URL
type	The Europeana material type of the resource.	Must be: TEXT or IMAGE or SOUND or VIDEO
provider	Name of the organisation that is delivering content to Europeana. If the **provider is not an aggregator** then use this	

	element. If the **provider is an aggregator** then use this element for the name of the aggregator. For the name of the **provider to the aggregator** use the element **source**.	

3.5 Modifying metadata standards

The following is strongly recommended. For the advantages of interoperability to work:

Do NOT adapt a published metadata standard during the creation of an in-house system.

There situations where an organisation needs to resist the temptation to change ('adapt') the standard:

- ***Element names do not follow in-house practice***
  This is when the standard does not use the same names for elements that an organisation is familiar with. A good standard will be aware of this and will give a set of 'non-preferred' names for its elements.
  For example *SPECTRUM* has a Unit of information *Object production person* with non-preferred names of: *Artist*; *Maker*; *Manufacturer*; and *Moneyer*. This list does not cover all non-preferred names but gives the user of the standard enough information to know the reasoning behind the standard name. In this case the standard name is non-specific.
- ***Elements are missing from the standard that is needed to record information***
  One source of this is similar to the last. Again it might be a case of the 'missing' element is 'hiding' under another name. Check with the publisher of the standard to confirm that you have really identified a missing element. They should be grateful either to help you with your implementation and your possible input to the development of the standard. Recently *SPECTRUM* has added Units which allow organisations to record information from past owners and viewers of objects.
- ***The standard is thought to be too large and complex***
  Small organisations are overwhelmed sometimes by a standard. They think that there must be something 'easier'. This is understandable.
  Here the organisation should consult with the standard publisher to see if there are any beginner's guides, training, or other support available. Also commercially available systems, and also increasingly open source systems, are on the market which will help with implementation of a standard.

The aim of the ATHENA best practice network is to create a stable network of expertise that can be beneficial for the sector and museums in particular. Therefore the network should act as an information point on standards. An organisation with a question about a standard will either find an answer supplied by the network or will be pointed in the correct direction to find an answer.

Examples of such services from the UK are:

JISC Digital Media[18] (formerly known as TASI), a technical service giving advice, guidance and training on:

- Creating digital content - images, video and audio;

[18] See http://www.jiscdigitalmedia.ac.uk

- Delivering content to users;
- Using Content for teaching, learning and research;
- Managing digitisation projects.

Collections Link[19] is a partnership over 20 organisations providing advice and support to museums, archives, libraries and other collections-holding organisations. The aim of the service is to provide a single point of access to best practice in the care and management of collections. It has three main parts:

- An online library of best practice guides and factsheets;
- A national database of training and skills development opportunities;
- A commissioning fund to support the development of new resources

Collections Link is managed by the Collections Trust, an ATHENA partner.

The success of such pan-European a service may provide a basis for sustainability after the end of the EC-funded project.

3.6 Creating rich metadata for the service environment

3.6.1 Using SPECTRUM as a framework

In order to create rich metadata for the service environment we ***recommend*** that you use *SPECTRUM*[20], or other rich national standards, as a framework to identify the types of information you should be recording. In *SPECTRUM* these types are called 'Units of information'[21].

To help organise the units we have arranged the relevant subset of units in two ways. Firstly by entities:

- ***What***
- ***When***
- ***Where***
- ***Who***

Secondly by events or processes in the 'lifetime' of the object:

- ***Association*** – an event (for example use) in the physical object's or group of objects' history (other than field collection);
- ***Collection*** – Important particularly for objects collected by archaeological excavation and natural science specimens;
- ***Creation***;
- ***Description*** – Information about the object including: physical characteristics, and content for visual works;
- ***Identification*** – Basic information about the object including permanent identifiers for physical and digital objects;
- ***Location*** – Where the physical object is in the collection holding organisation (could be on loan).

[19] See http://www.collectionslink.org.uk

[20] To download the English language for the UK, see: http://www.collectionstrust.org.uk/spectrum. There are also Dutch versions for Belgium and the Netherlands. Other language versions are in preparation.

