



38189

MILE

Appendix 1:

Image Metadata: The Way Forward – A Simple Guide to Picture Metadata

Deliverable number/name	<i>D1.2.3 MILE Final Report: Appendix 1</i>
Dissemination level	<i>Confidential</i>
Delivery date	<i>30 September 2009</i>
Status	<i>Final</i>
Author(s)	<i>The MILE Project, Stella Dextre Clarke, Dr. Richard Butterworth</i>



eContentplus

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.



Image Metadata: the Way Forward

A simple guide to picture metadata

The MILE Project,
Stella Dextre Clarke & Dr. Richard Butterworth



Co-funded by the European
Community eContentplus
programme





Table of Contents

Why metadata?	3
Why so many standards?.....	4
Key list: metadata/ vocabulary: standards for image libraries.....	6
Developing a metadata schema: Top Tips.....	8
Simple guide to the published schemas.....	9
Applying metadata: Top Tips.....	11
Mapping and crosswalks	13
Detailed metadata element and group descriptions	15
Metadata crosswalk example.....	40
Approaches to multilingual provision.....	45
Developing and applying controlled vocabularies: Top Tips.....	46

1. Why metadata?

How to make the case for investing in good metadata practices:

- ✓ helps you manage workflow economically and effectively
- ✓ speeds up the import of data from your suppliers, and reduces errors
- ✓ helps you to export your data in a format that complies with your customers' needs
- ✓ enables you to reutilise the same data/images for multiple purposes
- ✓ helps your data to reach potential users all over the world
- ✓ improves the quality and consistency of your data
- ✓ helps you provide multilingual access to your collections
- ✓ enables sophisticated search techniques, with subtle distinctions where needed
- ✓ promotes interoperability over the Internet and other networks
- ✓ helps you comply with copyright laws
- ✓ helps with management and preservation of records

To achieve all the above, you must plan carefully and exploit standards when possible.

2. Why so many standards?

We have always found it helpful to standardize practices and conventions in our cataloguing of museum pieces, photographs, etc., so that people can make sense of the display and keep our records in order. But in a networked world, computers rather than humans handle most of the transactions, and they need standards much more than we do.

Data for computer handling need to be presented exactly the same way every time, and different standards may be added at each stage in the process. MILE has not dealt with underlying codes and protocols such as Unicode, HTTP, HTML and XML, that underpin networked communications in all sectors. The focus is on those needed specifically for the metadata of images, works of art and other items of cultural heritage held in galleries and museums. Even for these, a huge number of different standards apply. The communication chain does require several, that is to say, combinations of standards that work together harmoniously. But try the wrong combination and you'll find conflicts which impede communication.

So why not select just one standard for each component of the communication chain, and ensure that each of these is compatible with the next? The simple explanation is that needs are not the same across the sector. More specifically:

- Some image libraries depict fine art, needing metadata of interest to art historians and curators, while others show botanical specimens (or medical cases, or oceanographic pictures, etc) with metadata needed by scientists, and others have simply stock photography for the media and general public.
- As well as needing different metadata schemas, the controlled vocabularies to be applied with the schemas need to vary from one sector to another. For example, medical images need terms from MeSH (Medical Subject Headings); art images need artist names from ULAN (Union List of Artist Names); and many libraries need a vocabulary specifically designed to cover the breadth and depth of their own collection. Because the terms in everyday language have so many ambiguities, and there are so many different ways of expressing the same concept, the controlled languages are typically incompatible with each other.
- The sector cannot operate in isolation. Images are often delivered embedded in text and accompanied by other materials. Across the Internet we connect with the multimedia world. Image libraries must be flexible in providing data in formats acceptable to users and systems from other sectors.
- However logical and convenient a new standard may seem, the reality is that existing collections have mostly been catalogued using older standards, some of

them homegrown. Retrieving items from the legacy collections cannot be neglected as new practices come in.

- Internally, every organisation has the idiosyncrasies of its own history, customer base and suppliers to accommodate.

Given all this variety, it is unrealistic to hope that everyone will conform to exactly the same schemas and/or vocabularies. The best hope of achieving interoperability across the sector relies on a twin strategy:

- encourage all image libraries, museums, galleries etc to be consistent in applying the standards they choose;
- develop crosswalks that enable mapping between schemas and vocabularies.

3. Key list: metadata/vocabulary standards for image libraries

Guides to practice

SPECTRUM: The UK Museum Documentation Standard
PAS 197: Code of practice for cultural collections management
CCO (Cataloging Cultural Objects)
AACR2 (Anglo-American Cataloguing Rules)
RDA (Resource Description and Access)

Metadata schemas

Dublin Core
IPTC
VRA Core
CDWA/CDWA Lite
IEEE-LOM
ESE (Europeana Semantic Elements)
museumdat
METS (Metadata Encoding and Transmission Standard)
PLUS (Picture Licensing Universal System)

Metasearch /harvesting protocols

OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting)
SRU (Search and Retrieval via URL)
SPARQL (Simple Protocol and RDF Query Language)
SKOS-API (Simple Knowledge Organization Systems – application programming interface)
ADL thesaurus protocol (Alexandria Digital Library)

Standards for controlled vocabularies

ISO 2788:1986 (for monolingual thesauri)
ISO 5964:1985 (for multilingual thesauri)
BS 8723 (in 5 parts, published 2005/2007, monolingual + multilingual)
ISO 25964 (in 2 parts, expected 2010+)
ANSI/NISO Z39.19 (2005, monolingual only)
IFLA Guidelines for multilingual thesauri (2009)

Vocabulary exchange formats

SKOS (Simple Knowledge Organization Systems)
DD 8723-5
Zthes



Controlled vocabularies designed for art/image collections

Iconclass

AAT (Art & Architecture Thesaurus)

ULAN (Union List of Artist Names)

TGN (Thesaurus of Geographic Names)

TGM (Thesaurus of Graphic Materials)

Reference models

CIDOC-CRM (Conceptual Reference Model)

FRBR (Functional Requirements for Bibliographic Records)

4. *Developing a metadata schema: Top Tips*

- Plan to use one master metadata schema throughout your internal processes, with mappings to and from other (external) schemas, which will be used at the output and input ends of your workflow, respectively.
- Do not expect to find a ready-made schema that meets your internal needs in full. But do look for one that comes close.
- Do not reinvent the wheel; base your internal schema on the existing published schema closest to your needs. You can adapt this by adding some elements and omitting others.
- Try to use exactly the same name and definition of each element as in the published schema or you will end up confusing everybody.
- If possible, set your schema up as an “application profile” of an existing published schema. The general approach and methodology for application profiles is well explained in the Dublin Core literature and you can copy this approach, even if you do not choose DC as your base schema.
- Developing the application profile may be easier if you draw on just one published schema, but optionally you can draw on two or more.
- In your schema you can exploit the “EXIF” (Extensible Image File Format) metadata recorded automatically for each photograph by almost all digital cameras
- Enlist the support of a champion at top management level.
- Involve representatives of all key interests: users; sales team; cataloguers; IT, etc.
- Take into account the requirements of the customers and collaborators to whom you supply data. This should include an understanding of the search preferences of users.
- Study the foibles of the data to be imported from your suppliers.
- Know your own in-house needs.



