38189

MILE

Metadata Classification Seminar 3 & Report

www.mileproject.eu

Deliverable number/name	*D2.3 Metadata Classification Seminar 3 and Report*
Dissemination level	*Public*
Delivery date	*February 28th 2009*
Status	*Final*
Author(s)	*Jessica Tier, Project Manager* *Lucy Geering, Project Assistant* *Celestine Bramley, Project Assistant*

e*Content*plus

This project is funded under the *e*Content*plus* programme[1], a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable.

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

Contents

MILE METADATA CLASSIFICATION CONFERENCE 2: Attendance

9.30am – 5.00pm, Wednesday 28th January, and 10.00am – 2.00pm, Thursday 29th January 2009
The Royal Thames Yacht Club, 60 Knightsbridge, London, SW1X 7LF, UK

A Picture's Worth a Thousand Euros' Delegate List

No.	Company	Contact name	Company Role
1	Messum's	Millie Messum	Research Assistant
2	Messum's	Dr Jane Hamilton	Archivist & Researcher
3	The Wallace Collection	Nell Carrington	Picture Librarian
4	Visual Media Services	Julia Bate	Picture Researcher
5	Visual Media Services	Nicole Kaczynski	Delegate
6	The Geffrye Museum	Christine Lalumia	Deputy Director
7	The Geffrye Museum	Mandy Williams	Web/Picture Library Administrator
8	Arquivo Internacional de Cor	Conceição Guimaraes	Delegate
9	Arquivo Internacional de Cor	Luis Ferraira	Delegate
10	Arquivo Internacional de Cor	Maria Guimaraes	Delegate
11	Brighton Museum & Art Gallery	Beverley Green	Image Reproduction Officer
12	Brighton Museum & Art Gallery	Kevin Bacon	Curator of Photographs
13	Elmbridge Museum	Sue Webber	Collections Officer
14	Elmbridge Museum	Rachel Perkins	Museum Manager
15	Nature Picture Library	Giles Manning	IT Manager / Picture Researcher
16	TopFoto	Flora Smith	Delegate
17	TopFoto	Callum McClellan	Delegate
18	TopFoto	Christine Kendrick	Delegate
19	Digital Humanities Observatory	Dot Porter	Metadata Manager
20	The Fitzwilliam Museum	Margaret Greeves	Head of Public Engagement
21	Adult and Community Services Libraries Archives and Information	Jason Siddall	Heritage Information Manager
22	Maidstone Museum & Bentlif Art Gallery	Fiona Woolley	Keeper of Fine & Applied Art
23	Dulwich Picture Gallery	Fulvio Rubesa	Picture Librarian
25	Venice Picture Library	Sarah Quill	Photographer
26	Tate Enterprises Ltd	Clive Coward	Tate Images Manager

27	London College of Communication	Dr. David Penfold	Senior Lecturer
28	Brighton University	Dr. Karina Rodriguez-Echavarria	Research Fellow
29	RIBA	Laura Whitton	Assistant Curator of Photographic Collection
30	Country Life	Justin Hobson	Picture Library Manager
31	Phocuus	Henry Horton	Delegate
32	DACS	Joy Stanley	Copyright Membership Executive
33	Focal International	Sue Malden	Chair
34	Alinari 24 ORE	Sam Minelli	Head of Research Applications Centre
35	Alinari 24 ORE	Andrea de Polo	Head of IT
36	Archetypon	George Kordelis	Project Manager
37	BAPLA	Angela Anderson	Acting CEO
38	BAPLA	Sarah Saunders	Chair of BAPLA Metadata Group
39	BAPLA	Simon Cliffe	Executive Director
40	Bildombudsmannen	Staffan Teste	Legal Advisor
41	Bridgeman Berlin	Annette Godefroid	General Manager
42	Capture Ltd	Abbie Enock	Managing Director
43	CEPIC	Sylvie Fodor	CEPIC Coordinator
44	Collections Trust	Nick Poole	CEO
45	Collections Trust	Gordon McKenna	Standards and Systems Manager
46	Electric Lane	Graeme Cookson	Digital Consultant
47	Europeana	Robina Clayphan	Metadata Expert
48	Fotofinder	Frank Toelle	Collections Manager
49	Fotofinder	Ali Paczensky	Managing Director
50	Imense Ltd	Dr Chris Town	Co-Founder & President of Technology
51	Imense Ltd	Tony Rowland	Delegate
52	Imense Ltd	Stuart Cox	Delegate
53	Knowledge Transfer Partnership	Kalliopi Vacharopoulou	KTP Associate
54	MILE	Jessica Tier	MILE Project Manager
55	MILE	Lucy Geering	MILE Project Administrator
56	MILE	Celestine Bramley	MILE Project Assistant
57	Photo 12.com	Suzanne Davies	Delegate
58	SILVER	Brian Kavanagh	Systems Operator
59	SILVER	Stella Dextre Clarke	Information Consultant