[21] The mapping between *SPECTRUM* and *LIDO* in the ATHENA system appears as an appendix to deliverable D3.3.

The tables below list the relevant *SPECTRUM* Units arranged in this way. A definition is included. See *SPECTRUM* itself for fuller information including examples.

Who

Event	Unit of information	Definition
Creation	Object production organisation	An organisation involved in the design, creation or manufacture of the object.
	Object production people[22]	A people involved in the design, creation or manufacture of an object.
	Object production person	A person involved in the design, creation or manufacture of an object. This may include the commissioner of an object.
Collection	Field collector	The person or organisation responsible for collecting a specimen or object in the field.
Description	Content – organisation	The organisation depicted in or described in an object.
	Content – people	A people depicted in or described by an object.
	Content – person	A person depicted in or described by an object.

[22] In *SPECTRUM* a 'people' is a cultural group of some type, e.g. the Maori.

Event	Unit of information	Definition
Association	Associated event organisation	An organisation associated with an event in an object's or group of objects' history (other than field collection or ownership).
	Associated event people	A people associated with an event in an object's or group of objects' history (other than field collection or ownership).
	Associated event person	A person associated with an event in an object's or group of objects' history (other than field collection or ownership).
	Associated organisation	An organisation associated with an object's or group of objects' history.
	Associated people	A people associated with an object's or group of objects' history.
	Associated person	A person associated with an object's or group of objects' history.
	Owner	Details of a people, person or organisation who owned an object before title was transferred to the organisation.

What

Event	Unit of information	Definition
Creation	Technique	Processes, methods, techniques or tools used to fabricate or decorate an object.
Collection	Field collection event name	The name of an event at which an object was collected.
	Geological complex name	The name of a geological complex from which a geological specimen was collected.
	Habitat	A term describing the surroundings and environment of the area where a specimen was collected in the field.
	Stratigraphic unit name	The stratigraphic unit from which a field collection was made.
Identification	Object number	A unique number identifying an object or specimens, including any separated parts.

	Object name	A description of the form, function or type of object.
	Title	The name assigned to an object or group of objects by the artist/creator or collector at the time of origin or subsequent titles either specifically assigned or generally understood to refer to it.
Description	Age	The numeric age of a natural science specimen when it died. Use Phase for a textual description of Age.
	Colour	The colour of an object.
	Content – activity	An activity depicted in or described by an object.
	Content – concept	A concept depicted in or described by an object.
	Content – event name	An event depicted in or described by an object.
	Content – object	An object depicted in or described by another object.
	Dimension	The method used to mount or preserve a specimen.
	Dimension value	The numeric value of the measurement of a Dimension.
	Form	The method used to mount or preserve a specimen.
	Material	The basic materials and media from which an object is constructed.
	Object status	A statement of the standing of a natural science specimen or other object in relation to others in existence.

Event	Unit of information	Definition
Description	Phase	A textual expression of the age or developmental phase of a natural science specimen.
	Sex	The gender of an animal specimen.
	Style	Styles or schools relating to an object.
	Technical attribute	The name of a technical attribute possessed by an object which can be described and quantified.
	Technical attribute measurement	The measurement of a named Technical attribute.
Association	Associated activity	An activity associated an object or group of objects.
	Associated concept	A concept associated with an object or group of objects.
	Associated cultural affinity	A wider cultural context to which an object or group of objects relates.
	Associated event name	An historical event associated with an object or group of objects, not including production and collections management events.
	Associated object	An object associated with an object or group of objects.
	Usage	A single term describing the use of a particular kind of object.

When

Event	Unit of information	Definition
Creation	Object production date	The date when a stage in the design, creation or manufacture of an object took place.
Collection	Field collection date	The date an object is collected in the field.
Description	Content - date	A date depicted in or described by an object.
Association	Associated date	A date associated with an object or group of objects.
	Associated event date	The date of an event in an object's history.

