

**38189**

**MILE**

**Metadata Classification Seminar 2 & Report**

[www.mileproject.eu](http://www.mileproject.eu)

<b>Deliverable number/name</b>	<i>D2.2 Metadata Classification Seminar 2 and Report</i>
<b>Dissemination level</b>	<i>Public</i>
<b>Delivery date</b>	<i>January 31<sup>st</sup> 2008</i>
<b>Status</b>	<i>Final</i>
<b>Author(s)</b>	<i>Jessica Tier, Project Manager Lucy Geering, Project Assistant</i>



***eContentplus***

This project is funded under the *eContentplus* programme<sup>1</sup>, a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable.

---

<sup>1</sup> OJ L 79, 24.3.2005, p. 1.



## **MILE METADATA CLASSIFICATION SEMINAR 2: Attendance**

9.30am – 5.00pm, Friday 18<sup>th</sup> January 2008-01-31  
 Assembly Hall, Church House, Dean's Yard, Westminster, London

<b>No.</b>	<b>Company</b>	<b>Contact name</b>	<b>Company Role</b>
1	Aisa-bcn	Xavier Castell	Director
2	Alice Badjan	Alice Badjan	Freelance
3	ALINARI	Andrea de Polo	Head of IT
4	ALINARI	Sam Minelli	R&D Engineer
5	Ann & Bury Peerless Picture Library	Ann Peerless	
6	Arcaid Picture Library	Mary Ashby	Metadata Manager
7	Archetypon	George Kordelis	Director
8	Arnolfini	Julian Warren	Archivist
9	Ashmolean Museum	Amanda Turner	Picture Library Manager
10	Ashmolean Museum	Declan McCarthy	Commercial Manager
11	BAA Aviation Photo Library	William Kolbé	Library Manager
12	BAPLA	Pauline Shakespeare	Online Editor
13	BAPLA	Sarah Saunders	Consultant / Metadata Panel Chair
14	BBC Photo Library	Esther Barry	Media Manager
15	BBC Photo Library	Jacqueline Faulkner	Media Manager
16	Bildombudsmannen	Staffan Teste	Legal Advisor / CEPIC Treasurer, IPTC Chair
17	Birmingham Central Library Photographic Collections	Jim Ranahan	Photographic Archivist
18	Brighton Museum & Art Gallery	Jenny Hand	Knowledge & Information Manager
19	Brighton Museum & Art Gallery	Jonathon Whitson	Assistant Curator (Documentation)
20	Brighton Museum & Art Gallery	Kevin Bacon	Curator of Photographs
21	Bristol University	Grant Young	
22	Bristol University	John Hargreaves	Technical Research Officer TASI
23	British Library	Velson Horie	Research Project Manager
24	Cambridge Imaging Systems	Ian Mottashed	
25	Camera Press	Paul Holliday	
26	Cardiff University	David Skilton	Research Professor in English
27	Central Office of Information	John Wigzell	IT Technician
28	Christies	Cordelia Jenkins	

Metadata Classification Seminar 2 & Report



<b>29</b>	Christies	Delegate	
<b>30</b>	Christies	Delegate	
<b>31</b>	Collections Navigator Project / Culture & Sport Glasgow	Kirsten H F Riley	Research Assistant
<b>32</b>	Culture 24 / 24 Hour Museum	Anra Kennedy	Head of Learning
<b>33</b>	Culture 24 / 24 Hour Museum	Jon Pratty	
<b>34</b>	Department for Business, Enterprise and Regulatory Reform	Chris Marshall	Manager, Publishing Industries
<b>35</b>	Department for Business, Enterprise and Regulatory Reform	Michel Woodman	Communications and Content Industries
<b>36</b>	Department for Business, Enterprise and Regulatory Reform	Tim Gettins	Digital Content Policy
<b>37</b>	DK Images	Katja Klier	DK Images Senior Search Data
<b>38</b>	DK Images	Margarita Rouki	DK Images Systems Manager
<b>39</b>	Dulwich Picture Gallery	Fulvio Rubesa	Picture Librarian
<b>40</b>	Edinburgh University Library	Dr Andrew Grout	Special Collections
<b>41</b>	Elysium Ltd / JPEG	Richard Clark	
<b>42</b>	European Commission	Christian de Bruyne	
<b>43</b>	Foto Libra	Gwyn Headley	
<b>44</b>	Fotofinder	Agnes Folaji	Collections Manager
<b>45</b>	Getty Institute	Murtha Baca	Head of Vocabulary Programme
<b>46</b>	Heini Schneebeili	Heini Schneebeili	
<b>47</b>	Heritage Images	Amanda Waite	Image Cataloguer
<b>48</b>	Hulton Getty	Karen Leach	Archival Search Data Manager
<b>49</b>	Hulton Getty	Sarah McDonald	Curator
<b>50</b>	Hungry Eye Ltd	Jenny Wicks	
<b>51</b>	Hungry Eye Ltd	Rob Brooks	
<b>52</b>	Huntington Whiteley	Magdalen Evans	
<b>53</b>	ICCE Photolibrary	Jacolyn Wakeford	
<b>54</b>	Image Source Ltd	Martin Jacobs	Keyword & Thesaurus Manager
<b>55</b>	Jaguar Daimler Heritage Trust	Karam Ram	Photographic Archivist
<b>56</b>	Jessica Strang Photo Library	Jessica Strang	
<b>57</b>	Jessica Strang Photo Library	Orlando Oliver	
<b>58</b>	JISC / King's College London	Dr Anna Bentkowska-Kafel	
<b>59</b>	John Thompson & Partners	Alice Leeves	Image Librarian
<b>60</b>	Kings College Archives	Geoffrey Browell	

Metadata Classification Seminar 2 & Report



<b>61</b>	King's College London	Malcolm Polfreman	Information Officer
<b>62</b>	Laurence King Publishing Ltd	Sue Bolsom	Picture Editor
<b>63</b>	Lee Miller Archives	Lance Downie	Digital Librarian
<b>64</b>	Lexara	Ben Hayman	
<b>65</b>	Lexara	Ben Hayman	
<b>66</b>	Lexara	Martyn Farrows	
<b>67</b>	Lexara	Martyn Farrows	
<b>68</b>	Lexara	Mike Bell	
<b>69</b>	Lexis Nexis	Diana Rose	Magazine Editor
<b>70</b>	Link Picture Library	Orde Eliason	
<b>71</b>	London College of Communication	Professor Lindsay MacDonald	Professor of Digital Media Director of Enterprise
<b>72</b>	Matrix Photos	Daniel Saunders	
<b>73</b>	MDA	Nick Poole	CEO
<b>74</b>	MediaFund	Piers Russell Cobb	Managing Director
<b>75</b>	MLA East Midlands	Caroline Moore	Digital Access Development Officer
<b>76</b>	National Army Museum	Ian Jones	
<b>77</b>	National Army Museum	Juliet McConnell	Picture Library
<b>78</b>	Peter Wilson	Peter Wilson	
<b>79</b>	Photofusion Picture Library	Liz Somerville	
<b>80</b>	Practical Pictures Ltd (part of Anness Publishing Group)	Natasha Mulder	Sales Manager
<b>81</b>	Public Catalogue Foundation	Andy Ellis	Director
<b>82</b>	Red Cover Picture Library	Mark Bezodis	Managing Director
<b>83</b>	Red Cover Picture Library	Polina Plotnikova	Technical Supervisor
<b>84</b>	Report Digital	John Harris	
<b>85</b>	Rex Features	Stuart Ellison	Keywording Coordinator
<b>86</b>	Rightscom Ltd	Francis Cave	ACAP Technical Project Manager
<b>87</b>	Royal Academy of Arts	Trine Lyngby Hougaard	Photographic Coordinator
<b>88</b>	Royal Albert Memorial Museum Exeter	Kay Holgate	Delegate
<b>89</b>	Royal Albert Memorial Museum Exeter	Rob Day	Digital Media Officer
<b>90</b>	Royal Asiatic Society	Alison Ohta	Curator
<b>91</b>	Royal Asiatic Society	Kathy Lazenbatt	
<b>92</b>	Science Museum	Chris Rowlin	Acquisitions Executive
<b>93</b>	Science Museum	David Exton	Senior Photographer
<b>94</b>	Sense World Wide	Jeremy Brown	



