Praise for *Metadata in the Digital Library*

‘Gartner’s latest book is written with his characteristic clarity and authority. He defines the concepts of metadata in an engaging and natural way and his book bridges the gap between theory and practice by using examples from his own work at the Warburg Institute Library. He has taken a fresh approach to explaining how XML works, as well as widely used metadata standards such as Dublin Core and METS. The chapters on Metadata Management (Chapter 3) and the Semantic Web (Chapter 10) are particularly informative. This quality of writing comes from a deep understanding of the subject coupled with direct experience of day-to-day management and use of metadata. *Metadata in the Digital Library* will be a valuable resource for postgraduate students studying information and knowledge organisation at our University.’

Dr David Haynes, *School of Computing, Edinburgh Napier University*

‘This book is a fantastic blend of technical detail, disciplinary expertise and introductory information. A general description and discussion about metadata contextualises more detailed analyses of good practice guidelines for creating interoperable metadata. While the book is most relevant to digital libraries, the concepts and standards discussed are relevant across the GLAM sector. Readers new to the adventures of metadata for digital collections will feel challenged (in the best possible way) and experts will have an authoritative reference they can use to support their work. I thoroughly enjoyed reading this book and plan on adding it to my modules’ core reading lists.’

Kristen Schuster, *Lecturer in Digital Curation, King’s College London*
Metadata in the Digital Library
Every purchase of a Facet book helps to fund CILIP’s advocacy, awareness and accreditation programmes for information professionals.
Metadata in the Digital Library

Building an Integrated Strategy with XML

Richard Gartner
Contents

Figures and tables xi
Acknowledgements xiii
List of abbreviations xv

1 Introduction, Aims and Definitions 1
   1.1 Origins 1
   1.2 From information science to libraries 2
   1.3 The central place of metadata 4
   1.4 The book in outline 6

2 Metadata Basics 9
   2.1 Introduction 9
   2.2 Three types of metadata 10
       2.2.1 Descriptive metadata 10
       2.2.2 Administrative metadata 14
       2.2.3 Structural metadata 18
   2.3 The core components of metadata 19
       2.3.1 Syntax 19
       2.3.2 Semantics 20
       2.3.3 Content rules 21
   2.4 Metadata standards 23
   2.5 Conclusion 23

3 Planning a Metadata Strategy: Basic Principles 25
   3.1 Introduction 25
   3.2 Principle 1: Support all stages of the digital curation lifecycle 25
   3.3 Principle 2: Support the long-term preservation of the digital object 27
   3.4 Principle 3: Ensure interoperability 29
   3.5 Principle 4: Control metadata content wherever possible 30
   3.6 Principle 5: Ensure software independence 31
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Principle 6: Impose a logical system of identifiers</td>
<td>32</td>
</tr>
<tr>
<td>3</td>
<td>Principle 7: Use standards whenever possible</td>
<td>32</td>
</tr>
<tr>
<td>3</td>
<td>Principle 8: Ensure the integrity of the metadata itself</td>
<td>33</td>
</tr>
<tr>
<td>3</td>
<td>Summary: the basic principles of a metadata strategy</td>
<td>34</td>
</tr>
<tr>
<td>4</td>
<td>Planning a Metadata Strategy: Applying the Basic Principles</td>
<td>37</td>
</tr>
<tr>
<td>4</td>
<td>Introduction</td>
<td>37</td>
</tr>
<tr>
<td>4</td>
<td>Initial steps: standards as a foundation</td>
<td>37</td>
</tr>
<tr>
<td>4</td>
<td>4.2.1 ‘Off-the shelf’ standards</td>
<td>38</td>
</tr>
<tr>
<td>4</td>
<td>4.2.2 Mapping out an architecture and serialising it into a standard</td>
<td>40</td>
</tr>
<tr>
<td>4</td>
<td>4.2.3 Devising a local metadata scheme</td>
<td>43</td>
</tr>
<tr>
<td>4</td>
<td>4.2.4 How standards support the basic principles</td>
<td>45</td>
</tr>
<tr>
<td>4</td>
<td>Identifiers: everything in its place</td>
<td>45</td>
</tr>
<tr>
<td>5</td>
<td>XML: The Syntactical Foundation of Metadata</td>
<td>53</td>
</tr>
<tr>
<td>5</td>
<td>Introduction</td>
<td>53</td>
</tr>
<tr>
<td>5</td>
<td>What XML looks like</td>
<td>54</td>
</tr>
<tr>
<td>5</td>
<td>XML schemas</td>
<td>56</td>
</tr>
<tr>
<td>5</td>
<td>Namespaces</td>
<td>57</td>
</tr>
<tr>
<td>5</td>
<td>Creating and editing XML</td>
<td>60</td>
</tr>
<tr>
<td>5</td>
<td>Transforming XML</td>
<td>62</td>
</tr>
<tr>
<td>5</td>
<td>Why use XML?</td>
<td>65</td>
</tr>
<tr>
<td>6</td>
<td>METS: The Metadata Package</td>
<td>67</td>
</tr>
<tr>
<td>6</td>
<td>Introduction</td>
<td>67</td>
</tr>
<tr>
<td>6</td>
<td>Why use METS?</td>
<td>67</td>
</tr>
<tr>
<td>6</td>
<td>The METS architecture</td>
<td>69</td>
</tr>
<tr>
<td>6</td>
<td>Identifiers within METS</td>
<td>70</td>
</tr>
<tr>
<td>6</td>
<td>The METS root element</td>
<td>71</td>
</tr>
<tr>
<td>6</td>
<td>The METS Header</td>
<td>72</td>
</tr>
<tr>
<td>6</td>
<td>Descriptive Metadata Section</td>
<td>73</td>
</tr>
<tr>
<td>6</td>
<td>Administrative Metadata Section</td>
<td>74</td>
</tr>
<tr>
<td>6</td>
<td>The File Section</td>
<td>77</td>
</tr>
<tr>
<td>6</td>
<td>The Structural Map</td>
<td>82</td>
</tr>
<tr>
<td>6</td>
<td>Structural Links and Behavior Section</td>
<td>85</td>
</tr>
<tr>
<td>6</td>
<td>Creating and using METS in practice</td>
<td>87</td>
</tr>
<tr>
<td>7</td>
<td>Descriptive Metadata: Semantics</td>
<td>89</td>
</tr>
<tr>
<td>7</td>
<td>Introduction</td>
<td>89</td>
</tr>
<tr>
<td>7</td>
<td>Dublin Core</td>
<td>89</td>
</tr>
</tbody>
</table>
7.3 MODS – the Metadata Object Description Schema 93
7.4 MARCXML 99
7.5 Other descriptive metadata standards 101
  7.5.1 VRA Core 102
  7.5.2 Text Encoding Initiative P5 Manuscript Description 102
  7.5.3 Schemas from the sciences and social sciences 103
  7.5.4 Using these schemas 103
7.6 Descriptive metadata: from semantics to content rules 104