- Go for granularity, i.e. subdivide each element into the smallest sub-elements you may need. (You can always dumb down by merging elements later; but you cannot so easily “dumb-up”!)
- Don’t expect too much of your cataloguers. If they can’t tell the difference between a geranium and a pelargonium, don’t expect them to make that distinction reliably in the metadata!
- Provide for more than one level of description within a single record. Very often the metadata associated with a photograph (e.g. date, IPR, name of photographer, etc.) differs from the metadata applicable to the object(s) shown in the photograph. This is a common problem with pictures of museum pieces or other works of art, and with composite resources such as a newspaper article or an educational package. The schema may need to “nest” the metadata for several objects within that of the image as a whole.
- Provide for user-generated metadata (such as bookmarks or social tagging) to be added at a later stage. Usually separate fields are needed for such data.

5. *Simple Guide to the published schemas*

This page picks out key features of some commonly available schemas, that you should bear in mind when choosing your base schema, or when mapping outputs for your customers. The schemas described here are only a selection especially relevant to images and cultural heritage. You should also consider schemas specific to your country (e.g. the e-GMS if you supply to the UK government) or sector (e.g. AGMES from FAO if you are into agriculture).

CDWA and CDWA Lite (Categories for the Description of Works of Art)

http://www.getty.edu/research/conducting_research/standards/cdwa/index.html

- Well adapted to works of art;
- Supported by good cataloguing guidance and other resources from the Getty Research Institute;
- CDWA Lite provides an XML schema for outputting the core elements in the full CDWA.

Overall: Could form the basis of your internal schema if you are in the business of fine art, and your customers may well request data formatted according to CDWA Lite.

VRA Core 4.0 (Visual Resources Association)

<http://www.vraweb.org/projects/vracore4/index.html>

- Well suited to images of cultural objects;
- Provides well for nested metadata ;
- Follows the same cataloguing rules as CDWA.

Overall: Could form the basis of your internal schema if you are in the business of photography, and your customers may well request data using this format.

IPTC Photo Metadata Standard (International Press Telecommunications Council)

[http://www.iptc.org/std/photometadata/specification/IPTC-PhotoMetadata\(200907\)_1.pdf](http://www.iptc.org/std/photometadata/specification/IPTC-PhotoMetadata(200907)_1.pdf)

- Comes in two flavours: IPTC Core provides the basics while IPTC Extension adds more fields;
- Updated 2009 to support metadata of artworks within the metadata of a photograph and to incorporate certain image rights properties of the PLUS standards;
- Popular with news agencies worldwide.
- **Overall:** Could form the basis of your internal schema if you are in the business of supplying images, and your customers may well request data using this format.

METS (Metadata Encoding and Transmission Standard)

<http://www.loc.gov/standards/mets/mets-home.html>

- Its strength is in managing and exchanging data about composite digital objects;
- Caters for embedding metadata drawn from another schema, e.g. VRA or DC;

Overall: Choose METS if your users need to find their way around complex digital “documents” made up of several chapters, images, tables, audio files, etc.

Dublin Core (DC)

<http://dublincore.org/documents/dces/>

- Popular with academic institutions and other organisations worldwide;
- Works best for text resources;
- Lacks many elements needed for images and for works of art.

Overall: Even if the schema is inadequate for your in-house needs, you should be able to supply outputs in DC format for customers who require it.

ESE (Europeana Semantic Elements)

<http://dev.europeana.eu/>

- Based on the Dublin Core, extended to meet needs of Europeana Project;
- Applicable to all kinds of cultural heritage, including literary works.

Overall: Not designed for your internal use, but for outputs if you wish to join in Europeana.

IEEE-LOM (Learning Object Metadata, Institution of Electrical and Electronics Engineers)

<http://ltsc.ieee.org/wg12/>

- Enables learners or instructors to search, evaluate, acquire, and utilize Learning Objects;
- Allows metadata for a simple object, e.g. an image, to be nested within a package of educational materials.

Overall: Unlikely to meet your internal needs, but useful for outputs if you supply to educational institutions or aggregators of materials.



museumdat

<http://www.museumdat.org/>

- Adapted from data elements in CDWA Lite and VRA Core;
- Used by museum portals.

Overall: Not designed for your internal use, but for outputs if you supply to certain museum portals.

PLUS License Data Format

<http://www.useplus.org> with technical specification at <http://ns.usePLUS.org>

- Applies only to IPR metadata for images;
- Accompanied by a useful glossary of image licensing terminology.

Overall: Could be integrated with your internal schema to allow access to and automated ingestion of image IPR metadata.

6. Applying metadata: Top Tips

- To accompany your schema you need a set of cataloguing rules. For each element that you have drawn from a published schema, your definition and cataloguing rule should be compatible with those of the source schema. (For example, if your schema includes elements drawn from CDWA or VRA Core, you should follow the corresponding rules from CCO.) You may need to extend the rules to cover particular issues that arise within your own image collection.
- Set up editorial and quality control procedures to ensure that all the cataloguing records you produce conform to the rules – consistently!
- Test the schema and the rules thoroughly before it is too late to change them. Use some of your quirkiest images for the testing. When the images are catalogued following the rules, does the metadata display correctly on your website and are your customers happy with the records you supply?
- Organize your workflow to maximize efficiency. For example:
 - Use automated processes wherever possible without jeopardizing quality
 - Avoid duplication of effort
 - Integrate controlled vocabularies wherever possible to speed data entry and avoid errors
- Beware of unnecessary perfectionism, which adds to costs without a proportionate increase in value. Subjective decisions are required to determine which rules/procedures are important to your business.
- Get the right balance between complexity and simplicity. Complexity is fine if it works effectively and is hidden in the backroom, but make sure that simplicity is delivered to the user. Likewise, high development (one-off) costs may be justified in building robust schemas, vocabularies and systems, provided they result in low costs at the operating (much-repeated) stage.
- Exploit the detail of your metadata in a specialised interface for users who want to distinguish between, for example, pictures of Vienna and works housed in Vienna, or between a recent painting in Renaissance style and one that was actually painted during the Renaissance period. But do not impose a complicated interface on the first-time user!
- Remember the potential of harvesting and metasearch for reaching new audiences. Your data need to be in the right format to reach portals and other networks across the globe.