60	SILVER	Richard Butterworth	Senior Metadata Analyst
61	SILVER	Hilary Bates	Project Manager
62	Stiftung Preussischer Kulturbesitz	Axel Ermert	Scientist
63	Stiftung Preussischer Kulturbesitz	Monika Hagedorn-Saupe	Deputy Director
64	System Simulation	Mike Stapleton	Technical Director
65	System Simulation	Graham Howard	Design Director
66	System Simulation	Alec Turner	Software Engineer
67	The Bridgeman Art Library	Harriet Bridgeman	Chairman
68	The Bridgeman Art Library	Pandora Mather-Lees	SILVER Project Manager / Managing Director (Sales)
69	The Bridgeman Art Library	Ed Vald	Vocabulist
70	The Bridgeman Art Library	Adrian Gibbs	Collections Manager
71	The Bridgeman Art Library	Felicity Page	Head Cataloguer
72	The Bridgeman Art Library	Hannah Armstrong	Controlled Vocabularies
73	The Bridgeman Art Library	Victoria Bridgeman	Managing Director
74	The Bridgeman Art Library	Steffen Wedepohl	German Rights & Sales Executive
75	The Bridgeman Art Library	Paul King	Head of IT
76	The Bridgeman Art Library	Susana Cespedes	International Agents Account Manager
77	The Image Business	Angela Murphy	Consultant
78	The Israel Museum	Amalyah Keshet	Head of Image Resources and Copyright Management
79	Trinity College Dublin	Garret McMahon	Institutional Repository Content Manager
80	Trinity College Dublin	Niamh Brennan	Programme Manager
81	Wardynski & Partners	Ewa Gillabert	Legal Advisor
82	Wardynski & Partners	Ewelina Sliwinska	Legal Advisor
83	Wardynski & Partners	Janusz Kolczynski	Legal Advisor
84	Wellcome Library / BAPLA	Catherine Draycott	Head of Wellcome Images

Introduction

The final seminar for Work Package 2, Metadata Classification, took place at the Royal Thames Yacht Club, London on the 28th and 29th January, 2009.

This seminar, "A Picture's Worth a Thousand Words Euros" built on two previous events dedicated to metadata classification, this time focusing on metadata crosswalks, or mapping. The seminar invited participation from experts involved in the development of metadata for digital image access, in order to further MILE's intention to try to bridge the gaps between commercial and cultural heritage users of digital images; to challenge current working practices; to explore stakeholders' and users' needs in relation to workflow systems and productivity; and to seek a unified approach to metadata classification. A workshop was designed to explore mapping practices, the results of which will feed into a proposal for a Memorandum of Understanding (MoU), committing industry members to use a list of core fields appropriate to specified metadata schemas.

Many users of digital image archives suffer from difficulties of access through unsuccessful search mechanisms. MILE's previous seminars dedicated to classification identified that the prevailing issue for mapping is concerned with the differing syntaxes of the various metadata schemas and element sets which are in use. Mapping is the matching-up of elements from different metadata schemas and standards, in order to facilitate semantic interoperability and cross-repository searching. Each schema contains a set of field names which are used to arrange and describe information relating to an image or object. The problem is that similar or the same content is described by different fields within different metadata standards, and that there are varied syntaxes for encoding metadata, i.e. the rules dictating where specific information is to be placed within fields differ from standard to standard, which impedes simple mapping between schemas and therefore leads to problems in a networked environment. In addition, standards differ according to user – standards used by researchers and archivists tend to be different to those used for resource discovery. So there are specific metadata interoperability problems within museum environments which are replicated through internal workflow systems, and which then follow through between museums and external entities.

Feedback generated from previous seminars suggested that a significant factor in the implementing of good metadata and integrating mapping systems is time and cost. Whilst it is not in MILE's remit to produce a business plan, the financial concerns raised suggested that a commercial justification would be very useful, and a key element in persuading organisations to take up MILE's recommendations for best classification practices.

