

Preserving Digital Information

Henry M. Gladney

Preserving Digital Information

With 43 Figures and 13 Tables

 Springer

Author

Henry M. Gladney
HMG Consulting
Saratoga, CA 95070, USA
hgladney@gmail.com
<http://home.pacbell.net/hgladney/>

Library of Congress Control Number: 2006940359

ACM Computing Classification (1998): H.3, K.4, K.6

ISBN 978-3-540-37886-0 Springer Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable for prosecution under the German Copyright Law.

Springer is a part of Springer Science+Business Media

springer.com

© Springer-Verlag Berlin Heidelberg 2007

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Typeset by the author

Production: LE-TeX Jelonek, Schmidt & Vöckler GbR, Leipzig

Cover design: KünkelLopka Werbeagentur, Heidelberg

Printed on acid-free paper 45/3100/YL - 5 4 3 2 1 0

Dedicated to
My parents, who sacrificed much to obtain
the best possible education for their children,
And schools whose generosity admitted me to an education
I could not otherwise have afforded:
Upper Canada College, Toronto,
Trinity College, University of Toronto,
and
Princeton University.

Preface

What might this book offer to people already confronted with information overload? It is intended to help citizens who want to understand the issues around digital preservation without reading the technical literature. It is also intended to help scholars who want depth quickly find authoritative sources. It is for

- authors, artists, and university faculty who want their digitally represented works to be durable and to choose information service providers that are committed and competent to ensure preservation of those works;
- attorneys, medical professionals, government officials, and businessmen who depend on the long-term reliability of business records that are today mostly in digital form;
- entertainment industry managers, because their largest enterprise assets are old performance recordings;
- archivists, research librarians, and museum curators who need to understand digital technology sufficiently to manage their institutions, especially those curators that are focusing on digital archiving;
- citizens who want to understand the information revolution and the attendant risks to information that might affect their lives; and
- software engineers and computer scientists who support the people just mentioned.

Ideally, a book about a practical topic would present prescriptions for immediately achieving what its readers want—in this case a durable existence for monographs, articles, performance recordings, scientific data, business and government records, and personal data that we all depend on. Doing so is, however, not possible today because software and infrastructure for reliably preserving large numbers of digital records have not yet been built and deployed, even though we know what software would work and what services repository institutions need to provide.

The software needed includes tools for packaging works for long-term storage and for extracting information package contents conveniently for their eventual consumers. Many useful components exist, and some are in use. Others are not yet represented by specifications that must precede peer criticism, selection, and refinement within communities that have specialized applications. Some of the agreements needed will ultimately be expressed as information interchange standards. The products of such work could be deployed in five to ten years.

The infrastructure needed includes institutional repositories (digital archives) that share methods and digital content and whose characteristics

are relatively well understood. Since large projects to create the required network and storage infrastructure exist in several countries (Australia, Germany, The Netherlands, the U.K., and the U.S.), the current book positions preservation within this infrastructure without describing the infrastructure in detail. It focuses on principles for reliable digital preservation and on what these principles teach about design for representing every kind of intellectual work.

Substantial deployment will not occur until interested communities achieve consensus on which proposed components to choose so that their clients, the producers and consumers of information resources, can share their works safely and efficiently. We intend this book to help the necessary discussions.

Trustworthy Digital Objects

The *Open Archival Information Systems (OAIS) Reference Model* and related expositions address the question, “What architecture should we use for a digital repository?” This is sometimes construed as all aspects of providing digital library or archive services—everything that might be pertinent to creating and managing a digital repository within an institution such as a university, a government department, or a commercial enterprise.

To address the *OAIS* question and the responsibilities of repository institution managers, doing so in the compass of a single monograph, seems to me a too-difficult task, partly because accepted best practices have not yet emerged from increasing research activities. In contrast, digital preservation is a tractable topic for a monograph. Among the threats to archival collections are the deleterious effects of technology obsolescence and of fading human recollection. In contrast to the *OAIS* question, this book addresses a different question, “What characteristics will make saved digital objects useful into the indefinite future?”

The book’s technical focus is on the structure of what it calls a Trustworthy Digital Object (TDO), which is a design for what the *OAIS* international standard calls an Archival Information Package (AIP). It further recommends TDO architecture as the packaging design for information units that are shared, not only between repository institutions, but also between repositories and their clients—information producers and information consumers.

In contrast, most research articles addressing digital preservation focus on the structure and operations of archival repository institutions and research libraries—what they sometimes call Trusted Digital Repositories. A critic has called this distinction still controversial in the sense that TDO methodology is not widely accepted as the path that must be taken. In fact,

TDO architecture seems to have been mostly ignored since 2001, when current and former IBM Research staff members started to publish its description.