- Remember the dangers of harvesting and metasearch. When your images are captured and displayed in a completely different environment, your context will be lost and extra metadata may be needed. For example, a UK collection of Victoriana might include pictures of “The Queen at home”; for an international audience the metadata needs more explanation of *which* queen.
- Plan where to store your metadata. If you embed it within the image file, it will accompany the image wherever it goes, helping to clarify issues such as IPR. If you hold it in a separate metadata database, that may improve efficiency and performance at search time. You may need to do both things; or to embed the metadata into the image files at the time when you export them from your system (rather than sending a separate XML file that could become detached from the images or even lost).
- If you use Adobe products, you can take advantage of XMP (eXtensible Metadata Platform), a proprietary technology to support embedding metadata with images. With XMP it is easy to incorporate IPTC and PLUS data into your image files.
- Be prepared for change as time passes, and design systems accordingly. For example you may need to add more metadata elements, or retrieve and correct all items referring to a certain copyright holder, or add another language to your search vocabularies.

7. Mapping and crosswalks

- Mapping is useful when moving data from one system to another, or when sending a query across the boundaries between systems. It is a bit like translation – you convert the data from one language or set of conventions into the language/conventions that will be understood by the neighbouring system.
- Just like ordinary translation, mapping suffers from “Chinese whispers”. In other words, each time the message passes from one system to the next, there is a risk of corruption of the signal, and the effect is multiplied the more mappings you apply to the same message. Chained mapping can bring disaster!
- With mappings it is important to specify direction, because a mapping that works well in one direction may not be suitable in the opposite direction. For example, it is easy to map “ducks” or “chickens” to “birds” (if the original terms are not available in the second vocabulary) but if you go in the opposite direction it is hard to choose which sort of bird to map to.
- You can apply mappings to metadata at two levels:
 - at the schema level, you rename the fields in one system to match the corresponding fields in another. For example, “Artist” in one system might be mapped to “Creator” in another. To make this possible, you map the names of schema elements. The pair of inter-mapped schemas is known as a “crosswalk”.
 - at the level of metadata values, you alter the data inside the fields. For example, the person known as “Winston Spencer Churchill” in one system might be named “Sir Winston Churchill” in another. To make this possible, you map the names or terms in the controlled vocabularies from which the metadata values are drawn.
- When you receive batches of images for input, use a schema crosswalk to feed the supplier’s data into the correct metadata field (with validation and other checks as necessary). This is easier if the supplier uses a standard schema such as IPTC. You may find it useful to develop an “input crosswalk” from each of the commonly used schemas to your own internal schema. But if your supplier uses a text-oriented schema such as Dublin Core, which does not make the distinctions commonly needed for works of art or for images, the result of the mapping may need hand-editing to fit it for your purposes.
- If the supplier has used a controlled vocabulary, you may also want to convert the data to your own vocabulary, for which you need a vocabulary crosswalk. Since a controlled vocabulary is very much bigger than a metadata schema, vocabulary crosswalks are expensive and hard to come by.



- Mapping is also useful when you are supplying completed records in the formats required by your customers. Because mappings are directional, the “output crosswalk” **from** your data **to** any of the commonly used schemas usually differs from the input crosswalk.
- The best quality crosswalks can be obtained by preparing mappings from your internal schema directly to each required output schema. However, you can save development time by consulting a published crosswalk (e.g. the “Metadata standards crosswalk” from CDWA to other standards, on the Getty Research Institute website). Avoid chained mapping if you can. (In other words, map directly from A to C rather than indirectly via B.)

8. Detailed metadata element and group descriptions

1. Identification

This cluster of elements identifies and names an image. It is assumed that in most cases the combination of Creator, Title and Year will all but uniquely identify an image, however there will be cases when they do not. The image id number will definitely uniquely identify an image.

1.1. Image id

Definition	Unique BAL identification number
Cardinality / obligation	Obligatory
Purpose	Combination of image id and agency prefix (derived from the 2.4. Image supplier element) form a unique key for each image.
Notes	
Not to be confused with	2.1. New id 2.2. Previous id(s) 2.3. Source id(s)
Examples	BAL000640 XIR930318 BON718181
Encoding schemes	Agency prefix is three capital letters. Image id is 6 numbers.
Mapped to	DC: identifier (should be labelled as a BAL id number, particularly if there are occurrences of 2.3. Source id.) IEEE LOM: 1.1. Identifier e-GMS: Identifier SPECTRUM: Object number



1.2. Title group

Definition	Defines the given title of an image, or if there is no given title a description of the image. The combination of the Title group and the Creator group should broadly uniquely identify an image if possible.
Cardinality / obligation	0 to many. Each image must have at least one title or Description (see 1.3 Description group). Different titles may represent the same title in different languages, commonly used variations on the title, or denote a different title for the image and work (see 1.2.5. Type). The primary title is given in the vernacular language of the creator.
Internal relationships	If a title is given, exactly one title must be marked as primary.
Purpose	Allowing the ability to store several titles for one image means that we can capture not only the 'official' title of the image, but also titles that we know may be searched for.
Notes	Only title to be included here. Year or era of creation should not be included.
Examples	Le Penseur would be given as the primary title, but The Thinker

	should be included because that is the title most likely to be searched for by English speaking clients. Where no title is available the description should be succinct and accurate. 'Carved wooden figurine from grave site.' for example.
Mapped to	DC: title (For more than one titles in the title group, the title flagged as primary should be listed first. In RDF the titles can be qualified as to language using the xml:lang tag.) IPTC Core: Title e-GMS: Title

1.2.1. Title

Definition	The title of the image
Cardinality / obligation	Obligatory
Examples	'Mona Lisa', 'Trajan's column'
Encoding schemes	Free text. Given that the title of an image is to be used in lists of search results, the title should be fairly succinct. However it would artificial to impose a fixed limit to the number of characters (unless it is absolutely necessary due to technology constraints.)
Mapped to	CDWA: 3.1. Title text. Because 3.1. Title text is core, if there are no entries in the Title group, the main entry from the 1.3. Description group should be offered as a title, as well as the CDWA 3.2. Title type field having the value 'descriptive' IEEE-LOM 1.2. Title (second element in LangString) SPECTRUM: Title

1.2.2. Language

Definition	Identifies the language of the title or description.
Cardinality / obligation	Obligatory
Examples	en, pt, fr, de, etc
Encoding schemes	ISO 639-1
Mapped to	CDWA: 3.4. Title Language IEEE-LOM 1.2. Title (first element in LangString)

	SPECTRUM: Object name/title language
--	--------------------------------------

1.2.3. Primary?

Definition	Flag denoting whether this member of the title group is the primary title.
Cardinality / obligation	Obligatory.
Purpose	To establish which title from the group is to be used (for example) when listing search results.
Notes	The primary title of an image is to be the title given to it by the creators in their vernacular language. Precisely one title in the title group must be primary.
Encoding schemes	True/false
Mapped to	CDWA: 3.3. Preference (true implies 'preferred', false implies 'alternative').



