Contents

Part I Methodology

1 Long-Term Preservation of Digital Documents 3
1.1 Blessing and Curse of Digital Media 3
1.2 Terms, Concepts, and Challenges 5
1.3 Preserving Character Streams 10
1.4 Technical Approaches to Long-Term Preservation 12
1.5 Legal and Social Concerns 19

2 OAIS and DSEP Organizational Models 21
2.1 The OAIS Reference Model 21
2.1.1 Background 22
2.1.2 The Information Model 22
2.1.3 Modeling Context and Processes 24
2.2 The DSEP Process Model 26

3 Migration 31
3.1 Migration: Definition and Goals 31
3.2 Migration in Long-Term Preservation 34
3.2.1 Target Data Formats 35
3.2.2 Digital Media Migration 37
3.2.3 Migration to Nondigital Media 39
3.2.4 Migration by Transformation 47
3.3 Archiving Processes of the Migration Approach 49
3.4 Strengths and Weaknesses of the Migration Approach 54

4 Emulation 57
4.1 Emulation: Notions and Goals 57
4.2 Emulation as a Means for Long-Term Preservation 64
4.2.1 What Exactly Should be Emulated? 64
4.2.2 Emulation Variants 66
4.2.3 Using Virtual Machines 69

4.3 Preservation Processes in Emulation Approaches 72
4.4 Chances and Risks of Emulation 77

5 Document Markup 79
5.1 Introductory Example 79
5.2 Markup Variants 80
5.2.1 Procedural, Structural, Semantic Markup 80
5.2.2 Embedded Markup Considered Harmful 84
5.2.3 Levels of Markup 86
5.3 Exploiting Markup for Long-Term Preservation 87
5.3.1 Requirements for Long-Term Preservation 87
5.3.2 Bibliographic Requirements 90
5.4 Persistence is a Virtue 91
5.4.1 Uniform Resource Identifier, -Name, -Locator 92
5.4.2 Referencing Documents 93
5.4.3 Handles and Digital Object Identifiers 96
5.4.4 Summary 97

6 Standard Markup Languages 99
6.1 Standards for Syntactic Document Markup 99
6.1.1 Tagged Image File Format (TIFF) 100
6.1.2 Portable Document Format (PDF) 101
6.1.3 HyperText Markup Language (HTML) 103
6.1.4 eXtensible Markup Language (XML) 104
6.2 Standards for Semantic Document Markup 114
6.2.1 Resource Description Framework (RDF) 114
6.2.2 Topic Maps 117
6.2.3 Ontologies: OWL 119
6.3 Vision: The Semantic Web 121

7 Discussion 123
7.1 Why Do We Need to Act NOW? 123
7.2 What Do We Know Already, What Remains to Be Done? 125
7.3 Facing Reality 128
7.4 A Combined Approach 130

Part II Recent Preservation Initiatives

8 Markup: Current Research and Development 135
8.1 The Dublin Core Metadata Initiative 135
8.2 The Metadata Encoding & Transmission Standard 142
8.3 The Victorian Electronic Records Strategy (VERS) ... 147

8.4 The Text Encoding Initiative (TEI) 154
8.5 The PANDORA Project 162

9 Migration: Current Research and Development 171
9.1 Migration in VERS-Compliant Recordkeeping Systems 171
9.2 Preserving the Whole 176
9.3 Risk Management of Digital Information 183
9.4 Database Migration 190
9.4.1 Introduction 190
9.4.2 Repository Architecture 192
9.4.3 Experiment 199
9.4.4 Lessons Learned 202
9.4.5 Summary 206

10 Emulation: Current Research and Development 207
10.1 An Emulation Experiment by Jeff Rothenberg 207
10.2 The Universal Virtual Computer (UVC) 213

11 Software Systems for Archiving 221
11.1 Introduction 221
11.1.1 Basic Conditions for a Product Assessment . . . 222
11.1.2 Criteria Catalog within the Decision-Making Process 224
11.2 Development of Criteria 225
11.2.1 Starting Points 225
11.2.2 Previous Analyses 226
11.2.3 Functional Criteria 228
11.2.4 Nonfunctional Criteria 230
11.3 Criteria Catalog 231
11.4 General Assessment of Current Archiving Systems . . . 231
11.5 Discussion 238
11.6 Product Examples 239
11.7 Rated Products 240

References 257

Index 263