8 Descriptive Metadata: Content Rules 105
  8.1 Introduction 105
  8.2 Why content rules are needed 105
  8.3 Cataloguing rules 107
    8.3.1 Established standards for cataloguing rules 108
  8.4 Devising local guidelines 110
  8.5 Controlled vocabularies 117
    8.5.1 Name authorities 118
    8.5.2 Subjects 121
    8.5.3 Codes and dates 123
  8.6 Creating local name authorities and thesauri: the MADS schema 124

9 Administrative and Preservation Metadata 127
  9.1 Introduction 127
  9.2 PREMIS: an overview 129
    9.2.1 Technical metadata: the PREMIS Object entity 130
    9.2.2 Digital provenance metadata: the PREMIS Event and Agent entities 133
    9.2.3 Rights metadata: the PREMIS Rights entity 136
    9.2.4 Using PREMIS with METS 138
  9.3 Other useful schemas for administrative metadata 139
    9.3.1 Technical metadata 139
    9.3.2 Rights metadata 143
  9.4 How much administrative metadata do we need? 145

10 Pathways to Interoperability 147
  10.1 Introduction 147
  10.2 Exchanging METS files 148
  10.3 Metadata harvesting 151
  10.4 The Semantic Web 153
  10.5 Conclusion 157
11 Implementing the Strategy: Two Case Studies 159
11.1 Introduction 159
11.2 The Warburg Digital Library 159
11.3 The Warburg Iconographic Database 166
11.4 Conclusion 172

12 Summary and Conclusions 173
12.1 Introduction 173
12.2 The strategy in outline 173
  12.2.1 The basic principles of a metadata strategy 173
      (Chapter 3)
  12.2.2 Standards and identifiers (Chapter 4) 174
  12.2.3 Syntax (Chapter 5) 175
  12.2.4 Packaging the metadata (Chapter 6) 175
  12.2.5 Descriptive metadata (Chapter 7) 175
  12.2.6 Content rules (Chapter 8) 176
  12.2.7 Administrative and preservation metadata
      (Chapter 9)
  12.2.8 Enabling interoperability (Chapter 10) 178
12.3 Conclusions 178

Appendix: Sample MODS File Serialised from Data Model 181

Useful Resources 183
  Key metadata standards referenced 183
  Supplementary documentation and resources 185

Further Reading 189

References 195

Index 197
Figures and Tables

Figures
2.1 Descriptive metadata in a standard catalogue card 10
2.2 Descriptive metadata in an online catalogue 11
2.3 Descriptive metadata in the Warburg Institute's Iconographic Database 12
2.4 Descriptive metadata from the video streaming service MUBI 13
3.1 DCC Curation Lifecycle Model 26
3.2 Open Archival Information System model 28
4.1 Inputting Dublin Core metadata into Omeka 38
4.2 Inputting MODS metadata into Islandora 39
4.3 Metadata map (highly simplified) for a digitised photograph of a work of art 41
4.4 Map in Figure 4.3 populated with metadata 42
4.5 Structure of Warburg Digital Library file system based on system of internal identifiers 49
5.1 Editing an XML file in Oxygen (inserting an element) 61
5.2 Creating an XML schema in Oxygen 61
6.1 The seven major subdivisions of a METS file 70
6.2 Title page of digitised book 70
6.3 The four subdivisions of the METS Administrative Metadata Section 75
6.4 Using the METS Structural Map and File Section to display the image of a page 86
7.1 The 15 fields of Simple Dublin Core 90
8.1 Oxford Digital Library cataloguing guidelines: introduction 112
8.2 Oxford Digital Library cataloguing guidelines: Title 113
8.3 Oxford Digital Library: summary of AACR2 rules for Title field 114
8.4 Oxford Digital Library Name Authority Procedures 115
8.5 Oxford Digital Library: Personal names authority list 116
8.6 LCNAF entry for a personal name 118
8.7 Entry for composer Bonaventura Aliotti on VIAF 120
METADATA IN THE DIGITAL LIBRARY