Where

Event	Unit of information	Definition
Creation	Object production place	A place where the design, creation or manufacture of an object took place.
Collection	Field collection place	The place where an object was excavated or collected in the field.
Location	Current location	The place within the organisation where an object is currently located.
Description	Content - place	A place depicted in or described by an object.
Association	Associated event place	A place associated with an event in an object's history.
	Associated place	A place associated with an object or group of objects.
	Ownership place	The place where an object was owned before title was transferred to the organisation.

3.6.2 Making metadata content interoperable

We ***recommend*** that you use these best practises to increase the interoperability of your metadata in an aggregated service use environment:

- ***Use standard terminologies***
  Where a metadata element should contain content based on a standard set of terms. Use such a set of terms. Preferably these should be a widely used and published set of terms. However an in-house set is better than no terminology control. Do not worry if they are 'wrong' consistency is the goal. See the work of WP4 for more details.
- ***Make descriptions independent of the service***
  Often there is a temptation to leave out information in descriptions that is 'obvious' in the context it was created. If a description is a major source for creation of search indexes for a service that aggregates metadata and content then the missing information cannot be inferred. For example photographs of a city with only street names will not allow hits to searches on the city from the description alone. Therefore include all relevant information in a description.
- ***Do not include negative information in descriptions***
  Only include positive information in a description. The system extracting index search terms from a description is unlikely to be able to understand that fact is 'not something'.

4. Fact Sheet – Technical Standards for Digitising Cultural Content in Museums

Basic advice

Use open standard formats when creating and delivering digital content.

Doing this will ensure the interoperability of your content and avoid you becoming dependent on proprietary systems.

Detailed advice

Detailed advice on the use technical standards is based on the environment the material is being used in. In broad terms there are three use environments:

- ***Master*** **–** Where the digital surrogate is created from an analogue original. Sometimes this is described as creating an archival master. This can be done by a number of techniques including: photography; scanning; sampling; OCR (optical character recognition), 3-D modelling, and so on. Born digital content will be archival.
- ***Service*** **–** Where users of the material are given **meaningful access** to a **single** piece of digital content. Delivery usually includes metadata describing the significance of the material being accessed.
- ***Discovery*** **–** Where users are given access to a set of digital content. The aim here is to review the results and move on to more detailed information. Delivery is usually part of the result set of a search and includes discovery metadata

Text recommendations

Parameter	**Use Environment**	
	Master	***Service***
File Format	XML [preferred] PDF; DjVu [alternative]	XHTML; HTML [preferred] PDF; DjVu [alternative] ODF; RTF; Microsoft Word [supplementary]

Image recommendations

Parameter	**Use Environment**		
	Master	***Service***	***Discovery***
File Format	TIFF	JPEG; PNG	JPEG; PNG
Colour Quality	8 bit greyscale 24 bit colour	8 bit greyscale 24 bit colour	8 bit greyscale 24 bit colour
Resolution (dpi)	600 (photographs) 2400 (slides)	150-200	72
Maximum dimension (pixels)	[not applicable]	600	100-200

Audio recommendations

Parameter	Master Use Environment
File Format	Uncompressed [preferred]: WAV; AIFF Compressed [alternative]: MP3; WMA; RealAudio; AU
Creation quality	24-bit stereo and 48/96 KHz sample rate

Parameter	Service Use Environment
File Format	Compressed [preferred]: MP3; RealAudio; WMA Uncompressed [alternative]: WAV; AIFF; AU
Delivery quality	256 Kbps (near CD quality); 160 Kbps (good quality)

The discovery use environment may be provided by a relevant image (see below)

Video recommendations

Parameter	Master Use Environment
File Format	Uncompressed [preferred]: RAW AVI Compressed [alternative]: MPEG (MPEG-1, MPEG-2 or MPEG-4); WMF; ASF; Quicktime.
Quality	Frame size of 720x576 pixels; Frame rate of 25 frames per second; 24-bit colour; PAL colour encoding

Parameter	Service Use Environment
File Format – for downloading	MPEG-1; AVI; WMV; Quicktime
File Format – for streaming	ASF; WMV; Quicktime

