1. INTRODUCTION

1.1 Background

There is a rapidly increasing volume of information which exists in digital form. Whether created as a result of digitising non-digital collections, created as a digital publication, or created as part of the day-to-day business of an organisation, more and more information is being created digitally and the pace at which it is being created is accelerating. In the UK, initiatives such as the People's Network and the National Grid for Learning have emerged with a consequent need for quality digital materials to support them. Numerous other digitisation initiatives either planned or underway are constantly being announced within the commercial, higher education, and public sectors. In addition, a Government white paper announced the intention to have all newly created public records electronically stored and retrieved by 2004 (**PRO 1998**).

This activity is occurring in an environment in which there is a growing awareness of the significant challenges associated with ensuring continued access to these materials, even in the short term. In the UK, a series of research reports were commissioned by JISC (Joint Information Systems Committee) and the NPO (National Preservation Office) which served to highlight various aspects of digital preservation (**see Reference 2**).The reports provided a broad overview of key issues and two in particular (**Beagrie and Greenstein 1998** and **Hendley 1998**) recommended that further research be undertaken to explore the issues they raised in more detail.

The combination of these two factors is both challenging and troublesome. On the one hand, there are considerable opportunities offered by digital technology to provide rapid and efficient access to information. On the other hand, there is a very real threat that the digital materials will be created in such a way that not even their short-term viability can be assured, much less the prospect that future generations will also have access to them. The need to create and have widespread access to digital materials has raced ahead of the level of general awareness and understanding of what it takes to manage them effectively.

The need for guidance

Both the JISC/NPO studies and the second workshop in digital preservation organised by the JISC and the British Library in 1999 at Warwick identified the need to improve guidance on digital preservation. At around the same time, a survey commissioned by RLG (Research Libraries Group) investigated the needs of member institutions(Hedstrom and Montgomery 1998). A clear picture emerged from both sets of activities of a complex and rapidly changing environment in which those creating and/or acquiring digital resources would require guidance on how to manage those resources most effectively.

All organisations in all sectors have been or will soon be creating digital materials.They may be created as part of their organisational records, they may be created by digitising non-digital collections in order to enhance access to them, or they may be created digitally (**born digital**). However, they come into being, they will need to be managed as early as possible in their life-

18

cycle, preferably at the design stage, but if not as soon as practicable thereafter, if they are to remain accessible as long as they are required. Practical experience and expertise in this area is still scarce so there is a clear need for guidance to ensure that the significant opportunities are not overwhelmed by the equally significant threats.

Given this conjunction of factors, it seemed timely to embark on a handbook which aimed at both identifying good practice in creating and managing and preserving digital materials and also providing a range of practical tools to assist in that process. This handbook is being produced at a time when an important body of experience is emerging from recent research projects into digital preservation and from established data archives in the sciences and social sciences. Although many challenges remain, it is now possible to point to many examples of good practice and to suggest ways in which institutions can begin to address digital preservation. By providing a strategic overview of the key issues, discussion and guidance on strategies and activities, and pointers to key projects and reports, the handbook aims to provide guidance for institutions and individuals and a range of tools to help them identify and take appropriate actions.

Development of the handbook

In 1999 the AHDS (Arts and Humanities Data Service) submitted a proposal to the Preservation of and Access to the Recorded Heritage Research Programme.The proposal aimed to build on work which has already taken place in identifying the broad issues and challenges associated with digital preservation, and to provide more detailed guidance to all those creating and/or acquiring digital materials.The AHDS has considerable experience in collecting and managing digital materials and has been active in providing guidance in creating digital materials for the arts and humanities. Many of the challenges associated with ensuring continued access to digital materials are identical regardless of how or where they are created, so it made sense to build on this practical experience and to aim at a wider audience.The project was awarded funding of £33,561 from Resource: The Council for Museums,Archives and Libraries, with contributing in kind funding from AHDS and JISC, institutions represented by the Advisory Group, and participating case studies. The work was undertaken between July 1999 and September 2000 by Neil Beagrie and Maggie Jones on behalf of the AHDS.