8.8 LCSH entry for Oratorios (beginning) 122
9.1 The four top-level entities of the PREMIS data model 129
10.1 Simple RDF ‘triple’ 153
10.2 The same RDF ‘triple’ with URIs for subject, predicate and object 154
11.1 The Warburg Digital Library home page 161
11.2 Browsing through a book in the Warburg Digital Library 161
11.3 Metadata generation workflow for the Warburg Digital Library 162
11.4 Data model, based on CIDOC-CRM, for the Warburg Iconographic Database 168
11.5 Mapping of central concepts of data model to XML metadata standards 170
11.6 The chain of references that map out the hierarchy for an entry in MADS 171

Tables
2.1 Definition of ‘subject’ in four metadata schemes 21
7.1 Simple Dublin Core record for oratorio libretto 91
Acknowledgements

I would like to acknowledge my friends and colleagues at the Warburg Institute, King’s College London, Dr Williams’s Library and the Bodleian Library who have provided me with such support and guidance over many years. The same thanks are due to many other colleagues here and abroad, including my friends on the METS Editorial Board with whom I have had the pleasure of working for almost 20 years.

I would also like to thank Pete Baker and Michelle Lau at Facet for their help in developing this book and the effortless way that they steered me through to its completion.
Abbreviations

AACR2 Anglo-American Cataloging Rules, Second Edition
AIP Archival Information Package
AVI Audio Video Interleave
CIDOC-CRM CIDOC (International Council of Museums International Committee for Documentation) Conceptual Reference Model
DACS Describing Archives: A Content Standard
DC Dublin Core
DCC Digital Curation Centre
DCMI Dublin Core Metadata Initiative
DIDL Digital Item Declaration Language
DIP Dissemination Information Package
FRBR Functional Requirements for Bibliographic Records
ID identifier
IPR intellectual property rights
ISAD(G) General International Standard Archival Description
ISO International Organization for Standardization
JPEG Joint Photographic Experts Group
LCNAF Library of Congress Name Authority File
LCSH Library of Congress Subject Headings
MADS Metadata Authority Description Schema
MARC MACHine Readable Cataloguing
METS Metadata Encoding and Transmission Standard
MIX NISO Metadata for Images in XML Schema
MODS Metadata Object Description Schema
MP3 Moving Pictures 3 (audio format) MPEG (Motion Picture Experts Group) Layer-3
MPEG Moving Picture Experts Group
OAI-PMH Open Archives Initiative Protocol for Metadata Harvesting
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>OAIS</td>
<td>Open Archival Information System</td>
</tr>
<tr>
<td>ODL</td>
<td>Oxford Digital Library</td>
</tr>
<tr>
<td>PREMIS</td>
<td>PREservation Metadata: Implementation Strategies</td>
</tr>
<tr>
<td>RDA</td>
<td>Resource Description and Access</td>
</tr>
<tr>
<td>RDF</td>
<td>Resource Description Framework</td>
</tr>
<tr>
<td>SGML</td>
<td>Standard Generalised Markup Language</td>
</tr>
<tr>
<td>SIP</td>
<td>Submission Information Package</td>
</tr>
<tr>
<td>TIFF</td>
<td>Tagged Image File Format</td>
</tr>
<tr>
<td>URI</td>
<td>Uniform Resource Identifier</td>
</tr>
<tr>
<td>URL</td>
<td>Uniform Resource Locator</td>
</tr>
<tr>
<td>XML</td>
<td>eXtensible Markup Language</td>
</tr>
<tr>
<td>XPath</td>
<td>XML Path Language</td>
</tr>
<tr>
<td>XSLT</td>
<td>eXtensible Stylesheet Language Transformations</td>
</tr>
<tr>
<td>W3C</td>
<td>World Wide Web Consortium</td>
</tr>
<tr>
<td>WAV</td>
<td>Waveform Audio File Format</td>
</tr>
<tr>
<td>WDL</td>
<td>Warburg Digital Library</td>
</tr>
<tr>
<td>W3C</td>
<td>World Wide Web Consortium</td>
</tr>
<tr>
<td>WMV</td>
<td>Windows Media Video</td>
</tr>
<tr>
<td>XML</td>
<td>eXtensible Markup Language</td>
</tr>
<tr>
<td>XPath</td>
<td>XML Path Language</td>
</tr>
<tr>
<td>XSLT</td>
<td>eXtensible Stylesheet Language Transformations</td>
</tr>
</tbody>
</table>