1.3. Description group

Definition	The textual description of a work. To be used when there is not a clear or explicit title to a work, or when there is some valuable contextual information about a work which is not naturally captured in other metadata fields.
Cardinality / obligation	0 to many. Each image must have at least one title or description. Different descriptions must be direct translations of each other.
Internal relationships	Exactly one of the descriptions must be (eventually) in English. If a description in English is not immediately available, there should be a workflow implemented to ensure that an English version is available with a given amount of time.
Purpose	Descriptions are expected to be in English. The ability to store more than one description in different languages is available for the case when image data is imported from a supplier in a foreign language. The description should be translated into English, but the original can be stored too.
Notes	The English description

	is considered to be primary, in that it will be the one made available in search results. If an English description is temporarily not available the available description should be shown in the search results, but with a clear indication of its language.
Mapped to	DC: description. (For more than one description in the description group, the description in English should be listed first. In RDF the description can be qualified as to language using the xml:lang tag.) CDWA: 18.1 Descriptive Note Text (or 3.1. Title Text, see Mapped to note in 1.2.1. Title). IPTC Core: Description e-GMS: Description SPECTRUM: Brief description

1.3.1. Description

Definition	The description of the image
Cardinality / obligation	Obligatory
Purpose	
Examples	'Female figure wearing a peplos'
Encoding schemes	Free text. It would be artificial to impose a fixed limit to the number of characters (unless it is absolutely necessary due to technology constraints.)
Mapped to	IEEE-LOM: 1.4. Description (second element of LangString)

1.3.2. Language

Definition	Identifies the language of the title or description.
Cardinality / obligation	Obligatory
Examples	en, pt, fr, de, etc
Encoding schemes	ISO 639-1
Mapped to	IEEE-LOM: 1.4. Description (first element of LangString)

1.4. Creator group

Definition	Defines the collection of individuals responsible for the creation of an image.
Cardinality / obligation	1 to many. All images must have at least one creator (even if the name of the creator is not explicitly known).
Internal relationships	There must be at least one primary creator, but there may be more than one (the responsibility for creating an image may be shared).
Purpose	Allows a moderately sophisticated description of 'who did what' to an image to be created.
Notes	
Examples	
Not to be confused with	
Mapped to	CDWA: 4.1. Creator Description e-GMS: Creator SPECTRUM: Object production person

1.3.1. Creator

Definition	Identifies a personality from the Personalities file who had some input into the creation of an image.
Cardinality / obligation	Obligatory
Purpose	Describes the 'who' part of 'who did what?'
Encoding schemes	Reference to Personalities authority file.
Mapped to	DC: creator CDWA: 4.1.3. Creator Identity IEEE-LOM: 2.3.2. Entity IPTC Core: Creator (?) IPTC Extension: Image Creator

1.3.2. Creator's role

Definition	Shows what role(s) in the creation of the image the creator played.
Cardinality / obligation	Obligatory
Purpose	Describes the 'what' part of 'who did what?'
Notes	May play some role in suggesting possible work types. eg if the creator's role was 'painter' this strongly suggests that the work type is 'painting'. However the role does not determine the work type, eg. a role of 'designer' does not determine any particular work type.
Not to be confused with	3.4. Work Type Group
Encoding schemes	Reference to a term in the controlled vocabulary for

	types of people (actually it is a subdivision of the general subject thesaurus). This vocabulary will include entries for all the professions, e.g. priests, gardeners, etc., and among these we should include sculptors, painters, etc.
Mapped to	CDWA: 4.1.4. Creator Role IEEE-LOM: 2.3.1. Role

1.4. Creation date group

Definition	Describes the dates between which an image was created.
Cardinality / obligation	Obligatory.
Internal relationships	Earliest date must be earlier than latest date. Display date must be congruent with the earliest and latest dates.
Notes	The earliest and latest dates are stored in a format compatible with ISO 8601, which means that dates earlier than 1582 are stored using the 'proleptic Gregorian calendar'. The cataloguing interface should prompt for and allow the cataloguer to input both Gregorian and Julian dates prior to 1582, but the system should convert them to proleptic Gregorian for storage.
Not to be confused with	4.4.12. Date/era depicted.

1.4.1. Earliest creation date

Definition	Describes the first date an image is known/believed to have been created.
Cardinality / obligation	Obligatory.
Purpose	Primarily for use in searching, giving an unambiguous earliest creation date. The date displayed in search results is 1.4.4. Display date.

Encoding schemes	ISO 8601 to the possible precision of days.
Mapped to	CDWA: 4.2.1. Earliest Date

1.4.2. Latest creation date

Definition	Describes the last date an image is known/believed to have been created.
Cardinality / obligation	Not obligatory.
Purpose	Primarily for use in searching, giving an unambiguous latest creation date. The date displayed in search results is 1.4.4. Display date.
Notes	If undefined, implicitly takes the value of the earliest date.
Not to be confused with	
Examples	
Encoding schemes	ISO 8601 to the possible precision of days.
Mapped to	CDWA: 4.2.2. Latest Date

1.4.4. Display date

Definition	A textual representation of the date range.
Cardinality / obligation	Non-obligatory, but if a specific display date is not entered by the cataloguer, an inferred display date is calculated from the earliest and latest creation dates.
Purpose	For ease of reading, particularly for approximate date ranges.
Notes	Could be automatically suggested from the earliest and latest dates.
Not to be confused with	
Examples	circa 12th Century
Encoding schemes	Free text, but guided by a display date syntax. (A warning to be given when the free text does not match the syntax, but can be overridden.)
Mapped to	DC: created (qualifier to date) CDWA: 4.2. Creation Date IEEE-LOM: 2.3.3. Date IPTC Core: Date Created e-GMS: Date.Created SPECTRUM: Object production date

2. Administration

Information held to ensure correct management of images. This information is not searchable by users unless otherwise stated.