The discovery use environment may be provided by a relevant still image from the video (see above)

Vector graphic recommendations

Parameter	Use Environment	
	Master	***Service***
File Format	SVG [preferred] SWF [alternative]	SVG [preferred]

The discovery use environment may be provided by a relevant image (see above)

Virtual reality recommendations

Parameter	**Use Environment**	
	Master	***Service***
File Format	X3D [preferred] QuickTime VR [alternative]	X3D [preferred] QuickTime VR [alternative]

The discovery use environment may be provided by a relevant image (see above)

Further information

This fact sheet is a short summary of the recommendations that can be found in the Minerva Project's:

Technical Guidelines for Digital Cultural Content Creation Programmes

See: **http://www.minervaeurope.org/interoperability/technicalguidelines.htm**

This document is highly recommended to any organisation already carrying out a digitization project and especially to those who are considering beginning one.

5. Fact Sheet – Cultural Metadata Standards in Museums

Basic advice

Use standards for creating and delivering metadata.

Doing this will:

- Maximise interoperability between systems;
- Ensure that content is reusable. It can be created and used in more than one system;
- Avoid dependency on a single system supplier or a limited set of staff familiar with your system.

Metadata use environments

Detailed advice on the use metadata standards is based on the environment the material is being used in. In broad terms there are three use environments:

Collections management – Where metadata is created. The information recorded comes from a number of sources:

- Collections management activities of the organisation (for example: acquisition; loans; conservation, rights management and use);
- Descriptions of the thing itself (for example: type; title; material; dimensions; subject of intellectual and visual content);
- Connections to events during its existence (for example: creation; field collection; use and association);
- Connections to persons, organisations, and places during its existence (these are often intimately connected to the events mentioned above).

Activity usually takes place at the collection holding organisation, in their systems, and with a lot of human effort.

Service – Where users are given **meaningful access** to a **single** piece of metadata describing an object or other piece of cultural material. Delivery usually includes a digital surrogate for the material. Metadata is a subset of the metadata in the collections management environment.

Discovery – Where users are given access to a set of pieces of metadata. Delivery is usually part of the result set of a search together with a thumbnail of some kind. Metadata is a subset of the metadata in the service or collections management environments.

Collection management standards

The choice of a collections management standard depends on the domain that your collection belongs to:

Domain	Recommended standard
Museums	*SPECTRUM*
Libraries	MARC
Archives	ISAD(G); EAD

If you have an in-house you should map your metadata elements to the metadata elements of these standards.

You should be able to export from the rich metadata in your collections management system into the schemas given in the next section.

Service standards
The choice of which standard depends on richness of the metadata that is in the system supplying data to the ATHENA aggregator:

Richness of metadata	Recommended standard
High	*LIDO*
Low	*ESE*

Partners supplying content will have make a decision on which of the above they will be able export their data into. For example partners with simple data, perhaps as a spreadsheet may choose to go directly to *ESE*.

Service standards
The metadata here will either be a subset of *LIDO* or *ESE*. In the former case the system will populate the subset automatically for ingestion.

Europeana requirements
The proper functioning of the Europeana portal depends on individual content suppliers and aggregators serving up their data in the *ESE* format. *ESE* is based on the Dublin Core format. This is relatively easy to map to from a collections holder's system via *LIDO*. However *ESE* has some elements that refine and extend simple DC which it is important to populate with data. Therefore partners will have this data available in their own systems.

