Preface

There has been a growing recognition of the need to address the fragility of the digital information that is deluging all aspects of our lives, whether in business, scientific, administrative, imaginative or cultural activities.

Society's growing dependence on the digital for its smooth operation as it becomes an information society provides the real urgency for addressing this issue. This case has been made very well in the large number of books and articles already published on the topic of digital preservation and therefore this case will not be expanded upon in this book.

Since there are many books about digital preservation why is there a need for yet one more?

At the time of writing the books and articles on digital preservation, for the most part, focus on consideration of documents, images and web pages; things which are normally just displayed by software for a human to view or listen to (or perhaps smell, taste or touch). We will refer to these as things which are *rendered*.

Yet there are clearly many more types of digital objects on which our lives depend and which may need to be preserved, such as databases, scientific data and software itself. These are things which are not simply rendered – they are processed and used in many different ways.

It should become clear to the reader that the tools and techniques used for preserving rendered objects are inadequate for all these other types of digital objects and we need to set our sights higher and wider. This book provides the concepts, techniques and tools which are needed.

Of course it is easy to make claims about digital preservation techniques – and there are many such claims! Therefore it is important that evidence is provided to support any such claims, which we do for our claims by using accelerated lifetime scenarios about the important changes which will challenge us. We use as examples a variety of digital objects from many sources and show tools and techniques by which they may be preserved.

1 Who Should Read This Book and Why?

This book is aimed at those who have problems in preserving digitally encoded information that they need to solve, especially where it goes beyond simply preserving rendered objects. The PARSE.Insight survey [1] suggests that while all researchers have documents and images, about half have non-rendered digital holdings such as raw data, scientific/statistical data, databases and software, therefore this book should be of wide interest.

It should also be essential reading for those who wish to audit their own archives, perhaps in advance of an independent audit, about how well they are doing in the preservation of the digitally encoded information which has been entrusted to them.

Researchers in digital preservation theory and developers of tools and techniques should also find valuable information here. Developers in the area of e-Science (also known as Cyberinfrastructure) may also gain a number of useful insights.

Some of the material in this book may be found to be too technical by some readers. For those readers we suggest that they skim over such material in order to at least be aware of the issues. This will allow them to advise more technical implementers who will certainly need such details.

To further help readers, the book is supported by other resources, including many hours of videos and presentations from the CASPAR project [2], which provides ❍ an elevator pitch for digital preservation, ❍ examples of digital preservation from several repositories, ❍ detailed lectures by the contributors to this book on many of the issues described here and ❍ lectures about, and video captures of, many of the software components. The open source software and further documentation is also available.

2 Structure of This Book

Part I of the book provides the concepts and theoretical basis that are needed, introducing, as examples along the way, digital objects from many sources. Since much of this book is based on the work of the CASPAR project, the examples will be derived from many disciplines including science, cultural heritage and contemporary performing arts.

The approach we take throughout is one of asking the questions which we believe a reasonably intelligent person may ask, and then providing answers to them. Sometimes, when there are some subtle but important points, we guide the reader towards the appropriate questions. As noted above, this will lead us into a number of technical issues which will not be to the taste of all readers but all topics are necessary for at least some readers.

Part II of the book shows practical examples of preserving a variety of specific objects and gives details of a range of tools and techniques. One obvious question, which an intelligent (but sceptical) reader may ask is “these tools and techniques may do something but why should I believe that they help to preserve things?”

After all, the only real way would be to live a long time and check the supposedly preserved objects in the future. However that is not very practical, and perhaps more importantly it does not help one to decide now whether to follow the ways proposed in this book. Choosing the wrong way could have a disastrous effect on what one intends to leave for future generations!

We provide what we believe is strong evidence that what is proposed **does** actually work for a wide variety of digital objects from many disciplines, through a number of accelerated lifetime scenarios, validated by members of the appropriate communities.

Part III provides answers to the questions about how to ensure that resources devoted to preserve digital objects are not wasted, showing a number of ways in which effort can be shared. In addition this part provides guidance on how to evaluate whether a particular repository (perhaps your own) is doing a good job, and where it might be improved. This part also describes the thinking behind the work carried out to produce the ISO standards on which the international audit and certification process can be based.

Throughout the book we indicate points where experience shows there is a danger of misunderstanding by the symbol

3 Preservation and Curation

This book is about digital preservation but there is another term which is being used, namely digital curation. The UK Digital Curation Centre [3] used to define this in the following way: "*Digital curation is maintaining and adding value to a trusted body of digital information for current and future use; specifically, we mean the active management and appraisal of data over the life-cycle of scholarly and scientific materials"*. This definition has been changed more recently to *"Digital curation involves maintaining, preserving and adding value to digital research data throughout its lifecycle"*. Sometimes the phrase "digital curation and preservation" is also used.

We prefer the term preservation in this book since we do not wish to restrict our consideration to "scholarly and scientific materials" nor "research data", because we wish to ensure we can apply our techniques to all kinds of digital objects including, for example, commercial and legal material. Nor do we wish to restrict ourselves to only a "trusted body of digital information" – since one might wish to preserve falsified data for example as evidence for legal proceedings. Moreover as we will see, our definition of preservation requires that if we are to preserve digitally encoded information we must ensure it remains understandable and usable. In other words preservation is the *sine qua non* of curation. For example it is possible to manage

and publish digitally encoded information without regard to future use; on the other hand if one wishes to ensure future as well as current use, one must understand the requirements for preservation.

4 OAIS Definitions

OAIS [4] plays a central role in this book. Many definitions, and some descriptive text, are taken from the updated OAIS; these are shown as ***bold italics***.

5 Acknowledgements

This book would not have been written without the work carried out by the many members of the CASPAR [2], DCC [3] and PARSE.Insight [1] projects, as well as the members of CCSDS [5] and others who have worked on developing OAIS [3] and the standards for certification of digital repositories [6], all of whom must be thanked for their efforts.

A fuller list of contributors may be found in "Contributors" at the end of the book.

Finally the editor and main author of this book would like to thank his family, in particular his wife Krystina and daughter Zoe, for their support and help in preparing this book for publication.

http://www.springer.com/978-3-642-16808-6

Advanced Digital Preservation
Giaretta, D.
2011, XXII, 510 p. 298 illus., 1 in color., Hardcover
ISBN: 978-3-642-16808-6