2.1. New id

Definition	Gives the up-to-date image id if this image has been superseded.
Cardinality / obligation	Non obligatory
Purpose	To ensure that a relationship can be maintained between current image and previously held copies of the same image.
Notes	
Not to be confused with	1.1. Image id, 2.2. Previous id(s), 2.3. Source id(s)
Examples	
Encoding schemes	Same as 1.1. Image id. Three capital letters as prefix, followed by six numeric digits.

2.2. Previous id(s)

Definition	List of image ids which have been superseded by this image.
Cardinality / obligation	Non obligatory
Purpose	To ensure that a relationship can be maintained between current image and previously held copies of the same image.
Not to be confused with	1.1. Image id, 2.1. New id, 2.3. Source id(s)
Encoding schemes	Semi-colon delimited list of id numbers.

2.3. Source id(s)

Definition	Identification number or code as used by the supplying institution.
Cardinality / obligation	Non obligatory.
Purpose	To ensure that images can be traced back to suppliers, and to provide adequate information so that the image can be discovered in its hosting institution.
Notes	Should be searchable by users.
Not to be confused with	1.1. Image id, 2.1. New id, 2.2. Previous id(s)
Encoding schemes	Semi colon delimited list of identifiers. Each identifier should clearly labeled with the institution that gave it that identifier.
Mapped to	DC: identifier (institution labels should be included to differentiate this identifier with the BAL 1.1. Image id.) CDWA: 21.3. Current Repository Numbers SPECTRUM: Other number

2.4. Image supplier

Definition	Defines the agency supplying the image.
Cardinality / obligation	Obligatory
Purpose	
Notes	2.4. Image supplier, 2.5. Copyright handler and 2.6. Copyright holder define the set of agents responsible for an image, with the copyright holder overriding the handler, overriding the image supplier in terms of priority.
Encoding schemes	A reference to the agency authority file.
Mapped to	DC: rightsHolder IPTC Extension: Image Supplier (possibly with the addition of BAL as a supplier).

2.5. Copyright handler

Definition	Defines the organisation/individual responsible for dealing with the rights issues for the image.
Cardinality / obligation	Obligatory (but may be the same as Image supplier).
Notes	See 2.4. Image supplier
Encoding schemes	A reference to the agency authority file.
Mapped to	DC: rightsHolder

2.6. Copyright holder

Definition	Defines the organisation/individual receiving rights payments for use of the image.
-------------------	---

Cardinality / obligation	Obligatory (but may be the same as Image supplier/Copyright Handler).
Notes	See 2.4. Image supplier
Encoding schemes	A reference to the agency authority file.
Mapped to	DC: rightsHolder IPTC Core: Source (?) IPTC Extension: Copyright Owner SPECTRUM: Rights in holder

2.7. Copyright status

Definition	Defines whether the image is in or out of copyright. The status can be calculated from the date of death and nationality of the primary creator, but may be overridden.
Cardinality / obligation	Obligatory.
Purpose	
Encoding schemes	In copyright / Out of copyright / Status unknown.
Mapped to	DC: rights IEEE-LOM: 6. Rights IPTC Core: Copyright notice e-GMS: Rights

2.8. Permissions

Definition	Defines the set of specific permissions, restrictions and conditions that relate to an image.
Cardinality / obligation	0 to many.
Notes	Inherited from 2.4. Image supplier, but may be overwritten
Encoding schemes	Set of references to the permissions / restrictions / conditions authority file.
Mapped to	DC: license (qualifier to rights) IPTC Core: Rights Usage Terms

2.9. Bacs agency

Definition	Defines the BACS agency for an image.
Cardinality	Non obligatory

obligation	
Purpose	To define the agency to whom fees should be paid in the case of a BACS artist,
Not to be confused with	2.4. Image supplier
Encoding schemes	Reference to Agency authority file.
Mapped to	DC: rights

2.10. Status

Definition	Defines the image's current place in the image lifecycle.
Cardinality / obligation	Obligatory
Encoding schemes	Being created / being edited / being proofed / accepted / withdrawn.

2.11. Withdrawal date

Definition	Defines the date that an image was or will be withdrawn from being actively available for purchase..
Cardinality / obligation	Non-obligatory
Encoding schemes	ISO 8601
Mapped to	DC: rights e-GMS: Disposal (?)

2.12. Credit line

Definition	Defines a line of text which must be included with an image when it is published by a client.
Cardinality / obligation	Obligatory (but may be blank).
Purpose	
Notes	Generated by a sequence of templates which take information from Image supplier, Copyright Handler/Holder fields and formats them. This field is in the form of a template which can overwrite the inherited templates.
Not to be confused with	2.13. Extended credit line.

2.13. Extended credit line

Definition	Allows for the inclusion of highly extended credit lines which assign credit to many people or agencies.
Cardinality / obligation	Non-obligatory
Not to be	2.12. Credit line.

confused with	
Mapped to	CDWA: 22.1. Copyright Statement IPTC Core: Credit line SPECTRUM: Credit line

2.14. Comp copies?

Definition	Flags whether the image supplier requires complimentary copies of this image when published.
Cardinality / obligation	Obligatory
Encoding schemes	True/false



3. Objective image information

3.1. Dimensions

Definition	Defines the physical dimensions of the original image.
Cardinality / obligation	Non obligatory
Purpose	
Notes	Adhering to the strict formatting rules defined in the encoding scheme assures that the dimensions can be machine read and processed, and therefore can be automatically recalculated in metric or imperial units.
Not to be confused with	
Examples	10" x 20" 100cm x 3m x 20cm 2' x ? 32cm Left panel: 30" x 6"; Centre panel: 30" x 2'; Right panel: 30" x 6"
Encoding schemes	Can be expressed using metric (mm, cm, or m) or imperial (' or ") units. All numbers must be qualified by a measurement. Units could be mixed, but this should be avoided. If the image is in several distinct parts (eg a triptych) then each part should be labelled (with free text) in a semi-colon delimited list, followed by a colon and the dimensions. Two dimensional

	measurements are expressed as height x width, three dimensional as height x width x depth. A single measurement is assumed to be the diameter of a two dimensional object. If a measurement is not known then a question mark is used.
Mapped to	DC: extent (qualifier to format) CDWA: 6.1. Dimensions Description SPECTRUM: Dimension, Dimension measured part, Dimension measurement unit, Dimension value

