

A How-To-Do-It Manual®

BUILDING DIGITAL LIBRARIES

Terry Reese, Jr. and Kyle Banerjee

CONTENTS

List of Figures	ix
Preface	xiii
1. Planning a Digital Repository	1
What Is a Digital Repository?	1
The Decision to Build a Digital Repository	2
Advantages of a Digital Repository	6
Selling the Project	7
Understanding the Purpose of the Repository	8
Anticipating Use Patterns	10
Image and Text Processing	12
Archival Image Formats	15
Display Image Formats	15
Sample Archival Settings	16
Rights Management	17
Accommodating Changing Formats and Data Structures	19
Protecting Integrity of Resources	23
LOCKSS (Lots of Copies Keep Stuff Safe)	24
Disaster Planning and Security	26
Managing with Available Resources and Change ...	26
Further Information	29
Summary	30
References	31
2. Acquiring, Processing, Classifying, and Describing Digital Content	33
Planning Workflow	33
Developing the Collection	36
Acquiring Digital Content	39
Processing and Organizing Digital Content	43

Organizing Related Works and Subcollections	46
Rights Management	48
Batch Processes	49
Ergonomics	52
Summary	54
3. Choosing a Repository Architecture.	55
Questions to Ask before Choosing an Architecture	55
Required Features	56
Desirable Features	57
Frameworks for Digital Repositories	58
Platforms Optimized for Specific Purposes	61
Evaluating Repository Functionality	69
Resource Identification and Ingestion.	69
Automating Management and Organization	76
Indexing for Easy Retrieval	77
Other Storage Challenges.	80
Repository Administration	81
Summary	83
References	83
4. General Purpose Technologies Useful for Digital Repositories.	85
The Changing Face of Metadata	85
XML in Libraries	86
XHTML	88
XPath.	89
XForms	90
XSLT.	90
XLink	90
XQuery	91
XPointer	91
XML Schema	91
Why Use XML-based Metadata?	97
XML Is Human-Readable	97
XML Offers a Quicker Cataloging Strategy	102
XML Can Represent Multi-formatted and Embedded Documents.	103

XML Metadata Becomes “Smarter”	105
XML Is Not Just a Library Standard	105
Future of Software Development	106
Web Services and SOAP	108
Sharing Your Services	113
Summary	114
References	115
5. Metadata Formats	117
Metadata Primitives	117
MARC	119
MARC21XML	121
Dublin Core	123
History	123
Elements	125
Strengths	128
Challenges	129
MODS (Metadata Object Description Schema)	130
History	130
Strengths	132
Challenges	133
METS (Metadata Encoding and Transmission Standard)	134
History	134
METS at a Glance	135
Applications	136
Semantic Web	138
Application Profiles	142
Summary	144
References	145
6. Sharing Data: Metadata Harvesting and Distribution	147
The Evolving Roles of Libraries	147
Metadata Wants to Be Free	149
Sharing Metadata	151
XSLT (eXtensible Stylesheet Transformation)	151
Metadata Crosswalking	157
Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)	161

OAI-PMH Verbs	162
OAI-PMH Application	169
Facilitating Third-Party Indexing	169
Metadata Repurposing	170
OSU Electronic Theses Process	171
Microformats	173
COinS	175
UNAPI	177
Summary	178
References	179
7. Federated Searching of Repositories	181
What Is Federated Searching?	182
Federated Search and Digital Libraries.	184
Federated Search versus Traditional Search Engines	185
Current Research	186
Recommender/Collaborative Filtering	186
Deduplication of Results	186
Knowledge-Base Management.	187
Automatic Data Classification	187
Ranking Systems	187
Need for Speed	188
Searching Protocols	188
Z39.50	189
SRU/SRW (Search/Retrieval/URL and Search/Retrieval Web Service).	192
OpenSearch	197
Linking Protocols	201
OpenURL	202
DOI (Digital Object Identifiers).	204
Search Engine Support.	206
Service Registries	207
Service.	209
Agent.	210
Collection	211
Evaluating Needs.	215
Developmental Needs	215
User Needs	215

Summary	216
References	217
8. Access Management	219
Copyright Issues	220
Copyright as Organizational Policy	222
Can It Be Archived? Can It Be Distributed?	226
Long-Term Rights Management	227
Allowing/Restricting Access	228
CONTENTdm	229
DSpace	231
Control Mechanisms	232
LDAP	233
Shibboleth	234
OpenID	235
Athens	235
Monitoring Repository Use and Statistics	236
Intellectual Property	236
Service Usability	236
Statistical Analysis	238
Web Spiders/Harvesters	238
Item Prefetching	239
Summary	239
References	240
9. Planning for the Future	241
Providing Information That People Need	241
Libraries' New Roles	243
Learning from the Past	244
Adapting to Change	247
Consolidation and Specialization	249
Federated Collection Management	251
Federated Vocabularies	255
Summary	256
References	257
10. Conclusions	259
Index	267
About the Authors	277