<b>95</b>	SILVER	Brian Kavanagh	Systems Operator
<b>96</b>	SILVER	Hilary Bates	Assistant Project Manager
<b>97</b>	SILVER	Pandora Mather-Lees	SILVER Project Manager / Managing Director (Sales)
<b>98</b>	SILVER	Richard Butterworth	Senior Metadata Analyst
<b>99</b>	SILVER	Stella Dextre Clarke	Information Consultant
<b>100</b>	Simon Conti	Simon Conti	
<b>101</b>	Sonia Halliday Photographs	Alison Floyd	
<b>102</b>	Sonia Halliday Photographs	Polly Buston	
<b>103</b>	Stiftung Preussischer Kulturbesitz	Axel Emert	Scientist
<b>104</b>	Stiftung Preussischer Kulturbesitz	Monika Hagedorn-Saupe	Deputy Director
<b>105</b>	System Simulation	Graham Howard	Creative Director
<b>106</b>	The Advertising Archives	Helen Mason	Picture Researcher
<b>107</b>	The Bowes Museum	Dr Howard Coutts	
<b>108</b>	The Bridgeman Art Library	Adrian Gibbs	Collections Manager
<b>109</b>	The Bridgeman Art Library	Annette Godefroid	General Manager German Office
<b>110</b>	The Bridgeman Art Library	Celestine Bramley	MILE Project Assistant
<b>111</b>	The Bridgeman Art Library	Felicity Page	Head Cataloguer
<b>112</b>	The Bridgeman Art Library	Gina Weston-Baker	Cataloguer
<b>113</b>	The Bridgeman Art Library	Hannah Armstrong	Controlled Vocabularies
<b>114</b>	The Bridgeman Art Library	Harriet Bridgeman	Chairman
<b>115</b>	The Bridgeman Art Library	Imogen Pasley-Tyler	Cataloguer
<b>116</b>	The Bridgeman Art Library	Jessica Tier	MILE Project Manager
<b>117</b>	The Bridgeman Art Library	Katherine Millar Craig	Cataloguer
<b>118</b>	The Bridgeman Art Library	Lucy Geering	MILE Project Administrator
<b>119</b>	The Bridgeman Art Library	Sophy Geering	PA to Chair
<b>120</b>	The Bridgeman Art Library	Steffen Wedepohl	German Rights & Research Executive
<b>121</b>	The Bridgeman Art Library	Susana Cespedes	Sales
<b>122</b>	The Bridgeman Art Library	Victoria Bridgeman	Managing Director
<b>123</b>	The British Museum	Ivor Kerlake	Photography & Imaging Manager
<b>124</b>	The British Museum Company	Beatriz Waters	Picture Library Manager
<b>125</b>	The British Museum Company	Clive Coward	Picture Library Manager
<b>126</b>	The British Museum Company	Kathryn Charles-Wilson	Picture Sales / Cataloguer
<b>127</b>	The British Postal Museum & Archive	Clare Stephens	Learning & Events Officer
<b>128</b>	The British Postal Museum & Archive	Siobhan O'Leary	
<b>129</b>	The Chelmsford Museum	Anne Lutyens-Humphrey	Keeper of Art, Chelmsford Museums



<b>130</b>	The Chester Beatty Library	Sinead Ward	Rights and Reproductions
<b>131</b>	The Courtauld Institute	Barbara Thompson	Witt & Conway Librarian
<b>132</b>	The Courtauld Institute	Matthew Percival	Senior Witt Library Assistant
<b>133</b>	The Geffrye Museum	Christine Lalumia	Deputy Director
<b>134</b>	The Geffrye Museum	Mandy Williams	Web Administrator
<b>135</b>	The Hatton Gallery	Natalie Reid	Curatorial Assistant
<b>136</b>	The Image Business	Angela Murphy	Consultant
<b>137</b>	The Interior Archive	Suki Sabine	
<b>138</b>	The Kennel Club Picture Library	Hannah Macnab	
<b>139</b>	The Kennel Club Picture Library	Karolina Hedstrom	Picture Librarian
<b>140</b>	The National Archives of Scotland	John Simmons	Image Library Manager
<b>141</b>	The National Gallery	Carol Hambleton	
<b>142</b>	The National Gallery	Charlotte Sexton	Deputy Head New Media
<b>143</b>	The National Gallery	Colin White	Head of Photographic Office
<b>144</b>	The National Gallery	Denise King	Photographic Office Manager
<b>145</b>	The National Gallery	Gillian Essam	Collection Information Manager
<b>146</b>	The National Gallery	Isabel Drummond	
<b>147</b>	The National Gallery	Marie-Therese Gramstadt	Slide Librarian
<b>148</b>	The National Gallery	Melissa Naylor	Senior Editor
<b>149</b>	The National Trust	Phillip Claris	Collections Information Manager
<b>150</b>	The National Trust Photo Library	Vicky Skeet	Photographic Resources Coordinator
<b>151</b>	The Photolibary Group	Dee Hurley	Keyword Coordinator
<b>152</b>	The Picture Desk	Anna Mosley	Art Archive Manager
<b>153</b>	The Pigmentum Project / University of Oxford	Dr Nicholas Eastaugh	
<b>154</b>	The Pigmentum Project / University of Oxford	Valentine Walsh	
<b>155</b>	The Royal Collection	Shruti Patel	Head of Photographic Services
<b>156</b>	The Royal Collection	Stephen Patterson	Head of Collections Information Management
<b>157</b>	The Royal Institution of Great Britain	Jane Harrison	Documentation Assistant
<b>158</b>	The Wallace Collection	Nell Carrington	Picture Librarian
<b>159</b>	The Wallace Collection	Phoebe Harris	Documentation Officer
<b>160</b>	The Wordsworth Trust	Jeff Cowton	Curator
<b>161</b>	Trinity College Dublin	Niamh Brennan	Programme Manager
<b>162</b>	UKIPO	Natasha Jenkins	Patent Examiner
<b>163</b>	V&A	Andrea Stern	Head of V&A Images
<b>164</b>	V&A	Michele Leong	V&A Images

Metadata Classification Seminar 2 & Report



<b>165</b>	VADS	Amy Robinson	Information Management Officer
<b>166</b>	Venice Picture Library	Sarah Quill	Freelance Photographer
<b>167</b>	Victoria and Albert Museum	Alan Seal	Head of Records and Collection Services
<b>168</b>	Victoria and Albert Museum	Douglas Dodds	Head of Central Services & Senior Curator, Computer Art
<b>169</b>	Wardynski & Partners	Ewa Ciechaniewicz-Gillabert	Lawyer, IP and Media
<b>170</b>	Willpower Information	Dr Leonard Will	Information Management Consultant



eContentplus



## *Semantics & Semiotics: Metadata Classification Seminar 2*

### **Morning Session**

- 9.00 a.m. Registration and coffee
- 9:30 a.m. Welcome, opening remarks, logistics (Jessica Tier, Harriet Bridgeman)
- 9:45 a.m. General introduction and overview; the evolving metadata landscape, and recent trends; CDWA Lite XML schema, VRA Core, and other schemas for expressing metadata for images; metadata mapping (Murtha Baca)
- 10:45 a.m. *Current Activities for Networking Museum Information in Germany and the Development of "museumdat"* (Monika Hagedorn-Saupe, Axel Ermert)
- 11:30 a.m. Coffee
- 12:00 p.m. The role of vocabularies; data content standards: CCO (Baca)
- 12.15 p.m. *Crossing the Boundaries: Interoperability between Vocabularies* (Stella Dextre Clarke)
- 12:45 p.m. Lunch

### **Afternoon Session**

- 2:00 p.m. Practical principles for metadata creation and maintenance (Baca)
- 2:15 p.m. Summary of main points from the morning; Q&A
- 2:30 p.m. Metadata harvesting; the roles of data providers and service providers (Baca)
- 3:00 p.m. Cataloguing exercise
- 3:30 p.m. Groups report on cataloguing exercise
- 4:00 p.m. General discussion: issues, challenges, practicalities (including staffing and workflow), Sarah Saunders & IPTC
- 4:45 p.m. Concluding remarks (Tier, Bridgeman); tea



## Welcome, opening remarks, logistics; introduction by Jessica Tier

Welcome to MILE – Metadata Image Library Exploitation. MILE is a 3-year project, which was proposed to the European Commission under the i2010 Digital Libraries initiative; the purpose of which is to increase the access to and exploitation of Europe’s cultural heritage.