3.2. Orientation

Definition	Describes the orientation of the original image.
Cardinality / obligation	Non obligatory
Purpose	
Notes	
Not to be confused with	
Examples	
Encoding schemes	Orientation CV.
Mapped to	DC: description CDWA: 12.1. Orientation/Arrangement Description

3.3. Image location

Definition	Describes the location of the original artwork/artefact.
Cardinality / obligation	Obligatory. (Must be obligatory to satisfy CDWA Core requirements.)
Purpose	
Notes	
Not to be confused with	4.4.4. Location depicted
Examples	
Encoding schemes	Reference to the Location/Landmarks authority file.
Mapped to	DC: Provenance (?) CDWA: 21.2. Current Repository / Geographic Location e-GMS: Location

3.4. Work type group

Definition	Describes the processes and materials involved in the creation of the image.
Cardinality / obligation	1 to many.
Internal relationships	At least one work type must be of primary significance.
Purpose	
Notes	
Examples	
Not to be confused with	
Mapped to	DC: type (note that the type information should include one of the following DCMI Type terms: Image or StillImage, and any further information should make it clear that the BAL holdings are photos/digitisations of the primary source). CDWA: 1.2. Object/Work Type (again, there should be a note added making it clear that BAL holds an image of the primary source, not the source itself.)

3.4.1. Work type

Definition	Describes a genre, process or technique involved in the creation of the image.
Cardinality / obligation	Obligatory
Purpose	

Notes	
Not to be confused with	
Examples	Sculpture, engravings
Encoding schemes	Work types/techniques CV.

3.4.2. Significance

Definition	Describes the significance of the work type in the overall process of creating the image.
Cardinality / obligation	Obligatory
Purpose	
Notes	
Not to be confused with	
Examples	
Encoding schemes	Primary / secondary / insignificant

3.5. Materials

Definition	A description of the physical materials used in the creation of the image.
Cardinality / obligation	1 to many.
Purpose	
Notes	3.6 Medium should be in some way congruent with 3.5. Materials
Not to be confused with	3.6. Medium
Examples	Carrara marble
Encoding schemes	Materials CV.

schemes	
Mapped to	DC: medium CDWA: 7.1. Materials/Techniques Description

3.6. Medium

Definition	A free text description of the physical materials used in the creation of the image.
Cardinality / obligation	Obligatory
Purpose	A free text expression of 3.5. Materials. Medium allows a more nuanced, and richer description of the materials used to create in image than the list of terms allowed in 3.5 Materials. If no text is entered, then this field defaults to a textual list of 3.5. Materials.
Notes	3.6 Medium should be in some way congruent with 3.5. Materials
Not to be confused with	3.5. Materials
Encoding	Free text.

4. Interpretative image information

4.1. Image type

Definition	Describes whether the image is in one of four broad browsing types.
Cardinality / obligation	Obligatory.
Purpose	
Notes	
Not to be confused with	
Examples	
Encoding schemes	Illustration / photo / artefact / other.
Mapped to	DC: type eGMS: Type

4.2. Style/period

Definition	Describes the style or period in which the image is expressed.
Cardinality / obligation	0 to many.
Purpose	
Notes	
Not to be confused with	1.4. Creation date group. Styles and periods are loosely related to specific time periods, but not definitively.
Examples	Op art, Bronze Age
Encoding schemes	Reference to Style/Period/Movement CV.
Mapped to	CDWA: 5.1. Styles/Periods Description IEEE-LOM: 1.6. Coverage(?) SPECTRUM: Associated cultural affinity(?)

4.3. Category group

Definition	Description of the subject categories which apply to the image.
Cardinality / obligation	0 to many.
Internal relationships	
Purpose	
Notes	A category is a term which describes an image overall, for example the BAL box numbers. A single image may belong to several different categories. Furthermore there may be several category schemes, for example the BAL box numbers, the Wikipedia classification scheme, etc., which we may wish to categorise images by.
Examples	
Not to be confused with	
Mapped to	DC: subject CDWA: 2.1. Classification Term IEEE-LOM: 1.5. Keyword(?) e-GMS: Subject

4.3.1. Category

Definition	A subject area within which the image can be placed.
Cardinality / obligation	Obligatory
Purpose	To facilitate successful

	browsing of the image collection.
Examples	history, geology
Encoding schemes	A reference to a category, taken from the category scheme.

4.3.2. Category scheme

Definition	The scheme which the category is drawn from.
Cardinality / obligation	Obligatory
Purpose	
Notes	
Not to be confused with	
Examples	
Encoding schemes	



4.4. Subject group

Definition	A collection of terms which describe the content of an image
Cardinality / obligation	0 to many
Internal relationships	Each of the subject terms must be associated with one of the subject types below, and be drawn from its associated controlled vocabulary. Each term may be drawn from an associated controlled vocabulary, may be a candidate for inclusion in a controlled vocabulary, or may be an 'uncontrolled' term, which has been considered, but rejected for inclusion in the controlled vocabulary, but is still included as a searchable term for the image as a one off. Some of the subject terms may be qualified, by number, gender, etc. In these cases there is a 1 to 1 relationship between the subject terms and the qualification.
Purpose	To facilitate successful searching of the image collection.
Notes	Terms that do not fit with the defined subject groups should be added to 4.4.1. Uncontrolled terms, and there should a work process which periodically checks the

	uncontrolled terms and creates new defined subjects (and associated controlled vocabularies) as necessary.
Examples	
Not to be confused with	
Mapped to	DC: subject. The identity of the controlled vocabulary from which the value is drawn should be specified too. Values should be appropriate to the level of specificity depicted in the image, and broader terms should not routinely be included. CDWA: 16.2. Subject Indexing Terms IEEE-LOM: 9.4. Keyword IPTC Core: Keywords e-GMS: Subject

4.4.1. Uncontrolled terms

Definition	A 'bucket' of terms which describe the content of an image, but which do not fit in the current subject groups.
Cardinality / obligation	0 to many.
Purpose	Provides a 'let out' so that terms are not foisted into subject headings that they do not really fit with.
Notes	This set of terms must be strictly managed, to ensure that it does not become a free text substitute that defeats the point of having controlled vocabularies.
Not to be confused with	
Examples	
Encoding schemes	Free text.

4.4.2. Event depicted

Definition	Name of a specific event depicted.
Cardinality / obligation	0 to many.
Purpose	
Notes	
Not to be confused with	4.4.3. Event type depicted
Examples	Battle of Waterloo.
Encoding schemes	Event CV.
Mapped to	DC: subject IPTC Extension: Event SPECTRUM: Content – event name

4.4.3. Event type depicted

Definition	Describes a type of event depicted.
Cardinality / obligation	0 to many.
Purpose	
Notes	
Not to be confused with	4.4.2. Event depicted.
Examples	Sea battles.
Encoding schemes	Types of Event facet of Subject Thesaurus.

