

Summary Table of Contents

Preface	VII
For whom is this book intended? What is its topical scope? Summary of its organization. Suggestions how to read it.	
Part I: Why We Need Long-term Digital Preservation	1
1 State of the Art	7
Challenges created by technological obsolescence and media degradation. Preservation as a different topic than repository management. Preservation as specialized communication.	
2 Economic Trends and Social Issues	23
Social changes caused by and causing the information revolution. Cost of information management. Stresses in the information science and library professions. Interdisciplinary barriers.	
Part II: Information Object Structure	53
3 Introduction to Knowledge Theory	57
Starting points for talking about information and communication. Basic statements that are causing confusion and misunderstandings. Objective and subjective language that we use to talk about language, communication, information, and knowledge.	
4 Preservation Lessons from Scientific Philosophy	77
Distinguishing essential from accidental message content, knowledge from information, trusted from trustworthy, and the pattern of what is communicated from any communication artifact.	
5 Trust and Authenticity	93
How we use <i>authentic</i> to describe all kinds of objects. Definition to guide objective tests of object authenticity. Object transformations. Handling dynamic information.	
6 Describing Information Structure	109
Architecture for preservation-ready objects, including metadata structure and relationships to semantics. Names, references, and identifiers. Ternary relations for describing structure.	

Part III: Distributed Content Management	135
7 Digital Object Formats	139
Standards for character sets, file formats, and identifiers as starting points for preservation.	
8 Archiving Practices	163
Security technology. Record-keeping and repository best practices and certification.	
9 Everyday Digital Content Management	181
Storage software layering. Digital repository architecture. Types of archival collection.	
Part IV: Digital Object Architecture for the Long Term	205
10 Durable Bit-Strings and Catalogs	209
Media longevity. Not losing the last copy of any bit-string. Ingestion and catalog consistency.	
11 Durable Evidence	219
Cryptographic certification to provide evidence that outlasts the witnesses that provided it.	
12 Durable Representation	235
Encoding documents and programs for interpretation, display, and execution on computers whose architecture is not known when the information is fixed and archived.	
Part V: Peroration	251
13 Assessment and the Future	251
Summary of principles basic to preservation with TDO methodology. Next steps toward reduction to practice. Assessment of the TDO preservation method against independent criteria.	
14 Appendices	265
Glossary. URI syntax. Repository requirements analysis. Assessment of TDO methodology. UVC specification. SW wanted.	
Bibliography	303

Detailed Table of Contents

Preface	VII
Trustworthy Digital Objects	VIII
Structure of the Book	IX
How to Read This Book	XI
 Part I: Why We Need Long-term Digital Preservation	 1
 1 State of the Art	 7
1.1 What is Digital Information Preservation?	8
1.2 What Would a Preservation Solution Provide?	11
1.3 Why Do Digital Data Seem to Present Difficulties?	12
1.4 Characteristics of Preservation Solutions	14
1.5 Technical Objectives and Scope Limitations	19
1.6 Summary	21
 2 Economic Trends and Social Issues	 23
2.1 The Information Revolution	23
2.2 Economic and Technical Trends	25
2.2.1 Digital Storage Devices	27
2.2.2 Search Technology	29
2.3 Democratization of Information	30
2.4 Social Issues	31
2.5 Documents as Social Instruments	33
2.5.1 Ironical?	34
2.5.2 Future of the Research Libraries	37
2.5.3 Cultural Chasm around Information Science	39
2.5.4 Preservation Community and Technology Vendors	41
2.6 Why So Slow Toward Practical Preservation?	43
2.7 Selection Criteria: What is Worth Saving?	45
2.7.1 Cultural Works	46
2.7.2 Video History	47
2.7.3 Bureaucratic Records	48
2.7.4 Scientific Data	50
2.8 Summary	50

Part II: Information Object Structure	53
3 Introduction to Knowledge Theory	57
3.1 Conceptual Objects: Values and Patterns	58
3.2 Ostensive Definition and Names	60
3.3 Objective and Subjective: Not a Technological Issue	63
3.4 Facts and Values: How Can We Distinguish?	65
3.5 Representation Theory: Signs and Sentence Meanings	68
3.6 Documents and Libraries: Collections, Sets, and Classes	70
3.7 Syntax, Semantics, and Rules	72
3.8 Summary	74
4 Lessons from Scientific Philosophy	77
4.1 Intentional and Accidental Information	77
4.2 Distinctions Sought and Avoided	79
4.3 <i>Information and Knowledge: Tacit and Human Aspects</i>	82
4.4 Trusted and Trustworthy	85
4.5 Relationships and Ontologies	86
4.6 What Copyright Protection Teaches	88
4.7 Summary	90
5 Trust and Authenticity	93
5.1 What Can We Trust?	94
5.2 What Do We Mean by 'Authentic'?	95
5.3 Authenticity for Different Information Genres	98
5.3.1 Digital Objects	98
5.3.2 Transformed Digital Objects and Analog Signals	99
5.3.3 Material Artifacts	101
5.3.4 Natural Objects	102
5.3.5 Artistic Performances and Recipes	102
5.3.6 Literature and Literary Commentary	103
5.4 How Can We Preserve Dynamic Resources?	103
5.5 Summary	105
6 Describing Information Structure	109
6.1 Testable Archived Information	110
6.2 Syntax Specification with Formal Languages	111
6.2.1 String Syntax Definition with Regular Expressions	111
6.2.2 BNF for Program and File Format Specification	112
6.2.3 ASN.1 Standards Definition Language	113
6.2.4 Schema Definitions for XML	114
6.3 Monographs and Collections	115

