METADATA in the DIGITAL LIBRARY

Building an Integrated Strategy with XML

RICHARD GARTNER

สำนักหอสมุด มหาวิทยาลัยเชียงใหม่

Metadata in the Digital Library

Building an Integrated Strategy with XML

สำนักหอสมุดมหาวิทยาลัยเชียงใหม่
CHIANG MAI UNIVERSITY LIBRARY

Richard Gartner

facet publishing

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

3.7 Principle 6: Impose a logical system of identifiers 32
3.8 Principle 7: Use standards whenever possible 32
3.9 Principle 8: Ensure the integrity of the metadata itself 33
3.10 Summary: the basic principles of a metadata strategy 34

4 Planning a Metadata Strategy: Applying the Basic Principles 37
4.1 Introduction 37
4.2 Initial steps: standards as a foundation 37
4.2.1 'Off-the shelf' standards 38
4.2.2 Mapping out an architecture and serialising it into a standard 40
4.2.3 Devising a local metadata scheme 43
4.2.4 How standards support the basic principles 45
4.3 Identifiers: everything in its place 45

5 XML: The Syntactical Foundation of Metadata 53
5.1 Introduction 53
5.2 What XML looks like 54
5.3 XML schemas 56
5.4 Namespaces 57
5.5 Creating and editing XML 60
5.6 Transforming XML 62
5.7 Why use XML? 65

6 METS: The Metadata Package 67
6.1 Introduction 67
6.2 Why use METS? 67
6.3 The METS architecture 69
6.4 Identifiers within METS 70
6.5 The METS root element 71
6.6 The METS Header 72
6.7 Descriptive Metadata Section 73
6.8 Administrative Metadata Section 74
6.9 The File Section 77
6.10 The Structural Map 82
6.11 Structural Links and Behavior Section 85
6.12 Creating and using METS in practice 87

7 Descriptive Metadata: Semantics 89
7.1 Introduction 89
7.2 Dublin Core 89

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 (Chapter 3) 173
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) 177
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