When I wrote the proposal for this project, I knew 2 things:

- There was a lack of consistency in the use (or lack of use) of metadata standards (or sets of cataloguing rules) in different image collections, which makes interoperability between image collections difficult;
- More crucially, there was a lack of consistency in the ways metadata is added to images at home.

I was naïve and idealistic when I submitted the proposal for this project since I proposed that the aim of the project should be to find a single metadata standard to which everyone should adhere, making it more easy and time and cost-efficient for everyone to catalogue their images, increasing interoperability between image collections, and making it easier for users to find the images they want.

With a little research and a little more time talking to the people who have devoted their working lives to developing the way we add data to images, I realised that it was not possible for a single harmonised standard to be a realistic outcome.

But with a little more research, I concluded that there are a number of standards which can be interoperable with each other and that maybe if we all conformed to some rules about data entry, we would be able to increase interoperability and ultimately increase access to and use of our images.

The first seminar in the series questioned our project partners about the problems they had experienced in working with standards or their own local authority files. This is a summary of those problems. Most of the problems came down to something I have already mentioned – INCONSISTENCY. This is manifested in a number of ways;

- Local rules:
  - Any local rules / standards are not always adhered to in adding keywords to images;
  - Often there have been generations of keywords added to any one image. The words we use naturally change over time so keywords need to be constantly re-evaluated for relevance;
  - Many image archives work with 3<sup>rd</sup> party image collections which may have their own keywords attached which may neither fit a template or be relevant or authoritative;
- We can all adhere to a single metadata standard – but no single standard exists that fits everyone’s needs;
- There is concern that there are already too many standards, and something which I was careful that we avoid in this project is inventing a new standard.



There are so many out there that we need to try and develop the ones we have so that they fit our needs;

- Returning to the way we add data to our images – avoid inputting too much data in any single field.

## SEMANTICS

- Image archives often rely upon the user or image searcher to know what they are searching for. Unfortunately, many users searching for images in our archives possess a very limited knowledge so our keywording / cataloguing needs to reflect this to enable the user to find images more easily;
- Keywords do not always make connections with what is being depicted in an image e.g. the personality in an image, the geographical location that is being depicted, the relationship between characters in the image;
- One question which arose from a previous discussion was – is semantic standardisation possible?



## **General introduction and overview; the evolving metadata landscape, and recent trends; CDWA Lite XML schema, VRA Core, and other schemas for expressing metadata for images; metadata mapping by Murtha Baca**

Murtha Baca, keynote speaker for MILE's Metadata Classification conference, kicked off with a presentation on *Using Metadata Schemas to Encode and Disseminate Information and Images* – an overview of what metadata does, and the available schema for those who are relatively familiar with metadata. Baca started by clarifying her terms, demarcating the difference between data structure, value, content, and format / technical interchange standards. The range of available schemas is wide, and Baca highlighted examples of some of the better-known metadata schemas – MARC2.1, SPECTRUM, MODS, METS, VRA Core, CDWALite and Dublin Core. She stressed that some schemas are more suitable for difference cataloguing purposes than others, e.g. MARC or Dublin Core are suitable for architecture, art objects, and similar cultural works. Whereas image libraries often either adapt schema to their particular needs, or create their own in-house schemas. The bottom line is to choose the schema which best suits your material.

Baca then defines Dublin Core – currently the preferred metadata schema for OAI harvesting –, MODS (Metadata Object Description Schema) and METS (Metadata Encoding & Transmission Standard), and illustrates how a digital object would be catalogued using METS as a flow diagram.

Emerging trends in metadata schema currently in development are metadata which

- Is “schema-agnostic”;
- Is shareable and re-purposeable;
- Is harvestable;
- Is non-exclusive or cross-cultural (using a combination of standards);
- Uses controlled vocabularies and authority files;
- Is practical and economically feasible.

The main factor affecting decisions on which schemas to use can be classified overall as “visible web” vs. “deep web,” and it is important to think within these terms. Metadata is an investment because it controls institutions' main assets; therefore it's crucial to invest the time and energy necessary to create “good” – consistent, standards-based, interoperable – metadata. Good metadata is vital for both non-profit-based collections (museums) and commercial collections (image libraries). Baca particularly recommended keeping the end-user / Web user in mind when choosing and applying metadata. She also explains the role of crosswalks, and stresses how crucial they are in helping create interoperability, with an illustration.

This led Baca naturally on to talk about CDWA Lite. This is a standard for cultural objects which was developed by The Getty Institute in association with ARTstor. It is a harvestable, XML schema, initially developed by Getty and ARTstor as an authoritative, way to collect and share information on images. Compared with Dublin Core, CDWA Lite can be more flexible about the depth of the metadata which is inputted by individual users, and as it was developed specifically for works of art and



material culture is more appropriate for cataloguing cultural objects than Dublin Core. It refers to *CCO – Cataloguing Cultural Objects*, a cataloguing guide produced by Getty. Baca illustrates how data is entered using CDWA Lite and CCO with a diagram, and also shows examples of mapping, web results, and the inherent elements within CDWA Lite. The basic premise behind a successful metadata schema or standard relies on the following tenets;

- Use the minimal data which will make easy access to your collection;
- Be consistent;
- Use fields which are compliant with CCO;
- Include core data irrespective of whether it is specific to your collection ;
- But direct users back to your online collection so they can view the image in the context of your collection.

Baca then further illustrates this with an example of a Getty Museum image record which is within the ARTstor database. The incentive for using a schema like CDWA Lite – based on standards, harvestable via OAI – is fresh, authoritative, richer metadata, which can then be very easily published, disseminated, harvested and mapped. Her conclusion was that this will be worth the large time and money investment which goes in to creating such metadata.



## ***Current Activities for Networking Museum Information in Germany and the Development of "museumdat" by Professor Monika Hagedorn-Saupe and Axel Ermert***

Professor Monika Hagedorn-Saupe and Axel Ermert deliver a general overview of the current cultural cataloguing scene in Germany, before focusing on *museumdat*, a standard currently in development in Germany. Other current projects include;

EUROMUSE;

FOTOERBE;

MICHAEL;

BAM;

[www.museumsvokabular.de](http://www.museumsvokabular.de).

**EUROMUSE** – [www.euromuse.net](http://www.euromuse.net) is a web-based network listing exhibitions in major cultural institutions throughout Europe. It gives comprehensive and up-to-date information on major temporary exhibitions throughout Europe, as well as background information on the organising museums and their collections. Its search facility is extensive. It is multi-lingual, covering Danish, Dutch, English, French, German, Greek, Italian and Swedish; exhibitions can be searched by topic, country and date; museums can be searched by city, country and initial name letter. However, it does not cover all leading museums at the moment, so its results are not exhaustive.

**FOTOERBE** – [www.fotoerbe.de](http://www.fotoerbe.de) is a German, web-based photographic archive. Collections are organised by 3 controlling fields; place, time and material. They can be searched alphabetically, and newly acquired collections and archives are highlighted. Information on each record is displayed clearly and contains links for further searches. In particular there is an extensive geographical search tool; however, it is monolingual, so only useable by German speakers.

**MICHAEL** – [www.michael-culture.org](http://www.michael-culture.org), Multilingual Inventory of Cultural Heritage in Europe. Essentially a portal for major digital cultural archives and collections, this is another EC-funded project, under the ETen proposal. Through MICHAEL the user can find and explore digital collections from major museums, archives, libraries and other cultural institutions from across France, Italy and the UK. Its multilingual facility is English, French and Italian. Like Euromuse, its search facility is extensive, and results can be sorted by language, subject, spatial coverage, period, institution type, location, audience and access type, and it has a much more comprehensive range of resources than Euromuse.

**BAM** – [www.bam-portal.de](http://www.bam-portal.de), **Bibliotheken Archiven Museen**. The BAM-Portal combines digital catalogues, search tools and inventories from the eponymous three kinds of institutions (Libraries, Archives and Museums) to offer a simultaneous research for literature, records and artefacts. Funded by the German Research Foundation, its data search results are broad and can bring up unwanted results, i.e. a search for the painter Tischbein brings up several artists by that name and as well as a record for an antique 'tisch' or table. However this does include all possible



synonyms, and BAM is also available through Wikipedia results. It has no multilingual facility, and therefore is only searchable by German speakers.