80 attendees registered for this seminar, with strong representation from both the cultural and commercial industry sectors. There was a lack of parity in EU attendee numbers; MILE partners represented Greece and Italy alongside French and German attendees. Whilst delegate numbers exceeded our target audience of 40, the lack of European representation was disappointing, especially in view of the number of invitations sent to EU industry members. Future invitations to seminars will be sent out earlier in an attempt to mitigate this problem; in addition, it is hoped that the significant

interest previously shown in the Memorandum of Understanding will result in further European commitment to the aims of the project.

A Picture's Worth a Thousand Euros: Report

Harriet Bridgeman, Chairman of The Bridgeman Art Library, introduced the day welcoming attendees, followed by Jessica Tier, the MILE Project Manager, who gave an overview of MILE's aims and a brief description of metadata standards and mapping.

This was followed by Nick Poole, CEO of Collections Trust. Poole gave a robust presentation arguing the case that applying commercial approaches to heritage collections needn't sacrifice their cultural integrity. Poole suggested that without good metadata, cultural institutions' internal management, public and educational access, and commercial gain are compromised. He suggested that both commercial and academic activity (such as the research and cataloguing carried out in museums) can drive value back into the public work of cultural institutions but to do this successfully, good metadata is essential. He reiterated that good metadata facilitates data mapping, rationalisation and harmonisation, and thus makes interoperability (federated searching and cross-cultural searching) possible.

Poole acknowledged the existence of idealistic disparities within an industry that is struggling to marry culture with commerce. His eloquent and persuasive delivery was received well; however, in its current state, the museum industry is fragmented and such visionary arguments are often described as unachievable, largely due to policy direction and perceived high start-up costs. However, if approached with integrity and an ethic acceptable to government and EU policy frameworks, a commercial approach has every chance of successfully delivering culture, fulfilling both the public-service remit and the commercial and economic needs that are needed to support the industry.

Whilst Poole presented a balanced argument for employing good metadata, the lack of tangible evidence supporting the commercial argument needs to be addressed; that notwithstanding, his presentation supports the European Commission's recommendation that it is imperative that both the cultural sector and commercial image libraries implement good metadata if they want to improve interoperability and access to digital images.

Mike Stapleton, Director of System Simulation, provided an excellent presentation demystifying the practical implementation of data structures, metadata standards, words and pictures. Stapleton emphasised the ease with which such systems can be used, and supported Nick Poole's argument that good metadata improves collection management, cataloguing, discovery and exchange. He demonstrated a variety of metadata categories used by the image industry, acknowledging differences in the way that the industry members use them for cataloguing purposes. For example, many different standards are used but best practice is often missing; appropriateness, consistency and bad management are all problems preventing good metadata. Stapleton acknowledged that standards can be mapped, but he stressed that the key to effective and automated systems is first and foremost good metadata - then planned architectures for mapping can be put into place.

Stapleton identified how crucial it is to decide on the information that you want in both an image editing application, and the image database system prior to cataloguing, as this will dictate the metadata that you need. He stressed how all user needs must be

established to facilitate good metadata, asking

“are the metadata categories that you are using adequate for all potential uses and users?”

He also advocates the customising of metadata schemas for individual organisation’s needs, recognising that ‘one size fits all’ is never going to be a realistic, or appropriate, solution to creating standardised metadata;

“Select and use appropriate metadata elements, these should be customised according to institutional needs.”

Instead, he emphasised the importance of rules and consistency for every data entry level. As in previous seminars, the audience was reminded of the importance applying keywords using controlled vocabularies thesauri and local authority files; the importance of knowing “who your users are” in terms of matching fields with users requirements was heavily stressed. Future implications include the development and adoption of best practice guides for improving internal workflow systems, and further consideration of end-users.

Sarah Saunders, Chairman, Electric Lane, shared her professional experience to give sound justification for good data management, and a visual demonstration clarifying mapping procedures. Saunders identified problems for mapping in the initial choice of standard and cataloguing protocols. Problems include the appropriateness of the standard or schema for the job it is going to do; technical differences – such as systems not entering dates before 1601 or 1904, circa dates, partial dates, or reading dates correctly; the image may be detached from its accompanying metadata or conversely there may be metadata but no image; there may be partial metadata; key metadata may be missing, e.g. core metadata such as copyright and source metadata; and there may be a lack of clarity within the metadata. Just to further complicate the issue, ambiguities are often found between the metadata describing the original artwork and the metadata describing the digital image. Such inconsistencies present problems for access before one even starts considering interoperable systems.