The relevant 'europeana' *ESE* elements are:

Element	*Definition and notes*	*Data requirements*
isShownBy	An unambiguous URL reference to the digital object on the content provider's web site in the best available resolution/quality. (i.e. a link to the content as a text, image, sound, or video file **not** to the webpage with it on) Data here will allow the full functionality of Europeana and the automatic generation of a thumbnail by them. If this cannot be given then you **must** provide data for **isShownAt**.	Must be valid URI (e.g. URL)
isShownAt	An unambiguous URL reference to the digital object on the content provider's website in its full information context. If this cannot be given then you **must** provide data for **isShownBy**.	Must be valid URI (e.g. URL)
object	For image thumbnails, if you can give a URL to a thumbnail on your website then give that URL here. However these thumbnails smaller than 110 pixels high then it will be scaled up to that size by Europeana. If you do not have a thumbnail then you may give the same data as in **isShownby** element.	Must be valid URL

type	The Europeana material type of the resource.	Must be: TEXT or IMAGE or SOUND or VIDEO
provider	Name of the organisation that is delivering content to Europeana. If the **provider is not an aggregator** then use this element. If the **provider is an aggregator** then use this element for the name of the aggregator. For the name of the **provider to the aggregator** use the element **source**.	

Modifying metadata standards

The following is strongly recommended. For the advantages of interoperability to work:

Do NOT adapt a published metadata standard during the creation of an in-house system.

There situations where an organisation needs to resist the temptation to change ('adapt') the standard:

- ***Element's name does not follow in-house practice***
  This is when the standard does not use the same names for elements that an organisation is familiar with. A good standard will be aware of this and will give a set of 'non-preferred' names for its elements
- ***Elements are missing from the standard that is needed to record information***
  One source of this is similar to the last. Again it might be a case of the 'missing' element is 'hiding' under another name. Check with the publisher of the standard to confirm that you have really identified a missing element. They should be grateful either to help you with your implementation and your possible input to the development of the standard.
- ***The standard is thought to be too large and complex***
  Small organisations are overwhelmed sometimes by a standard. They think that there must be something 'easier'. This is understandable. Here the organisation should consult with the standard publisher to see if there are any beginner's guides, training, or other support available. Also commercially available systems, and also increasingly open source systems, are on the market which will help with implementation of a standard

Creating rich metadata for the service environment

Using SPECTRUM to provide a framework

In order to create rich metadata for the service environment in the *LIDO* schema we ***recommend*** that you use the *SPECTRUM* standard[23], or other rich national standards, to form a framework for the types of information you record. In *SPECTRUM* these types are called *Units of information*.

Look at the relevant subset of units in two ways. Firstly by entities:

- ***What***;
- ***When***;
- ***Where***;
- ***Who***.

Secondly by events or processes in the 'lifetime' of the object:

- ***Association*** – an event in an object's or group of objects' history (other than field collection);
- ***Collection*** – Important for objects collected by archaeological excavation and natural science specimens;
- ***Creation***;
- ***Description*** – Information about the object including: physical characteristics, and content for visual works;
- ***Identification*** – Basic information about the object;
- ***Location*** – Where the physical object is in the collection holding organisation (could be on loan).

Making metadata content interoperable

We ***recommend*** that you use these best practises to increase the interoperability of your metadata in an aggregated service use environment:

- ***Use standard terminologies***
  Where a metadata element should contain content based on a standard set of terms. Use such a set of terms. Preferably these should be a widely used and published set of terms. However an in-house set is better than no terminology control. Do not worry if they are 'wrong' consistency is the goal.
- ***Make descriptions independent of the service***
  Often there is a temptation to leave out information in descriptions that is 'obvious' in the context it was created. If a description is a major source for creation of search indexes for a service that aggregates metadata and content then the missing information cannot be inferred. For example photographs of a city with only street names will not allow hits to searches on the city from the description alone. Therefore include all relevant information in a description.
- ***Do not include negative information in descriptions***
  Only include positive information in a description. The system extracting index search terms from a description is unlikely to be able to understand that fact is 'not something'.

[23] To download the English language version for the UK, see: http://www.collectionstrust.org.uk/spectrum. There are also Dutch versions for Belgium and the Netherlands. Other language versions are in preparation.

Further information

For further information on the organisational and project context in which metadata resides see Minerva Project's:

Technical Guidelines for Digital Cultural Content Creation Programmes

See: **http://www.minervaeurope.org/interoperability/technicalguidelines.htm**

This document is highly recommended to any organisation already carrying out a digitisation project and especially to those who are considering beginning one.