4.4.4. Personality depicted

Definition	Name (plus other distinguishing data) of an individual depicted.
Cardinality / obligation	0 to many.
Purpose	
Notes	
Not to be	4.4.12. People depicted.

confused with	
Examples	Winston Churchill (1874-1965)
Encoding schemes	Personality CV.
Mapped to	IPTC Extension: Person Shown in the Image SPECTRUM: Content -- person

4.4.5. Location depicted

Definition	Names a specific location or landmark depicted
Cardinality / obligation	0 to many.
Purpose	
Notes	
Not to be confused with	3.3. Image location group.
Examples	10 Downing Street (Westminster), Nelson's Column (Trafalgar Square)
Encoding schemes	Location/Landmarks CV
Mapped to	IPTC Extension: Location Shown in the Image SPECTRUM: Content -- place

4.4.6. Design element depicted

Definition	Describes a specific design element depicted.
Cardinality / obligation	0 to many
Purpose	
Notes	
Not to be confused with	
Examples	acanthus motif
Encoding schemes	Design elements facet of Subject Thesaurus.

4.4.7. Object depicted

Definition	Describes a specific object depicted.
Cardinality / obligation	0 to many.
Purpose	
Notes	
Not to be confused with	
Examples	armchairs
Encoding schemes	Objects facet of Subject Thesaurus.
Mapped to	IPTC Extension: Artwork or Object in the Image SPECTRUM: Content -- object

4.4.8. Activity depicted

Definition	Describes a specific activity depicted.
Cardinality / obligation	0 to many.
Purpose	
Notes	
Not to be confused	

with	
Examples	walking
Encoding schemes	Activities facet of Subject Thesaurus.
Mapped to	SPECTRUM: Content -- activity

4.4.9. Abstract concept depicted

Definition	Describes an emotion, mood or other abstract concept depicted.
Cardinality / obligation	0 to many.
Purpose	
Notes	
Not to be confused with	
Examples	pain; excitement; peace
Encoding schemes	Emotion/Mood/Abstract ideas facet of Subject Thesaurus.
Mapped to	SPECTRUM: Content -- concept

4.4.10. Organism depicted

Definition	Describes a specific organism depicted.
Cardinality / obligation	0 to many.
Purpose	
Notes	
Not to be confused with	
Examples	roses, tigers, mosquitoes
Encoding schemes	Organisms facet of Subject Thesaurus.

4.4.11. Organisation depicted

Definition	Description of a specific organisation depicted.
Cardinality / obligation	0 to many
Purpose	
Notes	
Not to be confused with	
Examples	
Encoding schemes	Organisations facet of Subject Thesaurus.
Mapped to	IPTC Extension: Name of Organisation Featured in the Image SPECTRUM: Content -- organisation

4.4.12. People depicted

Definition	Description of types of people depicted.
Cardinality / obligation	0 to many.
Purpose	
Notes	
Not to be confused with	
Examples	
Encoding schemes	People facet of Subject Thesaurus
Mapped to	SPECTRUM: Content -- people

4.4.12.1. Number

Definition	Qualifier to 4.4.12. People depicted, showing the number of persons depicted.
Cardinality / obligation	Obligatory.
Examples	
Encoding schemes	Number

4.4.12.2. Gender

Definition	Qualifier to 4.4.12. People depicted, showing the gender of persons depicted.
Cardinality / obligation	Obligatory.
Examples	
Encoding schemes	Male / Female / Mixed

4.4.12.3. Ethnicity

Definition	Qualifier to People depicted, showing the
-------------------	---

	ethnicity of persons depicted.
Cardinality / obligation	Non obligatory.
Examples	
Encoding schemes	?

4.4.12.4. Age group

Definition	Qualifier to People depicted, showing the age group(s) of persons depicted.
Cardinality / obligation	Non obligatory.
Examples	children, adults, elderly
Encoding schemes	

4.4.13. Date/era depicted

Definition	Description of a specific date/era depicted.
Cardinality / obligation	0 to many.
Purpose	
Notes	
Not to be confused with	
Examples	
Encoding schemes	Style/Period/Movement CV in some cases, ISO 8601 in others.
Mapped to	SPECTRUM: Content -- date

4.4.14. Work of art depicted

Definition	Name of a specific work of art depicted, including literary and musical works.
Cardinality / obligation	0 to many.
Purpose	
Notes	
Not to be confused with	
Examples	Book of Kells, A Midsummer Night's Dream
Encoding schemes	Named works CV.
Mapped to	IPTC Extension: Artwork or Object in the Image

5. Relational information

5.1. Related image group

Definition	Describes interrelationships between images.
Cardinality / obligation	0 to many.
Internal relationships	
Purpose	
Notes	
Examples	
Not to be confused with	5.2. Image set. 4.3. Category group.
Mapped to	DC: relationship CDWA: 20.1. Related Work Label / Identification IEEE-LOM: 7. Relation e-GMS: Relation

5.1.1. Image

Definition	The image with which there is a relationship.
Cardinality / obligation	Obligatory.
Purpose	
Notes	
Not to be confused with	
Examples	
Encoding schemes	The same as 1.1. Image id.
Mapped to	IEEE-LOM: 7.2. Resource

5.1.2. Relation type

Definition	Description of the type of relationship between images.
Cardinality	Obligatory.

obligation	/
Purpose	
Notes	The type of relationship determines whether it is reciprocal or not. (Therefore reciprocity must be encoded in the relation types.)
Not to be confused with	
Examples	
Encoding schemes	?
Mapped to	IEEE-LOM: 7.1. Kind



9. Metadata schemas crosswalk

In-house schema	Dublin Core	CDWA	IEEE-LOM	IPTC	e-GMS	SPECTRUM
1. Identification						
1.1. Image id	identifier		1.1. Identifier		Identifier	Object number
1.2. Title group	title			Title (Core)	Title	
1.2.1. Title		3.1. Title text	1.2. Title			Title
1.2.2. Language		3.4. Title Language	1.2. Title			Object name/title language
1.2.3. Primary?		3.3. Preference				
1.3. Description group	description	18.1 Descriptive note text		Description (Core)	Description	Brief description
1.3.1. Description			1.4. Description			
1.3.2. Language			1.4. Description			
1.4. Creator group	creator	4.1. Creator Description			Creator	Object production person
1.4.1. Creator		4.1.3. Creator Identity	2.3.2. Entity	Creator (Core) Image Creator (Extension)		
1.4.2. Creator's role		4.1.4. Creator role	2.3.1. Role			
1.5. Creation date group						
1.5.1. Earliest date		4.2.1. Earliest Date				