Saunders created a colour-coded mapping framework for key standards used by the image industry (Dublin Core, VRA Core, CDWA Lite and IPTC). This was based on the Getty Institute’s respected Crosswalks. She demonstrated how field names (metadata elements) and their descriptions can differ between standards but went on to show how, even with differences, it is possible to map from one standard to another – and how confusing this can be! Saunders illustrated how, due to the lack of consistency across standards and their use, mapping standards can result in a user interface full of ambiguities.

Questions from the audience included an important point made by Stella Dextre Clarke, that when developing mapping we need to be able to map both ways. Stella asked “how this was going to be done?” Saunders agreed that this was important but made no suggestions.

Niamh Brennan and Garret McMahon, Trinity College Dublin, gave a presentation on how to increase points of access to image collections - a guide to metadata harvesting.

The emphasis of the presentation was interoperability -

"….the ability of systems, services and organisations to work together seamlessly toward common or diverse goals".

Brennan talked about Open Archives Harvester PMH, one of a number of free options facilitating the discovery of and access to images via an internet portal. Harvesters such as OAI allow any image repository using open standards and metadata schemas to create a searchable index of metadata. Harvesting is the process of collecting metadata from different repositories (a network-accessible server that opens up metadata to harvesters). In order to harvest core data from several different information systems resources a common format is needed, this may be based on a standard. OAI Harvester PMH is based on the Open Archives Initiative standard. Once metadata has been harvested, it can be stored in a "union" environment which links it back to both the fuller metadata record, and any related resources.

Users have to conform to certain rules, usually specified field names from a particular standard, to be able to subscribe. Problems with metadata harvesting are similar to mapping problems and include duplication of information and fields producing ambiguous results, and a lack of consistency in results between harvesting systems. If harvesters are to be useful to the image industry, agreement on a core set of metadata elements would need to be based on an appropriate standard such as Dublin Core.

Future recommendations for harvesting include promoting agreement on the same terms, where possible - this will improve standardisation and mapping – removing duplication of terms, and providing intuitive understanding and re-labelling of data.

Brennan's presentation was followed by Robina Clayphan, Europeana's metadata expert. Clayphan gave a concise and factual overview describing Europeana – a portal developed to ensure a common and multilingual access to Europe's libraries, archives and museums. Europeana is the largest EC funded projects aimed at delivering cultural access via portals, aggregating cultural heritage portals from smaller cultural projects across Europe. Europeana's vision encompasses an alliance of Europe's cultural heritage and information associations; engagement with bodies that represent domain memberships ensures that Europeana is endorsed at the highest level by their associations, This include bodies to promote cooperation and interoperability such as

- Audio visual collections
- Including film, video, TV and sound
- museums and archives from national to local level
- Library networks.

Clayphan went on to talk about metadata and access to Europeana. All organisations with digitised collections, or with an interest in sharing knowledge and expertise, are welcome to join the network. Europeana's team has developed a Dublin Core-based application profile with specially developed extensions (ESE), which is currently in use on the beta version of their website. Europeana harvests metadata using the OAI-PMH protocol and its development included a huge amount of manual mapping. Providers are required to offer metadata which conforms to ESE.

Clayphan gave her general advice for preparing metadata for interoperability:

- Know what you are describing – one-to-one model
- Conform to a metadata standard
- Ensure consistency in applying the standard to your data
- Know how your standard related to other standards
- Use recognised, publisher controlled vocabularies and standard encodings for dates etc.
- Granularity issues – when sharing data it is easier to dumb down that to dumb up!
- Consider the functions you want to support
- Keep interoperability in mind

Europeana is still in its prototype phase, as demonstrated during the website launch and subsequent failure of the portal due to the unforeseen number of hits received during its first hour of public operation. A large network of aggregator portals is foreseen as driving Europeana forward, by groups of institutions getting together to aggregate their data and create either sector or thematic websites which will then be harvested by Europeana. At this stage it is hard to make concrete observations about Europeana, although in theory its potential is huge. However, whether such a cumbersome machine will ever facilitate anything more than superficial and monolinguistic access to digitised images remains to be seen. The onus is firmly on users to provide good interoperable metadata, and as we have seen, much needs to be done to achieve this.