Most of today's digital preservation research literature focuses on a small segment of what is created in digital form—the kinds of information that research libraries collect. It pays little attention to the written output of governments and of private sector enterprises. It hardly mentions the myriad documents important to the welfare and happiness of individual citizens—our health records, our property records, our photographs and letters, and so on. Some of these are tempting targets for fraud and other misfeasance. In contrast, deliberate falsification does not seem to be a prominent problem for documents of primarily cultural interest. Protecting against its effects has therefore received little attention in the cultural heritage digital preservation literature.

The book therefore explains what I believe to be the shortfalls of preservation methodology centered on repository institution practices, and justifies my opinion that TDO methodology is sound. Its critique of the trusted digital repositories approach is vigorous. I invite similarly vigorous public or private criticism of TDO methodology and, more generally, of any opinion the book expresses.

Structure of the Book

The reader who absorbs this book will understand that preservation of digital information is neither conceptually difficult nor mysterious. However, as with any engineering discipline, “the devil is in the details.” This motivates a typical engineering approach—breaking a problem into separate, tractable components.

Software engineers will recognize details from their own projects and readily understand both the broad structure and also the choice of principles emphasized. Readers new to digital preservation or to software engineering might find it difficult to see the main threads within the welter of details. Hopefully these readers will be helped by the Summary Table of Contents that can remind them of the book's flow in a single glance, the introduction that precedes each group of chapters, and also the summary that ends each chapter by repeating its most important points.

The book is laid out in five sections and a collection of back matter that provides detail that would have been distracting in the main text. The order of the sections and chapters is not especially significant.

The book proceeds from broad social circumstances to methodological details for preserving any kind of digital object whatsoever. It describes architectural abstractions without refinements that many people would

demand before they called them designs. This choice is intended to emphasize what might be obscured by the detail required in design for implementations. Before it begins with technical aspects, it summarizes the soundest available basis for discussing what knowledge we can communicate and what information we can preserve.

Throughout, the book emphasizes ideas and information that typical human users of information systems—authors, library managers, and eventual readers of preserved works—are likely to want. Its first section, **Why We Need Long-term Digital Preservation** describes the challenge, distinguishing our narrow interpretation of digital preservation from digital repository design and archival institution management.

Preservation can be designed to require no more than small additions to digital repository technology and other information-sharing infrastructure. The latter topics must respond to subtle variations in what different people will need and want and to subjective aspects of knowledge and communication. In contrast to the complexity and subjectivity of human thinking, the measures needed to mitigate the effects of technology obsolescence can be objectively specified once and for all.

Chapter 2 sketches social and computing marketplace trends driving the information access available to every citizen of the industrial nations—access that is transforming their lives. These transformations are making it a struggle for some librarians and archivists to play an essential role in the information revolution. Their scholarly articles suggest difficulties with digital preservation partly due to inattention to intellectual foundations—the theory of knowledge and of its communication.

The second section, **Information Object Structure**, reminds readers of the required intellectual foundation by sketching scientific philosophy, relating each idea to some aspect of communicating. It resolves prominent difficulties with notions of *trust*, *evidence*, *the original*, and *authenticity*. It emphasizes the distinction between *objective facts* and *subjective opinions*, which is not as evident in information practice as would be ideal. The section core is a communication model and an information representation model. These lead to our recommending structuring schemes for documents and collections.

The third section, **Distributed Content Management**, sketches electronic data processing standards that are essential starting points. It continues by discussing repository infrastructure aspects that comprise context for preservation software. Since most of this material is well known and well handled in previous works, Chapters 7 through 9 are limited to sketching the aspects essential for preservation and to providing citations intended to help readers who want more detail.

The fourth section, **Digital Object Architecture for the Long Term**, suggests how to solve those preservation challenges that technology can address. Chapters 10 through 12 present the TDO approach in a form permitting objective and specific critiques. It depends on well known elements of scientific and engineering methodologies: (1) careful attention to the interplay between the objective aspects (here, tools that might be employed) and what is necessarily subjective (human judgments, opinions, and intentions that cannot flourish in circumstances controlled too tightly); (2) focus on the wants and actions of individual people that balances and illustrates abstractions such as authenticity, integrity, and quality; (3) identification of possible failures and risk reduction; and (4) divide and conquer project management with modest pieces that build on other people's contributions and that facilitate and encourage their future contributions to address weaknesses and provide missing elements.

Specifically, Chapter 10 teaches replication to protect against losing the last copy of any bit-string. Chapter 11 describes signing and sealing to provide durable evidence about the provenance and content of any digital object, and of its links to other information. Chapter 12 shows how to encode bit-strings to be interpretable within any future computing system, even though we cannot today know such systems' architectures.