**MUSEUMDAT** – museumdat is an XML schema specifically designed for cultural collections data; a format for publishing the core data of museum objects. It is in development by Museumsvokabular, a working group funded by German cultural institutions. Hagedorn-Saupe and Ermert illustrate the schema's workflow with a flowchart, from data entry to end presentation. It has 27 fields, including use, collection type, and type creation. Museumdat is already in use by BAM, DigiCult; Schleswig-Holstein Museum; Bildarchiv Foto Marburg; GBV Library Network; Stadtgeschichtliches Museum Leipzig; Ethnologisches Museum der SMB-PK; and the Deutsches Historisches Museum.

Museumdat can be used in conjunction with a newly developed controlled vocabulary format, museumvok. Museumvok is a Simple Knowledge Organisation System (SKOS),

“...a family of formal languages designed for representation of thesauri, classification schemas, taxonomies, subject-heading systems, or any other type of structure controlled vocabulary...”

<http://en.wikipedia.org/wiki/SKOS>

Downloadable thesauri, systems and worklists are available from Museumsvokabular's website, [www.museumsvokabular.de](http://www.museumsvokabular.de). Museumsvokabular are currently planning to improve retrieval by linking museum subject vocabularies, and linking these with the Subject Heading Authority File from various German scientific libraries. Hagedorn-Saupe and Ermert illustrate the elemental structure of museumvok, which has 21 fields. It can be usefully applied as a tool for editing vocabularies.

In conclusion, Germany is clearly investing in its cultural metadata on a national scale via the Museumsvokabular working group, and trying to instill consistency and compatibility throughout its systems. Creating multilingual facilities will enable these schema to be adopted internationally.



## **The role of vocabularies; data content standards: CCO by Murtha Baca**

Murtha Baca went on to talk about the importance of using controlled vocabularies and authority files to enhance access to information and images. She opened her talk with some definitions and a typology of data standards currently available.

**Vocabularies** are the grouping of terms or names. They are said to be **controlled** if they are limited to a particular domain of set of concepts, and designate a **preferred form** or heading for each concept, person, or entity represented.

**Controlled vocabularies** are organized collections of words, phrases and/or names, structured to show the relationships between terms and concepts.

**Authority files and thesauri** are types of controlled vocabularies, but a simple picklist can also be controlled.

**An authority file** is a type of controlled vocabulary, typically in electronic form that serves as a source of standardized forms of names, terms, or titles. **Authority files** include references or links from variant forms to preferred forms.

A **thesaurus** is a type of controlled vocabulary in which each entry or record is restricted to a single meaning, and in which hierarchical semantic relationships (whole/part, genus/species, equivalence and associative relationships) are explicitly expressed in the structure of the vocabulary itself.

**A folksonomy** is an assemblage of concepts represented by terms and names (called "tags"). It is not a taxonomy - an orderly classification that explicitly expresses the relationships between and among things being classified. Folksonomies are the result of social tagging. "Steve" is a social tagging project for creating folksonomies for art museum images: <http://www.steve.museum/>. If combined / mapped to structured vocabularies, folksonomies could provide enhanced access especially subject access) to visual materials. However, although the potential here is huge, especially when you consider how useful feedback from end users could be, folksonomies are also very time-consuming.

Baca then discussed problems commonly found in cataloguing and creating vocabularies. She gave some examples: a "cartonnier" may also be called a "desk", "cabinet" and a "chest" but which one is right; should they all be included in a vocabulary? Users may call the same artist by various names; misspellings are common e.g. O'Keeffe, Georgia, Georgia O'Keeffe, O'Keefe, Georgia and Steiglitz, Alfred, Mrs. There is enormous potential for confusion. Published data files may be implemented in a local collection management system where the vocabulary may be controlled by the system and populated with published vocabularies as well as local terminology but this is no panacea.

Murtha then went on to describe "*Cataloguing Cultural Objects*" (CCO) – *A Guide to Describing Cultural works and Their Images*. A data content standard that she suggests would enhance access, it is specifically intended for cataloguing cultural heritage materials and their images. It follows on the development of metadata



element sets (e.g., CDWA, VRA Core) and controlled vocabulary standards (AAT, TGN etc) specifically destined for art and cultural objects information, providing guidance for how to populate data elements or fields based on the VRA Core and CDWA elements. Baca suggested that the strength of CCO lies in the “authority files” which can provide terms not found in published authorities, including non-expert and even “wrong” terms and names.



## ***Crossing the Boundaries: Interoperability between Vocabularies*** by Stella Dextre Clarke

Stella Dextre Clarke is an independent consultant specialising in the design and implementation of knowledge structures, including thesauri, classification schemes and taxonomies. The taxonomy she developed recently for the UK e-Government Unit is part of a metadata standard that is now mandatory across the UK public sector. Dextre Clarke's presentation focussed on the need for interoperability and how it can be developed specifically at metadata schema and vocabulary levels, the importance of mapping different metadata standards to facilitate interoperability. She also identified some of the concomitant issues, benefits, and necessary future steps.

Dextre Clarke highlighted the need for consistency in order to develop interoperability, identifying two steps:

- Apply a metadata schema consistently to all records, export via a standard metadata format;
- Implement a metadata crosswalk e.g. The Getty crosswalk.

She suggested the need for interoperability at two levels:

- Between the metadata schemas, e.g.: Artist – Creator – Maker – Location – Place – Coverage – Spatial – Keywords – Subject;
- Between vocabulary terms, e.g.:  
rowing boats – rowboats - pulling boats  
gramophone records – phonograph records  
garments – clothes - clothing

She highlighted the differences between descriptions in various vocabularies and made the following suggestions for achieving interoperability;

- Apply a controlled vocabulary consistently to all your records;
- Implement a vocabulary crosswalk (i.e. map the different sets of data). However crosswalks are not so easy to find; you may have to build your own.

Finally, she emphasised the necessity for standard formats and protocols.

Dextre Clarke demonstrated how it is necessary to plan architectures to structure mapping and discussed the pitfalls of different types of mapping, in particular the potential for disrupted context. She also made the point that human input and computer support are vital for building mappings; that practical needs for interoperability are not limited to metadata schema and vocabulary, and how wide-ranging applications necessary to implementing metadata would be of benefit.

Ms Dextre Clarke concluded that in a networked world interoperability at the vocabulary level is an achievable necessity.



## ***Practical Principles for Metadata Creation and Maintenance*** by Murtha Baca

Taken from the revised edition of *Introduction to Metadata*, forthcoming from Getty Publications.

### ***1. Metadata creation is one of the core activities of collecting and memory institutions.***

Quality metadata creation is just as important as the care, preservation, display, and dissemination of collections; adequate planning and resources must be devoted to this ongoing, mission-critical activity.

### ***2. Metadata creation is an incremental process, and should be a shared responsibility.***

A metadata record may begin its life cycle as a “place holder” consisting of core data, and then be enriched as it moves through the various stages of its use within an institution. By the same token, metadata creation and management should be a shared responsibility, distributed in a practical, reasonable way throughout the appropriate units of an institution, including but not limited to staff in acquisitions, cataloging and processing units, the registrar’s office, digital asset management units, digitizing units, and conservation and curatorial departments. “Ad hoc” user-created metadata may be generated from work done by visiting researchers and scholars as well as other users, including non-expert users.

### ***3. Metadata rules and processes must be enforced in all appropriate units of an institution.***

Inefficiencies, gaps in mission-critical metadata, poor quality metadata, and negative “downstream” effects on metadata creation and workflow can be avoided by establishing and enforcing processes and procedures in all the participating units throughout an institution.

### ***4. Adequate, carefully thought-out staffing levels including appropriate skill sets are essential for the successful implementation of a cohesive, comprehensive metadata strategy.***

An adequate number of appropriately trained staff with a variety of expertise and skill sets (e.g., subject expertise, cataloging experience, technical knowledge, research skills, knowledge of rights issues, etc.) is necessary for implementation of a successful, institution-wide metadata strategy.

### ***5. Institutions must build heritability of metadata into core information systems.***

To avoid redundant data entry and lack of synchronization of metadata in core enterprise systems, and to ensure sharing of reliable, mission-critical information among the relevant units throughout the institution, interoperability for the automated transfer and validation of metadata from one core system to another must be achieved.