On the second day of the conference, three presentations from software developers and technology providers demonstrated their company's latest advances, and the variety of products available for image asset management and search & retrieval tools, with talks from Abbie Enock of Capture, Dr Chris Town of Imense, and Alec Turner from System Simulation (who are part of the MILE partnership).

First was Abbie Enock, Director of Capture. Capture is a UK producer and developer of fully integrated image and digital asset management systems which incorporate an image database/browser with web capabilities to enable images to be uploaded, captioned, searched, viewed and marketed over the Internet. Enock demonstrated how the Capture systems, through series of flexible modules, can be adapted to suit almost all needs - from a large picture library requiring sophisticated image management, sales and marketing tools, to a single user who just needs to caption and view images or digital assets in a database; from a corporate or government agency to an image production company. The elements from each package can be mixed and matched to produce a flexible and cost effective system. She described their recent partnerships with both PLUS and Imense, both emerging industry organisations which are developing new technologies and systems for streamlining workflow processes.

Capture is currently working to combine its workflow systems with Imense's image recognition and retrieval technology and PLUS's universal picture licensing system. Both partnerships aim to maximise cost-efficiency and streamline workflow for both image providers and their end users.

Appropriately, Enock's presentation was followed by Dr Chris Town on Imense's technology. As in his previous presentation at MILE's *Speaking in Tongues* conference, Dr Town explained that the software works using ontologies to link text – such as keywords, tagging, search terms – to the relevant actual image content. It also uses methods such as probabilistic relevance assessment, spatial queries and auto-tagging. This automated annotation has great advantages over textual search & retrieval, because it reduces costs, increases accuracy and consistency, and is not tied to any particular language. Automated image retrieval such as that offered by Imense can offer automated image analysis that is;

- Automated, therefore lower costs
- Consistent and probabilistic
- Language independent (concepts)
- Clear semantics
- Spatial localisation
- Object specific
- Relevance is quantifiable

Dr Town demonstrated Imense's search results using text search & retrieval with two sample web searches for simple items such as 'purple flower', and Imense's technology, and gave diagrammatic illustrations of the software's retrieval process. And as before, Town points out that Imense would work alongside human cataloguers to help make their lives easier. It can locate shapes, colours, and simple designs, but as Town did not demonstrate any results using art images it was difficult to see how well this technology would work for many of the organizations represented by the delegates present, a mixture of cultural heritage and commercial image databases.

The session closed with Alec Turner's demonstrations of two asset management systems which System Simulation has developed – ImageIndex + and AssetIndex + – and further information on their work developing a more intuitive way of working with digital images through their participation in FABRIC project. Image and Asset Index + are useable tools aimed at photographers working with their images at digitisation source. The great strength of these systems is their simplicity and flexibility. FABRIC is an ongoing project geared at creating a digital swatchboard of fabric samples for designers to work with, and although still in development so far is receiving positive responses from surveyed designers.

Overall, emerging trends for mapping can be seen in current attempts to expand existing crosswalks to include new standards, and in projects such as Europeana. Alternatives to mapping include the development of mono-protocol searching, metadata harvesting and metadata searching.

Metadata Mapping Exercise

The final part of the day was dedicated to a workshop introduced by Dr Richard Butterworth, Metadata researcher, Project SILVER, and co-hosted by Robina Clayphan. A mapping exercise devised and led by Sarah Saunders, and assisted by MILE and BAL teams as group leaders, was designed to generate discussion around common problems when mapping images - which do not easily fit into predetermined fields - from one existing standard to another. Saunders chose the four standards most commonly used by commercial image libraries – CDWA Lite, IPTC Core and Extended Versions, VRA Core and Dublin Core. From these, the delegates divided into seven subgroups as follows;

- **VRA Core Digital Image**
- **VRA Core Art Work**
- **Dublin Core Digital Image**
- **Dublin Core Art Work**
- **IPTC Core**
- **IPTC Extended Version**
- **CDWA Lite**

Each group was given a set of cards with the corresponding field names for their standard, and Saunders then listed a series of brief field descriptions. For example, the generic description *Rights* was found to correspond to *Copyright Notice* in IPTC, *Rights (Artwork)* in Dublin Core, *Rights* in VRA Core and *20. Rights for Work* in CDWA Lite. The full set of fields for each schema are attached in **Appendix C**.