6.4	Digital Object Schema	117
6.4.1	Relationships and Relations	118
6.4.2	Names and Identifiers, References, Pointers, and Links	120
6.4.3	Representing Value Sets	122
6.4.4	XML “Glue”	123
6.5	From Ontology to Architecture and Design	124
6.5.1	From the OAIS Reference Model to Architecture	125
6.5.2	Languages for Describing Structure	127
6.5.3	Semantic Interoperability	128
6.6	Metadata	129
6.6.1	Metadata Standards and Registries	130
6.6.2	Dublin Core Metadata	131
6.6.3	Metadata for Scholarly Works (METS)	132
6.6.4	Archiving and Preservation Metadata	133
6.7	Summary	133
Part III: Distributed Content Management		135
7	Digital Object Formats	139
7.1	Character Sets and Fonts	139
7.1.1	Extended ASCII	140
7.1.2	Unicode/UCS and UTF-8	140
7.2	File Formats	142
7.2.1	File Format Identification, Validation, and Registries	143
7.2.2	Text and Office Documents	145
7.2.3	Still Pictures: Images and Vector Graphics	146
7.2.4	Audio-Visual Recordings	147
7.2.5	Relational Databases	150
7.2.6	Describing Computer Programs	151
7.2.7	Multimedia Objects	151
7.3	Perpetually Unique Resource Identifiers	152
7.3.1	Equality of Digital Documents	153
7.3.2	Requirements for UUIDs	154
7.3.3	Identifier Syntax and Resolution	156
7.3.4	A Digital Resource Identifier	159
7.3.5	The “Info” URI	160
7.4	Summary	160

8 Archiving Practices	163
8.1 Security	163
8.1.1 PKCS Specification	164
8.1.2 Audit Trail, Business Controls, and Evidence	165
8.1.3 Authentication with Cryptographic Certificates	165
8.1.4 Trust Structures and Key Management	169
8.1.5 Time Stamp Evidence	171
8.1.6 Access Control and Digital Rights Management	172
8.2 Recordkeeping Standards	173
8.3 Archival Best Practices	175
8.4 Repository Audit and Certification	176
8.5 Summary	178
9 Everyday Digital Content Management	181
9.1 Software Layering	183
9.2 A Model of Storage Stack Development	185
9.3 Repository Architecture	186
9.3.1 Lowest Levels of the Storage Stack	187
9.3.2 Repository Catalog	189
9.3.3 A Document Storage Subsystem	191
9.3.4 Archival Storage Layer	194
9.3.5 Institutional Repository Services	195
9.4 Archival Collection Types	196
9.4.1 Collections of Academic and Cultural Works	196
9.4.2 Bureaucratic File Cabinets	197
9.4.3 Audio/Video Archives	199
9.4.4 Web Page Collections	201
9.4.5 Personal Repositories	202
9.5 Summary	202
Part IV: Digital Object Architecture for the Long Term	205
10 Durable Bit-Strings and Catalogs	209
10.1 Media Longevity	210
10.1.1 Magnetic Disks	211
10.1.2 Magnetic Tapes	211
10.1.3 Optical Media	212
10.2 Replication to Protect Bit-Strings	213
10.3 Repository Catalog ↔ Collection Consistency	214
10.4 Collection Ingestion and Sharing	215
10.5 Summary	217