### ***6. There is no “one-size-fits-all” metadata schema or controlled vocabulary or data content (cataloging) standard.***

Institutions must carefully choose the appropriate suite of metadata schemas and controlled vocabularies (including collection-specific thesauri, local picklists, and so



on), along with the most appropriate cataloging standards (including local cataloging guidelines based on published standards) to best describe and provide access to their collections and other resources.

**7. *Institutions must streamline metadata production and replace manual methods of metadata creation with “industrial” production methods wherever possible and appropriate.***

Time - and labor-intensive procedures for metadata creation should be evaluated and streamlined wherever possible (e.g., creation of core records rather than exhaustive records; focus metadata work and vocabulary control on a very few core elements or access points; eliminate redundant and outdated workflows; etc.). Automated tools (e.g., use of templates, picklists, built-in thesauri, automated metadata generation or metadata mining, etc.) should be carefully researched and implemented as appropriate.

**8. *Institutions should make the creation of shareable, re-purposable metadata a routine part of their workflow.***

Creation of consistent, standards-based, continuously refreshed and updated metadata enables institutions to publish information about their collections and other resources and activities in a timely, efficient manner, and to more broadly disseminate that information through union catalogs and other “federated” resources via protocols such as the Protocol for Metadata Harvesting (OAI-PMH).

**9. *Research and documentation of rights metadata must be an integral part of an institution’s metadata workflow. This metadata should be captured and managed in an appropriate information system that is available to the all of the individuals in the organization who need to contribute to it, as well as those who need to use it.***

**10. *A high-level understanding of the importance of metadata and buy-in from upper management are essential for the successful implementation of a metadata strategy.***

Without a general understanding of principles 1-9 above on the part of the decision makers of an institution, it will be difficult if not impossible consistently to create adequate, appropriate metadata to enable access and use by core constituents (including internal users, the general public, and expert researchers).



## **Metadata harvesting; the roles of data providers and service providers by Murtha Baca**

Baca's final formal presentation focused on harvesting metadata as a way to create integrated access. She presents an overview on the three methods available for doing this and describes the processes inherent within each one –

- 1) Monoprotocol searching of records contributed to a union database,**
- 2) Metadata harvesting,**
- 3) Metadata searching.**

### **1) Monoprotocol searching of “union catalogues” built from contributed records**

Baca defined her terms; a union catalogue is a single database which gathers together data from different sources. They work with one schema, one platform and one searching protocol. There are 2 methods of contributing records to a union catalogue; one is to supply data using a standard metadata format such as Dublin Core, CDWA Lite, or the other is an informal approach using any method or a variety of metadata formats.

### **2) Metadata Harvesting**

Harvesting is the process of collecting metadata from different repositories (a repository is a network-accessible server that opens up the metadata to harvesters). Once metadata has been harvested, it can be stored in a “union” environment which links back to both the fuller metadata record, and any related resources. The goal of harvesting metadata is to open up access to digital material for sharing, publishing and archiving. The OAI-Protocol (Open Archives Initiative) for Metadata Harvesting is a mechanism for harvesting metadata from repositories.

The “Archives” referred to in the acronym “OAI” actually means digital resources, not only archival data. Some searches, such as the Arts and Humanities Data Service (AHDS) use a combination of meta-searching and metadata harvesting.

Baca then illustrated her definition with diagrams showing metadata harvesting, and goes on to define a standard called OAI-ORE (Open Archives Initiative Object Reuse and Exchange). This is a developing standard which uses a URI (a unique identifier).

### **3) Metasearching**

This is federated searching, i.e. it searches across collections, and uses a search portal. It can cover a wide variety of collections in terms of format and content, and can search many databases at once. Baca illustrated how this works with a diagram showing the process and flow of information from the initial search by a user through to the pages of web search results. Metasearching uses technical information exchange protocols, which can range from proprietary protocols within a digital library system such as ENCompass, to HTTP searching (“screen-



scraping”). Content creators and data providers do most of the work, rather than service providers. However, in order for metasearching to work, lots of people have to use it, so vendor participation is absolutely necessary. In the longer term, metasearching is more economical and involves fewer steps than the aggregating process. It will also relieve the burden on both the content creators and the aggregators. It’s still being developed, so a question of watching this space.

**Metasearching** (cross-collection searching)

Searches diverse collections

Searches in real time, therefore data is always fresh

**Harvesting**

Searches exposed databases & gathers results into one local database to be searched

Data providers must regularly post updated metadata, & service providers have to re-harvest this

In both cases access to commercial content is a real issue, and both methods involve significant mapping and vocabulary challenges. More issues regarding creating integrated access are how to convey to users the nature of what they have retrieved –

- is it a bibliographic record, archival materials, commercial data, etc?
- how to contextualise results both in terms of their source and intellectually?
- and anyway what’s it all for?

Ultimately, good metadata is key. Searching can only bring up what has been catalogued, and therefore the quality of search results depends primarily on the quality – consistency, controlled vocabularies, authority files – of the metadata. Baca nodded towards the problem of uncontrolled vocabularies, which were covered more fully by Stella Dextre Clarke, and highlights the 2 main solutions:

Data providers include variants, broader and narrower terms – this is labour-intensive and redundant; or

Service providers / aggregators employ the appropriate controlled vocabularies & thesauri as “search assistants” – this is promising and labour-saving, but not a real option at the moment.

Therefore, Baca sees the solutions as follows; service providers should be more demanding of their content providers, should always use appropriate standard schemas in their local systems and should consider adding value via services like vocabulary mapping, query expansion, vocabulary-assisted searching, user-added metadata, post-harvest subsetting, metadata enhancement, etc, etc...



## Groups report on cataloguing exercise

Further to comments and suggestions by many of the MILE network partners, the coordinators decided to incorporate a workshop style cataloguing exercise during the afternoon of the 'Semantics and Semiotics' seminar.

The 140 delegates were divided into 14 tables and each table was assigned a facilitator. Each facilitator was given the task of encouraging participation from each delegate, generating discussion and reporting any highlighted issues to the floor.

Every delegate was given the following from which to work:

Colour photocopies of works by Zoffany and Nam June Paik;  
Brief descriptions on the details of each image;  
Simple cataloguing template from the CDWA-lite schema;  
Questionnaire on individual usage of metadata standards/ schemas.

### AIMS

The purpose of the cataloguing exercise was to generate discussion around common problems when cataloguing images which do not easily fit into predetermined categories of any metadata standard or schema. We sought to take advantage of the eclectic variety of industry backgrounds represented at the seminar by considering opinions from many different viewpoints, including commercial picture archives, national art museums and galleries, stock photography libraries and professional photographers. It is widely known that cataloguing methods are dictated by users needs and that there is no metadata standard in existence which fits everyone's needs. By seeking out the specific problems 'on the ground' when cataloguing a variety of different images, we aimed to discuss practical solutions to the persons responsible for adding metadata to images on a daily basis.

### REVIEW

The images were chosen for their complex nature to encourage discussion and raise issues which any cataloguer would have come across at their desk when trying to catalogue according to any set of rules (metadata standard).

The main problems/ issues raised were as follows:

1. Confusion over what format should be catalogued. Is it the painting, the photograph, the sculpture within the photograph, the still frame within the photograph?
2. Indecision over dates which should be entered. Should it be the date of the sculpture or still frame within the photograph or the dates of the photograph itself since that too is considered a work of art?
3. Where to catalogue keywords which are useful for commercial clients such as dominating colour or mood or atmosphere?



4. Where to catalogue personality information? for example, narrative and relevance of the paintings within the painting in the case of the Zoffany and the political relevance of the sculpture within the photograph in the case of the Nam June Paik.
5. General misunderstanding of the catalogue template field headings. Some fields seemed to be similar to others but it was clear that this was not intentional but not clear how they differed.
6. Problems of subjectivity. Every cataloguer brings their own opinions to their work, thereby influencing the search results according to their own beliefs and thoughts.

The resolutions which were discussed and suggested after all of these issues were debated included:

1. Necessity of a dictionary/ glossary for the metadata standard or cataloguing rules in use. If each cataloguer is well versed in how each field should be populated, this would alleviate some confusion and dissatisfaction with that particular metadata standard or cataloguing rules. This should be augmented with solid training on how to use that given set of rules or standard.
2. It was also clear that although many image archives were in favour of using a popular standard such as CDWA lite to catalogue their works, that standard would have to be adapted/ amended according to each institution's needs to satisfy the searching requirements of their users. This creates problems in one sense because it is labour intensive for the information scientists who manage the metadata standards or cataloguing rules in each institution, but it largely solves the interoperability problem, because the data can still (in the case of CDWA) be mapped back to the standard.