This exercise demonstrated the difficulties of trying to map between just four standards, and it delivered some interesting results. The aim was to produce two mappings; one as a direct result of the workshop exercise during the conference, and one draft mapping to show a core set of field elements for use over *all* standards, not just the four used in the exercise. MILE's aim for this Work Package is to produce a core set of field elements which has been agreed upon by MILE partners and selected high-level industry professionals as a definitive set. So the resultant overall mapping from this exercise is displayed as **Appendix A**, and the first draft of these core fields is attached as **Appendix B**, and will also be published on MILE's website. Fields from each schema which did not correspond to any of the field names in the exercise are listed underneath each schema's individual results.

The Artwork schemas (VRA, CDWA) are rich in areas relating to the artwork, and low on terms to describe the digital image and its related rights. IPTC is rich in terms to describe the digital image and has limited terms to describe the artwork, while Dublin Core has broad field titles which need qualifying for both artwork and digital image use.

IPTC

The IPTC standard was developed by the ubiquitous organisation, the International Press Telecommunications Council. IPTC Photo Metadata was created to describe and administrate stock photographs, and to provide the most relevant rights related information. This allows visual content to be easily accessible by human language terms

or machine readable codes. There are currently three IPTC standards; IPTC IIM (Information Interchange Model), IPTC Core (now developed into IPTC Photo Metadata 2008 standard), and IPTC Extension - a supplementary standard to be used with IPTC Core which provides new additional fields and refines some existing ones. Saunders chose IPTC Core and Extended versions, as these are used together to create a superstandard.

IPTC has integrated technological advances by including the field *Globally Unique Identifier*, a number which should stay with the image through all suppliers. This is created at source by the camera or scanner. It is *not* usually the suppliers' reference number; the *Supplier Picture Number* is the correct field to use for this. Spare fields were left over from IPTC Core only; these were *Title, Location Created,* and *Source*.

DUBLIN CORE: Art Work* and *Digital Image

The Dublin Core Metadata Element Set is a vocabulary of fifteen properties for use in resource description, so-named 'core' because its elements are supposed to be generic enough for describing a wide range of resources. However this does of course conversely mean that the broadness of the field names does not always make it appropriate or easily applicable for art image uses. As the delegates found, there was some confusion about the definition of the following fields: *Contributor*, *Type*, *Publisher*, *Language*, and *Relation*. It is worth noting that Dublin Core is often used - as by Europeana - by adding qualifiers to the fields. A number of fields were mapped across to *Keywords* and *Search Terms*; these could be mapped from the relevant fields such as *Description*, *Language*, *Subject*. *Creator* and *Contributor* were used as alternative creators; perhaps *Contributor* would be an additional layer of *Creator*? *Description* was used for *Headline*, and delegates suggested that *Title* is better as a short description. It is also important to note that the physical location of the artwork and of the digital image may be different, as in *Source* (artwork) and *Source* (digital image).

As with IPTC above, the main message is still the importance of using consistent standards and mapping rules. Spare fields left over from ***DUBLIN CORE Digital Image*** were *Coverage, Format,* and *Relation*. Spare fields left over from **DUBLIN CORE Artwork** were *Contributor, Type, Publisher, Language,* and *Relation*.

VRA CORE: Art Work and Digital Image

Like IPTC, VRA Core was developed by an interested organisation – the Visual Resources Association. VRA Core is a data standard for the cultural heritage community which consists of a metadata element set, as well as an initial blueprint for how those elements can be hierarchically structured. The element set provides a categorical organization for the description of works of visual culture as well as the digital images that document them. Despite this, delegates pointed out that, from a curator's point of view, VRA Core is too simplistic. However, it is possible to have authoritative, controlled multiple fields defining a core set of data for sharing, and at least VRA's elements are less confusing and therefore more easily applied than Dublin Core or IPTC were found to be, so it could be argued that this simplicity is ultimately a strength rather than a weakness. There were many spare fields left over with VRA Core; from ***VRA CORE Digital Image*** were *Measurements, Type, Location, Style/Period, Material, Technique, Relation, Source, Culture, Subject, Record Type, Description.* And from

VRA CORE Artwork were *Relation, Type, Technique, Source, Record Type,* and *Style/Period.*