In the **Peroration**, Chapter 13 suggests open questions and work yet to be done. The questions include, "Is every detail of what we call TDO methodology correct and optimal? Are there missing pieces? What would be the architecture and design of satisfactory implementations? How can we make these convenient for users with little technical experience?" Such questions lead to suggestions for projects to create lightly coupled software modules.

How to Read This Book

Precise communication is unusually important for this book's topic. Accordingly, its diction is particularly cautious. Nevertheless, definitions are not given in the text except for unusually sensitive cases. The careful reader is referred to the Glossary.

How an individual word or phrase is used differs from community to community. For key words, we signal what we intend. A word in italics, such as *model*, has a relatively precise, technical meaning that is so important that this word has a Glossary entry. A word or phrase between double quotes, such as "model," is used to quote a source. Single quotes enclosing a word indicate that the word itself is being discussed as an object—as a symbol for something other than itself.

Some readers will find their objectives best met by reading this book out of order. This should not be surprising in a topic as complex and subtle as human communication. First time readers are encouraged to ignore the references, especially those to other sections of the book.

Some readers might be impatient with philosophical discussions that seem to them to expound little more than common sense. Such readers might proceed directly from the introductory chapters to **Digital Object Architecture for the Long Term**, consulting the **Information Object Structure** chapters only if they start to wonder how to improve upon what the fourth section proposes, or whether the whole work is soundly based.

Some readers will prefer to understand where we are leading them before they join us on the journey. We suggest that such readers might prefer to start with Chapter 13, which is devoted to an assessment of the merits of the TDO digital preservation approach.

Some readers will want more detail, others less. For those who want an introduction to preservation issues and to technology that can help address its challenges, we recommend generally ignoring the footnotes and the citations. For readers who want technical detail, possibly because they are skeptical about what the main text propounds, the footnote citations attempt to identify the most authoritative works. These citations are selections from about three times as many books and articles considered. By consulting these and the literature that they in turn cite, the reader can quickly learn what other people believe about digital preservation.

Some readers will want to decide quickly whether or not to inspect a cited work. The footnotes and an accompanying Web page are designed to help them. The objective is that a reader will be able to decide from each footnote alone whether or not to look at the cited work, i.e., decide without looking at any other page. Web page citations include the Web address, and are not repeated in the formal Bibliography at the end of the book. Instead they will be provided as actionable links in a supporting Web page.¹ Footnote citations of hard copy works are abbreviations of formal citations included in the Bibliography; they begin with the last name of the author and the publication date to make finding their Bibliography entries easy. Every footnote citation includes enough of the work's title for the reader to decide how interested he is in this source.

A few works are cited so often that it has been convenient to indicate them by abbreviations.² A few phrases are used so often that it is conven-

¹ This Web page is available at <http://home.pacbell.net/hgladney/pdilinks.htm>. As a fixed Web address is likely to be ephemeral, we suggest that readers locate a copy by a Web search for "Preserving Digital Information citations" or for the ISBN "3-540-37886-3" or "3540378863".

² See Appendix A for the abbreviations LSW, NDIIPP, OAIS, PDITF, PI, PK, RLG, TLP, and W2005.

ient to represent them with acronyms. All these are tabulated in the Glossary.

The professional literature cited extends to autumn 2006. The reader will understand that work that appeared shortly before this cut-off could not be considered as thoughtfully as earlier work. Recent articles selected for citation are suggested for their insights beyond what the book includes.

When this book's manuscript was nearing completion, there appeared the final report and recommendations of the Warwick Workshop, *Digital Curation and Preservation: Defining the research agenda for the next decade*.³ European experts across the full spectrum of the digital life cycle mapped the current state of play and future agenda. They reconsidered recommendations of a 1999 Warwick workshop and reviewed the progress made in implementing them. Their report concisely reflects the insights of many earlier discussions, making it a yardstick with which any reader can judge *Preserving Digital Information*. Appendix D uses its table of technical preservation components to assess TDO methodology.

Acknowledgements

I am grateful to John Bennett, Tom Gladney, Richard Hess, Peter Lucas, Raymond Lorie, John Sowa, and John Swinden for five years of conversation about topics represented in this book, including many suggestions for amendment of its draft versions. Their contributions and authoritative views are often acknowledged by the use of "we" in its text. I am particularly indebted to John Bennett for his patient inspection of several manuscript versions and his suggestions about how to communicate.

³

Warwick workshop 2005, http://www.dcc.ac.uk/training/warwick_2005/Warwick_Workshop_report.pdf.

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 Ironic?	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</i> : Tacit and Human Aspects	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