1.5.2. Latest date		4.2.2. Latest Date				
1.5.3. Display date	date.created	4.2. Creation Date	2.3.3. Date	Date Created	Date.Created	Object production date
2. Administration						
2.1. Current id						
2.2. Previous id(s)						
2.3. Source id(s)	identifier	21.3. Current Repository Numbers			Identifier	Other number
2.4. Image supplier	rightsHolder			Image Supplier (Extension)		
2.5. Copyright handler	rightsHolder				Rights.Custodian	
2.6. Copyright holder	rightsHolder			Source (Core) Copyright Owner (Extension)	Rights.Copyright	Rights in holder
2.7. Copyright status	rights		6. Rights	Copyright notice (Core)	Rights.Copyright	
2.8. Permissions	rights.licence			Rights Usage Terms (Core)		
2.9. Bacs agency	rights					
2.10. Status						
2.11. Withdrawal date	rights				Disposal	
2.12. Credit line						
2.13. Extended credit line		22.1. Copyright statement		Credit line (Core)		Credit line



2.14. Comp copies?						
3. Objective image information						
3.1. Dimensions	format.extent	6.1. Dimensions Description			Format.extent	Dimension, Dimension measured part, Dimension measurement unit, Dimension value
3.2. Orientation	description	12.1. Orientation / Arrangement Description			Description	
3.3. Image location	provenance	21.2. Current Repository / Geographic Location			Location	
3.4. Work type group	type	1.2. Object / Work Type			Type	
3.4.1. Work type						
3.4.2. Significance						
3.5. Materials						
3.6. Medium	format	7.1. Materials / Techniques Description				
4. Interpretative image information						
4.1. Image type	type				Type	
4.2. Style/period	temporal	5.1. Styles / Periods Description	1.6. Coverage		Coverage.temporal	
4.3. Category group	subject	2.1. Classification term	1.5. Keyword		Subject.	
4.3.1. Category						

D1.2.3 MILE Final Report: Appendix 1

4.3.2. Category scheme						
4.4. Subject group	subject	16.2 Subject Indexing Terms	9.4. Keyword	Keywords	Subject.	
4.4.1. Uncontrolled terms						
4.4.2. Event depicted				Event (Extension)		Content – event name
4.4.3. Event type depicted						
4.4.4. Personality depicted				Person Shown in the Image (Extension)		Content – person
4.4.5. Location depicted						Content – place
4.4.6. Design element depicted						
4.4.7. Object depicted				Artwork or Object in the Image (Extension)		Content – object
4.4.8. Activity depicted						Content – activity
4.4.9. Abstract concept depicted						Content – concept
4.4.10. Organism depicted						
4.4.11. Organisation depicted				Name of Organisation Featured in the Image (Extension)		
4.4.12. People depicted						Content – people
4.4.12.1. Number						



4.4.12.2. Gender						
4.4.12.3. Ethnicity						
4.4.12.4. Age group						
4.4.13. Date/era depicted						Content – date
4.4.14. Work of art depicted				Artwork or Object in the Image (Extension)		
5. Relational information						
5.1. Related image group	relation	20.1. Related Work Label / Identification	7. Relation		Relation	
5.1.1. Image			7.2. Resource			
5.1.2. Relation type			7.1. Kind			
5.2. Image set						

10. Approaches to multilingual provision

Several different approaches are possible, with different implications for metadata:

- a. Ignore the needs of users whose first language is not the same as the main language used in your records. In this way you avoid doing anything special or incurring extra costs with your metadata. The main disadvantage is loss of market share.
- b. Provide for users who search in different languages, by enabling them to get their queries translated automatically while they search. Unless you enable automatic translation of the outputs, the records they retrieve will be presented in your original language, which may or may not be acceptable to the users. With this approach, there may be doubts about the quality of the translation and there will be a significant development cost and running cost of the automatic translation, but you avoid the (initially larger) cost and delay of translating the metadata.
- c. Translate all the metadata into the languages wanted by your customers, and provide a user interface in the same languages. This way you have significant up-front development costs, and a small overhead associated with translation of vocabulary updates, but the extra running costs are minimal. In the long run, it is the cheapest way of enabling multilingual access to your collection.

Assuming you choose option c, the simplest way to translate the metadata is to translate the controlled vocabularies. Thus the same translations can be used over and over again, each time the same names or keywords are used in cataloguing. Only the fields which do not use multilingual controlled vocabularies will be left without translations.

11. *Developing and applying controlled vocabularies:*

Top Tips

- Typically you will need 3-4 different types of vocabulary:
 - simple lists (at most 20 items in each, as a flat list)
 - name authority lists (e.g. of artist names)
 - thesauri (to support search functions)
 - taxonomies (to support browse functions)
- All but the simple lists are expensive to build and maintain. Take advantage of existing lists wherever possible.
- Maintenance/updating should be built into your procedure and your budget, in perpetuity.
- The overhead cost of a controlled vocabulary is justified over time by:
 - a. reduction in operating costs, as the vocabulary supplies synonyms, broader terms, narrower and related terms automatically
 - b. similarly, a multilingual vocabulary supplies term translations automatically
 - c. improved consistency and error detection
- For geographic names, as well as normal synonyms and translations, your vocabulary should supply geographical coordinates, latitude and longitude, postcodes, and any other popular coding system.
- Thesaurus construction needs specialized software. A modest purchase saves you a lot of time and errors.
- Thesaurus construction needs tight editorial control. Appoint someone meticulous, good at spelling, who enjoys playing with words and has relevant experience.
- When developing your own vocabularies, follow the national and international standards. (See list)
- Include common misspellings (so long as they are unambiguous) among the other lexical variants of preferred terms that you admit.
- A multilingual vocabulary should give equal status to all the languages. To avoid the dominance of one language, take all the required languages into account from the start.

D1.2.3 MILE Final Report: Appendix 1

- Some of your customers may want to view and select terms from your vocabularies, but mostly you should try to integrate the vocabularies into your search and cataloguing interfaces so that they convert terms automatically behind the scenes.
- While mapping from one metadata schema to another is usually feasible, mapping from one controlled vocabulary to another is usually a much bigger job. The development expense of mapping between vocabularies is only justified if you routinely have to provide outputs or search capabilities in more than one.