## RESULTS OF THE METADATA CLASSIFICATION QUESTIONNAIRE

Questions 1 – 4 asked for personal details.

### QUESTION 5. WHAT METADATA STANDARD/S ARE YOU CURRENTLY USING?

A broad range of standards were reported to be in use including:

MARC, OWN, IPTC, DUBLIN CORE, MUSIMS, ISAD (G), SPECTRUM, DC TERMS, TEI, DDI, CDWA Lite, CDWA, CCO, URA, IEEE LOM, OAI-PMH, MARC 21, TGM, DRUER, ULAN, AAT, LCSH, AIM 25, ALINARI, UDC, AAC

Many organisations used multiple standards, whilst there were others who did not use a standard at all (*see page no.28*). “Own” standards were largely developed from published standards.

#### a. How useful would you rate the standards you use as a percentage?

Values ranged between 40% and 100% with SPECTRUM achieving the greatest use of a published standard (*see pages 5 & 6*). The median percentage for usefulness was calculated at 78.3%.

#### b. What issues / problems have you experienced with these standards?

- Not completely suitable for different media
- Creating hierarchical data
- Not enough crosswalks
- Consistency in general and specifically of metadata
- Controlled vocabularies
- Consistency of cataloguers
- Not flexible and or broad enough
- Standards too varied
- Loss of detail mapping from one standard to another
- Not enough fields
- Application of standards causes problems
- Pre-digital
- Linguistic ambiguities and limited thesaurus

There was much variety in the written comments and an awful lot of confusion with regards to understanding and perceptions about standards. This was reflected by pleas for training and more dissemination.



**c. What specific improvements / developments would you like to see made to the standards you currently use?**

- Crosswalks
- Semantic indexing i.e., by meaning not keyword
- Consistent mapping to IPTC
- Workshops
- Ability to create hierarchical data in MARC
- Creation of industry and content standard
- Consistency of data entry
- Including verbal content (contained within a work of art) in the metadata
- Controlled Vocabulary
- More fields and specificity
- Greater interoperability
- A thesaurus of standards more consistent for photographic images, not works of art
- More standardised hierarchies
- More consistency in metadata fields and keywording rules
- Interoperability
- Simplified and digital

A variety of answers were offered. Consistency, interoperability and the unification of standards were the most popular suggestions.

**d. If you do not use any standards – why not and what would it take for you to use a metadata standard to organise your data?**

- Infrastructure
- Better consistency
- One recognized standard
- Archaic databases held by organizations require better definition
- Standards able to cope with broader selection of categories e.g. western and non-western rare books, manuscripts and decorative objects

**QUESTION 6. WHAT PUBLISHED CONTROLLED VOCABULARIES ARE YOU USING?**

LCSH, OWN, IPSV, TGM, UKAT, AAT, ULAN, LSCH, TEN, HASSET, TGN, AAAF, SHIC, MARC, ALINARI, MDA ARCHAEOLOGICAL, MDA, ORGANISATIONAL CONTROL VOCAB, IPTC NEWS, UDC

AAT and ULAN were the most popular. Many people used multiple vocabularies (*see pages 7, 8, 9*). A few standards were referred to as vocabularies suggesting a lack of distinction between the metadata.



**a. How useful would you rate your controlled vocabularies as a percentage%?**

The median was 80%.

**b. What issues / problems have you experienced with these controlled vocabularies?**

- Too specialist
- Can limit indexing
- Diacritics and non -Roman scripts
- Cultural, regional and translation bias
- Lack of control
- Integration into cataloguing interface
- Interoperability
- Implementation and software
- Training
- Taxonomy

Cultural, regional and translation bias; lack of control and interoperability were the most commonly reported issues.

**c. What specific improvements / developments would you like to see made to the controlled vocabularies you currently use?**

- Better controlled
- Mapping user-generated tags to controlled taxonomy
- Integration
- Data more easily exploitable
- Introduction of translation systems
- Improved flexibility within terms
- Web service
- Easier access, real time integration
- Being flexible enough to adapt to unexpected subjects whilst remaining controlled
- Controlled vocabulary as high up the workflow as possible

Introduction of translation systems and improved access to culturally specific vocabularies were the most requested improvements.

**d. If you do not use a controlled vocabulary – why not and what would it take for you to use controlled vocabularies to organize your data?**

- A hybrid – 90% controlled and 10% not because the list is not extensive and does not contain all words needed

*This was the only answer for this section.*



## **QUESTION 7. DO YOU HAVE ANY LOCAL AUTHORITY FILES?**

Local authority files were used by all categories however, reported non-use figures were almost equal to the used figure, reasons for this were undetermined (*see page 10*).

## **QUESTION 8. IF YES, HOW DO YOU USE THEM?**

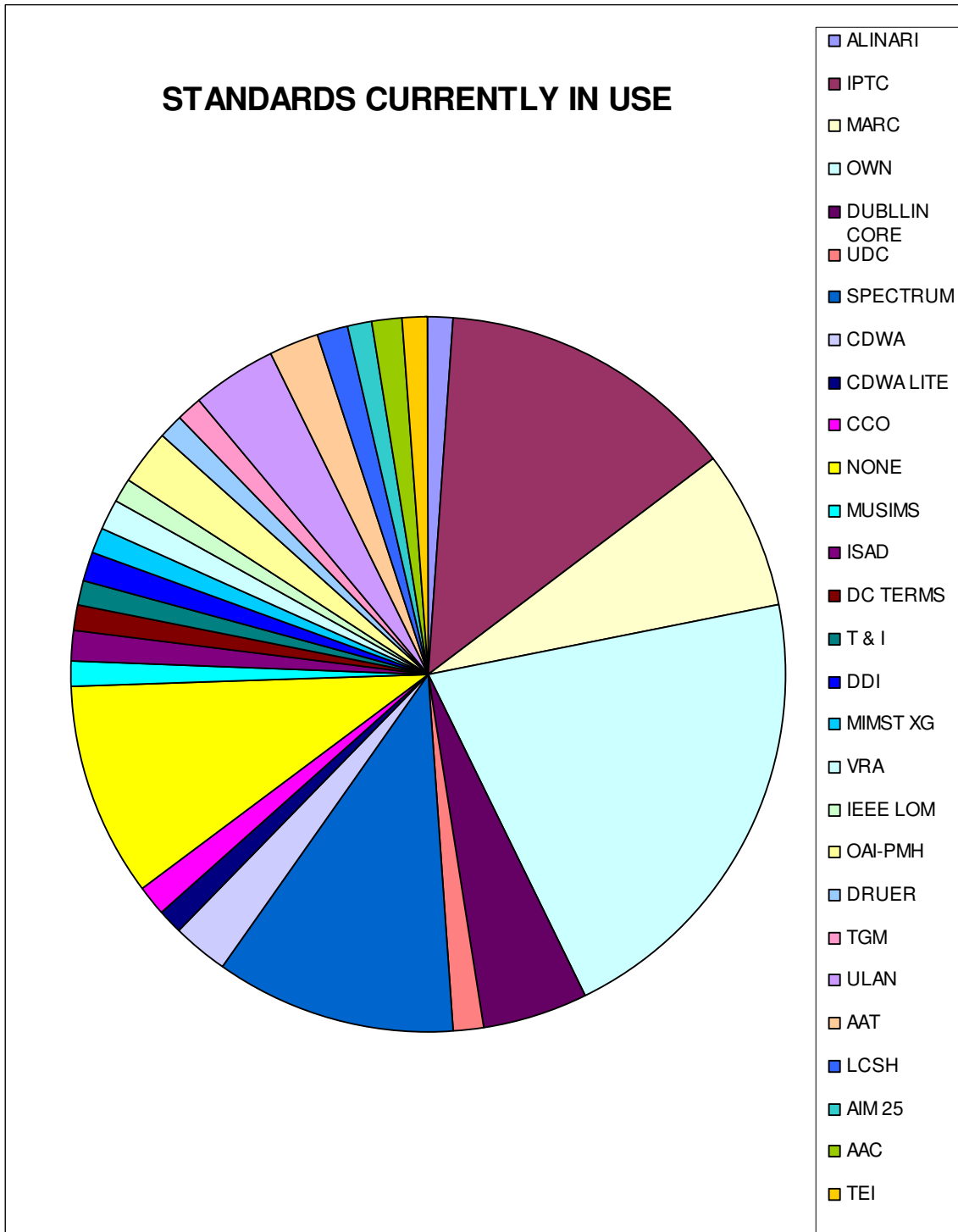
- For our database
- For sale to local authority
- Cataloguing
- Collections, names of objects, standardisation
- Repository
- Subject hierarchy
- To annotate content collections
- For preferred terms of keyword /terms and technical information on the image format
- Operate as pop-up lists / control files in data management system (MIMSY)
- Embedded in workflow
- To control cataloguing information and collection management information in collection management system
- To allow cataloguers to enter information consistently to facilitate data retrieval
- Like ULAN – related to museum index