11 Durable Evidence	219
11.1 Structure of Each Trustworthy Digital Object	220
11.1.1 Record Versions: a Trust Model for Consumers	222
11.1.2 Protection Block Content and Structure	222
11.1.3 Document Packaging and Version Management	224
11.2 Infrastructure for Trustworthy Digital Objects	227
11.2.1 Certification by a Trustworthy Institution (TI)	228
11.2.2 Consumers' Tests of Authenticity and Provenance	230
11.3 Other Ways to Make Documents Trustworthy	232
11.4 Summary	233
12 Durable Representation	235
12.1 Representation Alternatives	236
12.1.1 How Can We Keep Content Blobs Intelligible?	236
12.1.2 Alternatives to Durable Encoding	237
12.1.3 Encoding Based on Open Standards	238
12.1.4 How Durable Encoding is Different	241
12.2 Design of a Durable Encoding Environment	242
12.2.1 Preserving Complex Data Blobs as Payload Elements	243
12.2.2 Preserving Programs as Payload Elements	245
12.2.3 Universal Virtual Computer and Its Use	245
12.2.4 Pilot UVC Implementation and Testing	247
12.3 Summary	248
Part V: Peroration	251
13 Assessment and the Future	251
13.1 Preservation Based on Trustworthy Digital Objects	252
13.1.1 TDO Design Summary	252
13.1.2 Properties of TDO Collections	253
13.1.3 Explaining Digital Preservation	254
13.1.4 A Pilot Installation and Next Steps	255
13.2 Open Challenges of Metadata Creation	256
13.3 Applied Knowledge Theory	259
13.4 Assessment of the TDO Methodology	261
13.5 Summary and Conclusion	263

Appendices	265
A: Acronyms and Glossary	265
B: Uniform Resource Identifier Syntax	280
C: Repository Requirements	282
D: Assessment with Independent Criteria	284
E: Universal Virtual Computer Specification	289
E.1 Memory Model	289
E.2 Machine Status Registers	290
E.3 Machine Instruction Codes	291
E.4 Organization of an Archived Module	296
E.5 Application Example	297
F: Software Modules Wanted	300
Bibliography	303

Figures

Fig. 1: OAIS high-level functional structure	16
Fig. 2: Information interchange, repositories, and human challenges.	17
Fig. 3: How much PC storage will \$100 buy?	27
Fig. 4: Schema for information object classes and relationship classes	59
Fig. 5: Conveying meaning is difficult even without mediating machinery	64
Fig. 6: A meaning of the word 'meaning'	69
Fig. 7: Semantics or 'meaning' of programs	73
Fig. 8: Depictions of an English cathedrals tour	78
Fig. 9: Relationships of meanings;	79
Fig. 10: Bit-strings, data, information, and knowledge	84
Fig. 11: Information delivery suggesting transformations that might occur	99
Fig. 12: A digital object (DO) model.	116
Fig. 13: Schema for documents and for collections	118
Fig. 14: A value set, as might occur in Fig. 12 metadata	122
Fig. 15: OAIS digital object model	124
Fig. 16: OAIS ingest process	126
Fig. 17: Kitchen process in a residence	126
Fig. 18: Network of autonomous services and clients	135
Fig. 19: Objects contained in an AAF file	148
Fig. 20: Identifier resolution, suggesting a recursive step	159
Fig. 21: MAC creation and use	167

Fig. 22: Cryptographic signature blocks	168
Fig. 23: Trust authentication networks:	169
Fig. 24: Software layering for “industrial strength” content management	182
Fig. 25: Typical administrative structure for a server layer	184
Fig. 26: Repository architecture suggesting human roles	186
Fig. 27: Storage area network (SAN) configuration	188
Fig. 28: Replacing JSR 170 compliant repositories	193
Fig. 29: Preservation of electronic records context	195
Fig. 30: Workflow for cultural documents	197
Fig. 31: Workflow for bureaucratic documents	198
Fig. 32: MAC-sealed TDO constructed from a digital object collection	220
Fig. 33: Contents of a protection block (PB)	223
Fig. 34: Nesting TDO predecessors	225
Fig. 35: Audit trail element—a kind of digital documentary evidence	226
Fig. 36: Japanese censor seals: ancient practice to mimic in digital form	229
Fig. 37: A certificate forest	230
Fig. 38: Durable encoding for complex data	244
Fig. 39: Durable encoding for preserving a program	245
Fig. 40: Universal Virtual Computer architecture	289
Fig. 41: Exemplary register contents in UVC instructions	291
Fig. 42: UVC bit order semantics	292
Fig. 43: Valid UVC communication patterns	296

Tables

Table 1: Why should citizens pay attention?	3
Table 2: Generic threats to preserved information	10
Table 3: Information transformation steps in communication	18
Table 4: Metadata for a format conversion event	97
Table 5: Dublin Core metadata elements	132
Table 6: Closely related semantic concepts	134
Table 7: Samples illustrating Unicode, UTF-8, and glyphs	142
Table 8: Sample AES metadata	144
Table 9: Reference String Examples	155
Table 10: Different kinds of archival collection	204
Table 11: NAA content blob representations	240
Table 12: TDO conformance to InterPARES authenticity criteria	284
Table 13: Comments on a European technical research agenda	286