CDWA Lite

CDWA Lite (Categories for the Description of Works of Art) is the fullest schema of the four within MILE's exercise, and was developed by The Getty Institute. MILE's previous conference in this Work Package ran an exercise specifically on CDWA Lite, which was led by its chief devised, world-renowned metadata classification expert Murtha Baca, so MILE delegates already had some experience of this standard. Essentially CDWA Lite is an XML schema to describe core records for works of art and material culture based on The Getty Institute's ***CDWA***, and intended to accompany the ***CCO*** (***Cataloguing Cultural Objects***). ***Cataloging Cultural Objects: A Guide to Describing Cultural Works and Their Images*** provides prescriptive guidelines for selecting, ordering, and formatting data used to populate catalogue records. It deals with information related to a subset of the CDWA Categories and the VRA Core Categories. CDWA itself describes the content of art databases by articulating a conceptual framework for describing and accessing information about works of art, architecture, other material culture, groups and collections of works, and related images. CDWA includes 512 categories and subcategories, with a small subset of core categories which represent the minimum information necessary to identify and describe a work. CDWA also includes discussions, basic guidelines for cataloguing, and examples.

CDWA Lite records are intended for contribution to union catalogues and other repositories using the Open Archives Initiative (OAI) harvesting protocol. Delegates found that the digital image categories were difficult to match, and *Rights Information* needs additional user-generated qualifiers in order to separate out different data types. Both *Display* and *Indexing* options were given in most fields; presumably the chosen option would depend on the user's internal standards? As with Dublin Core, a number of fields could be used for *Keywords* or search terms, and these could be mapped from the relevant fields such as *Description*, *Language*, and *Subject*. *Nationality* was not included under *Culture*, though this may be the best fit. Spare fields from CDWA Lite were *Resource/ID Number, Resource Type, Display State/Edition. Related Works,* and *Inscriptions*.

Overall, these individual results enabled MILE to conclude some more general facts. Firstly, that metadata mapping is important so that data can be exchanged and harvested between different sets of people. Different metadata schemas have different strengths. A precise fit is not always possible, and most organisations create metadata schemes to suit their own internal requirements.

Saunders particularly stressed the importance of the following guidelines for metadata mapping;

- know what you are describing, apply consistent standards and know how your metadata relates to other standards
- mapping in one direction is not always exactly mirrored in the reverse direction
- pragmatism should guide your metadata strategy - consider how your data will be used

- signify the terms describing the digital image as separate from those describing the artwork.

The main issue for metadata interoperability is consistent mapping to other schemas. Dumbing down from rich metadata may be necessary for harvesting and interoperability, while maintaining the rich data and still making it available in the source archive. A lightweight scheme will provide access to images and remain cost-effective.

Technology providers are becoming intrinsic in this field as it will be increasingly important to carry metadata in the image file itself, e.g. automation, Orphan Works legislation.

From the results of the mapping exercise, MILE has created the first draft of a set of core fields which could be used across the different schemas and could be carried within the image file. Simply listed, these are;

Digital image

1. Supplier picture number
2. Headline
3. Photographer / creator
4. Copyright notice
5. Credit
6. Date created
7. Licensing contact
8. Instructions
9. Rights usage

Artwork

10. Title
11. Artist / creator
12. Subject / keywords
13. Description
14. Date created
15. Medium / format
16. Rights
17. Nationality
18. Location
19. ID number
20. Size of the artwork

Draft set of core fields

Comment will be invited, particularly by specialists in the artwork schemas used for the exercise, and MILE is working with these metadata experts on refining the mapped fields to produce a set of ultimate core fields, which will then be circulated among MILE's mailing list as part of its MoU on metadata classification. The draft of these fields, mapped across the four schemas, is viewable as **Appendix B**.

CONCLUSION

It is clear that mapping is crucial to successful interoperability and improving access to digital image archives. Looking to the future, developments in harvesting, large-scale projects such as Europeana, and building on existing mappings such as Getty's Crosswalks, will all contribute to making significant headway in improving mapping of standards. However, we are still some way off reliable schema which allows the integration of all standards, a necessary tool if commercial and non-commercial archives are to be made available and compatible. It is hoped that defining a core set of elements, industry agreement via an MoU, and making recommendations for improvements at metadata schema and vocabulary levels will help to create uniform access to images – it would be a beginning. This needs to be supported by demanding service providers, well-trained data providers (cataloguers), consistency, use of appropriate standards and schemas in local systems, standard formats and protocols, and planned architectures to structure mapping.

These recommendations require a cultural shift and are unlikely to be imbued overnight and whilst mapping, in its current form is not the perfect solution, it is being proven to be perhaps the most realistically achievable and cost-effective one, and MILE's aim is that its definitive core set of fields will save organisations time and money in doing this.