## APPENDICES

### APPENDIX 1:

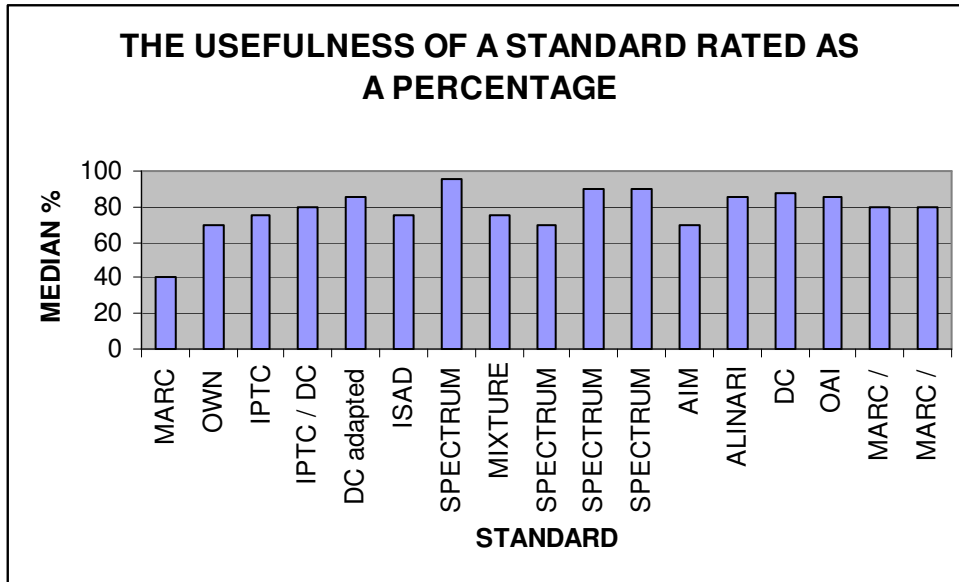
#### QUESTION 5. WHAT METADATA STANDARD/S ARE YOU CURRENTLY USING?





**APPENDIX 2:**

**QUESTION 5.a HOW USEFUL WOULD YOU RATE THE STANDARDS YOU USE AS A PERCENTAGE?**

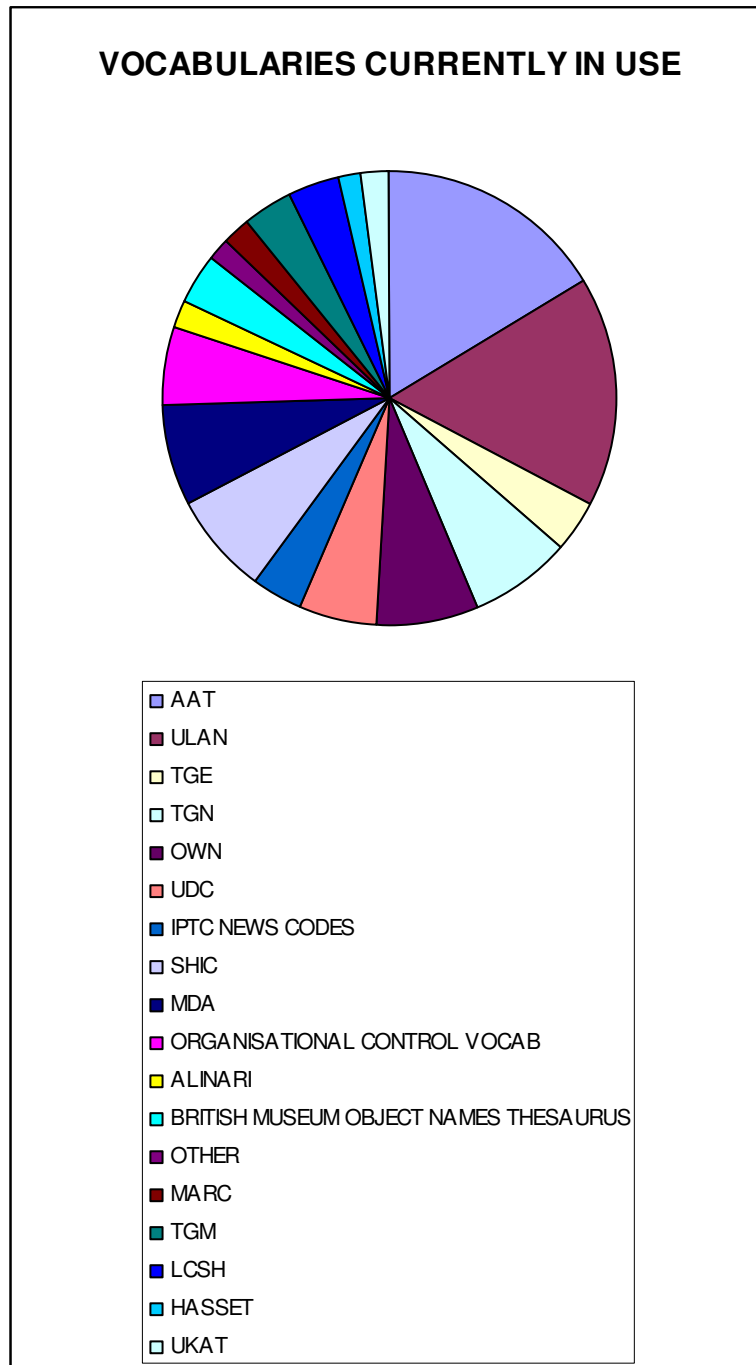




### APPENDIX 3

#### APPENDIX 3a;

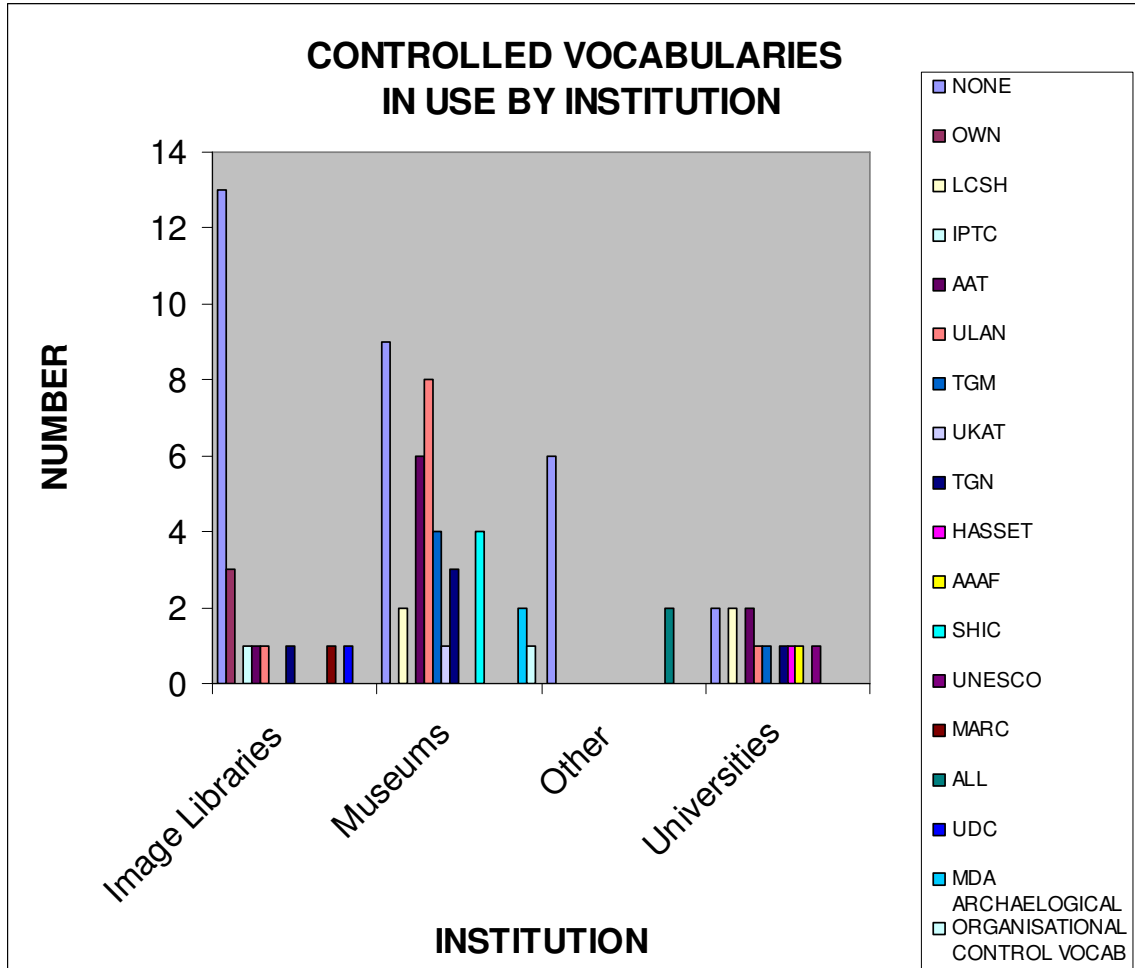
#### QUESTION 6. WHAT PUBLISHED CONTROLLED VOCABULARIES ARE YOU USING?





