Contents

List of Figures ix
List of Abbreviations xi
Preface xv
Acknowledgments xxi

Part I. Digital Curation: Scope and Incentives 1

Chapter 1. Introduction 3
Why There Is a Need for Digital Curation 3
What Digital Curation Is 5
Why We Should Be Interested in Digital Curation 8
Incentives for Digital Curation 11
Direct Benefits to Data Creators 12
"Public Good" Obligations 13
Compliance Reasons 13
Digital Curators 14
Summary: Main Characteristics of Digital Curation 15
References 16

Chapter 2. The Changing Landscape 19
Cyberscholarship: New Ways of Working 19
Cyberscholarship in Practice 21
E-science 21
Cyberscholarship's Requirements and Challenges 22
Content 22
Tools and Services 23
Expertise 23
Digital Curation: A New Profession, New Requirements 25
Educating and Training Digital Curators 28
Summary: Meeting the New Demands 29
References 29

Chapter 3. Conceptual Models 33
The DCC Curation Lifecycle Model 34
The Digital Curation Centre 35
Other Lifecycle Models 36

The OAIS Reference Model 38
OAIS Functions 38
Actors and Objects 39
OAIS Information Packages 40
OAIS and the DCC Curation Lifecycle Model 41
Summary: The Importance of Models 43
References 43

Chapter 4. Defining Data 45
Data as Digital Heritage 45
Born-Digital *and* Digitized Data 46
Data—And Much More 47
Metadata Is Data Too 49
Databases 50
Summary: New Kinds of Data 50
References 51

Part II. Key Requirements for Digital Curation 53

Chapter 5. Curation and Curators 55
Aims of Digital Curation 55
Scope of Digital Curation 56
Ensuring Longevity 57
Ensuring Integrity 57
Maintaining Accessibility 57
Roles of Digital Curators 58
Funding Bodies 59
Discipline Groups 59
Data Creators 60
Data Users and Reusers 60
Data Curators 60
Summary: Managing Curation 62
References 62

Chapter 6. Description and Representation Information 65
The Need for Description and Representation Information 66
Definitions 67
Standards for Description and Representation Information 68
Description Information 69
Preservation Metadata 70
Persistent Identifiers 72
Metadata Schemas and Standards 72
PREMIS 73
METS 74
MODS and MADS 75
Representation Information 76
OAIS and Representation Information 77
Sharing Representation Information 78
Policies for Description and Representation Information 79

Summary: Curation Needs Metadata 79
References 80

Chapter 7. Preservation Planning and Policy 83
Risk Management as the Context for Preservation Planning 83
Key Planning Steps 84
Planning for Sequential Actions 85
Planning for Full Lifecycle Actions 85
Policy for Curation 86
What Policies Address 86
Kinds of Policies Required 88
Costs of Curation 89
Summary: Planning for Active Management 91
References 91

Chapter 8. Sharing Knowledge and Collaborating 93
Keeping Up-to-Date 93
Starting Points 94
Online Tutorials 94
Project Websites 95
Blogs and E-mail Lists 95
Online Journals 96
Other Sources 96
Collaboration: Intrinsic to Digital Curation 96
Standards: Essential for Digital Curation 98
Tools and Toolkits 99
Summary: Collaboration Is the Key 100
References 101

Part III. The Digital Curation Lifecycle in Action 103

Chapter 9. Designing Data 105
Designing Curation-Ready Data 106
Importance of Standards 108
Designing Projects with Curation in Mind 109
Three Examples 110
Summary: Planning Data for Curation 112
References 112

Chapter 10. Creating Data 115
Policies for Creating and Receiving Data 116
Creating Data for Curation 117
Structuring Data for Use and Reuse 118
Open Formats and Open Source 119
Significant Properties and Authenticity 120
Documentation 122
Influencing Data Creators 123
Structuring Data for Management 124
Data Management 124
Data Quality 125

Structuring Data for Discoverability 126
Receiving Data for Curation 127
Summary: The Positive Effects of "Good" Data 128
References 128

Chapter 11. Deciding What Data to Keep 131
What Is Appraisal? 131
What Data Do We Want to Keep? 133
Drivers for Keeping Data 134
Why We Can't Keep Everything 136
How Long Do We Want to Keep Those Data? 137
Appraisal and Selection Policies 138
Who Decides? 139
Appraisal Tools 140
Two Examples 142
Selection and Archiving of Websites 142
Appraisal of Scientific Data Sets 143
Reappraisal 144
Disposal of Data 145
Transfer of Data 146
Destruction of Data 146
Summary: The Necessity for Appraisal and Selection 147
References 147

Chapter 12. Ingesting Data 151
OAIS and Ingest 152
Ingest Processes in More Detail 152
Submitting SIPs 153
Receiving SIPs 153
Generating AIPs 155
Ingest Tools 156
Policies for Ingest 158
Summary: Automation Is the Key 158
References 158

Chapter 13. Preserving Data 161
Digital Preservation Methods 162
Migration in Practice 165
Implementing Migration 170
Migration Changes Data 170
Automating Preservation Actions 171
Tools 171
Metadata Tools 173
Format Validation, Format Registry, and Obsolescence Notification Tools 173
Normalizing and Encapsulation Tools 174
Migration Tools 174
Emulation Tools 175
Web Archiving Tools 175
Other Curation Tools 175
Tool Development 176

Summary: Methods and Tools 178
References 178

Chapter 14. Storing Data **181**
Storage Requirements 182
Organizational Structure and Continuity 183
Technical Infrastructure and Practices 184
Best Practice in Data Storage 184
Ensuring Quality of Storage 186
OAIS Reference Model 186
Trusted Digital Repositories 187
Audit and Certification 188
Backing Up Data 188
Data Security 190
Physical Security 191
Network and File Security 191
Repository Software and Storage Solutions 192
Fedora, DSpace, EPrints, and Other Repository Software 192
Storage Solutions 194
New Models for Collaboration 195
Summary: Storing Data Securely 196
References 197

Chapter 15. Using and Reusing Data **199**
Access, Use, and Reuse 199
Sharing Data 200
Building Blocks for Sharing and Reusing Data 202
Standards 203
Standards for Repository Functionality 203
Standards for Interoperability 204
Structuring Data for Access 204
Citing Data 205
Legal Issues 206
Collaboration Processes 207
Annotation 207
Provenance 209
Access Controls and Authentication Procedures 210
Transform 211
Migration 211
Creating New Data 213
Conclusion: The Lifecycle Continues 213
References 214

Index 217
About the Author 225