**APPENDIX 3.b:**

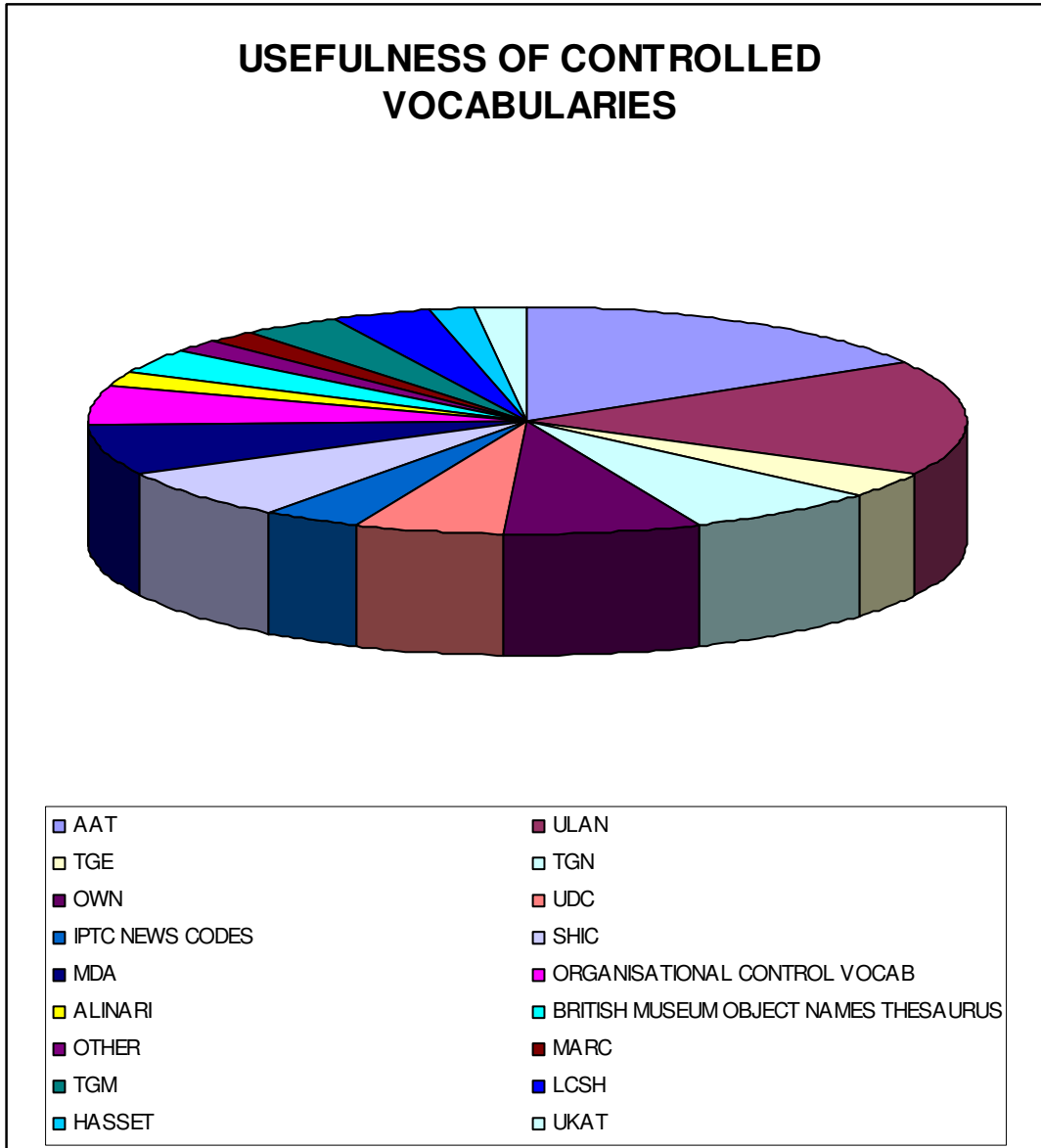
**QUESTION 6. WHAT PUBLISHED CONTROLLED VOCABULARIES ARE YOU USING? By institution**





**APPENDIX 4:**

**QUESTION 6 a. HOW USEFUL WOULD YOU RATE THE CONTROLLED VOCABULARIES YOU USE AS A PERCENTAGE?**

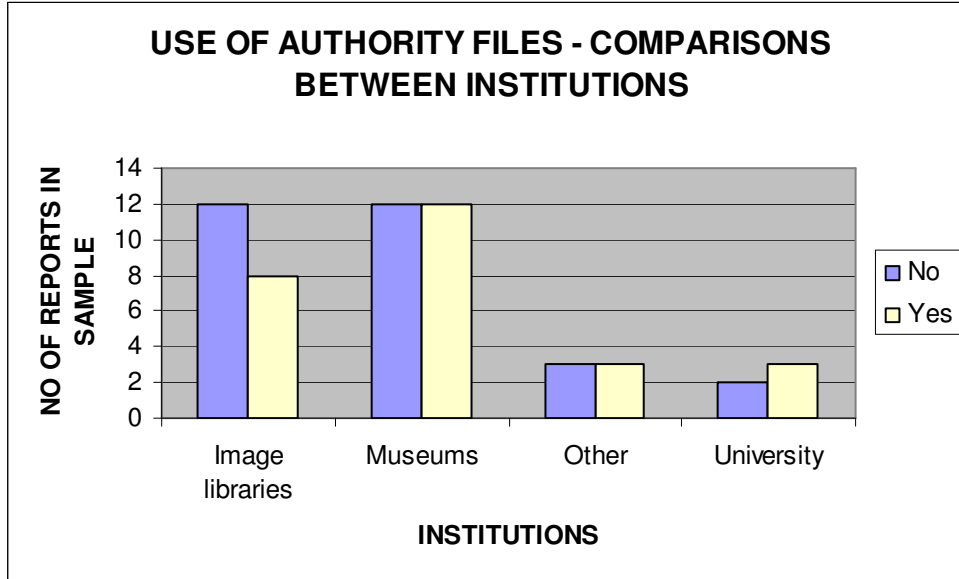


*Chart represents median percentages.*



**APPENDIX 5:**

**QUESTION 7: DO YOU HAVE ANY LOCAL AUTHORITY FILES?**





## Conclusions

The main conclusion from this conference was that the adherence to some sort of cataloguing rules or metadata standards was a necessity. This would increase labour and cost efficiency and increase the chances of interoperability. As for whether to and how to make metadata interoperable, most delegates agreed that as long as the data is clean and in organised, it can be mapped to a metadata standard. Mapping is another lengthy and expensive process which is why it is advisable for any image archive considering the use of an interoperable standard, reviews and chooses that standard very carefully.