An Advisory Group consisting of experts in the field of digital preservation was formed, all of whom had first hand knowledge of the range of complex issues involved. An early decision was that a handbook would be the most appropriate mechanism to provide the range of advice and guidance required for such a diverse audience. Research to compile the handbook combined traditional desktop research, utilising the world-wide-web as a source of freely available current information, as well as subscription-based print and electronic journals, supplemented with case studies and specialist interviews. Three very different case studies were selected to help develop the practical nature of the handbook and to ensure that it addressed key issues currently being faced by organisations. Through structured interviews with selected specialists, workshops and conference presentations, and the case studies, it was possible to gauge the overall level of awareness and understanding of digital preservation and to transfer that knowledge to the development of the handbook.

19

A consultation period for peer review and assessment was provided between 8 August and 4 September 2000. Comments were also accepted up until the end of the project to allow those wishing to comment to do so.

In general, the research for the handbook showed that the level of awareness of, and interest in, digital preservation is gradually increasing but is not keeping pace with the level of digital resource creation. In particular, institutions that have not played a role in preserving traditional collections do not have a strong sense of playing a role in preserving digital materials. Individual researchers were keen to "do the right thing" but frequently lacked the clear guidance and institutional backing to enable them to feel confident of what they should be doing.The difficulties of allocating responsibilities for preservation and maintenance in an environment in which digital resource creation is frequently a by-product of collaborative projects, which may well be funded by yet another external agency, was also mentioned. Overall, it appears that there is still a need to raise the level of awareness of digital preservation, particularly among funding agencies and senior administrators with responsibility for the strategic direction of an institution. This needs to be combined with more detailed guidance and training at the operational level. Moreover, the guidance needs to be able to accommodate people with varying levels of awareness and understanding of digital preservation, in a wide range of institutional settings, all of whose staff have significant constraints on their time.

Audience and purpose

Digital preservation has many parallels with traditional preservation in matters of broad principle but differs markedly at the operational level and never more so than in the wide range of decision makers who play a crucial role at various stages in the lifecycle of a digital resource. Consequently, this handbook is aiming at a very broad audience. In the first instance it is intended to provide guidance to institutions at national, regional and local levels who are involved in or contemplating creation and/or acquisition of digital materials. Within those institutions, the handbook is aiming at both administrators and practitioners and is accordingly structured to include a mix of high level strategic overviews and detailed guidance. In addition, the handbook is aimed at service providers who may be in a position to provide all or part of the services needed to preserve digital materials. It is also relevant to funding agencies who will need to be aware of the implications of the creation of digital materials. Finally, it will be of interest to data creators whose involvement in the preservation of their digital materials is still crucial, despite being restricted by the overarching business needs of their organisation.

The handbook fully recognises that these groups may have different interests and involvement with digital materials at different times. By adopting the life-cycle approach to digital preservation it aims to help identify dependencies, barriers, and mechanisms to assist communication and collaboration between these communities.

The need to tailor the handbook to the needs of individual institutions, including those where digital preservation may be outsourced and those where digital preservation may only be

short-term, means that the handbook needs to be seen as acting as a catalyst for further concerted action within and between institutions.

The broad issues associated with digital preservation are global in nature and examples of good practice, research activity and sources of advice and guidance have been drawn from around the world. However, there is a UK focus in terms of the background to the study and some examples, e.g. legislation, are UK specific. The text of the handbook will indicate a UK focus whenever relevant. It is still hoped that the handbook will be relevant to an international audience as many of the models and references provided are not UK based and are in any case applicable to any country.Wherever their country of origin, the users of the handbook will need to tailor it to their specific needs.

The overall theme of the handbook is that while the issues are complex and much remains to be clarified (and may never be definitively resolved), there is nevertheless much that has already been achieved and much that can be undertaken immediately by all involved in creating and/or acquiring digital materials.This activity will help to protect the initial investment in digital materials creation and offer considerably improved prospects for the long-term.

Web version of the handbook

In such a rapidly developing environment, it will clearly be necessary to maintain the handbook to ensure its currency.This will be achieved by having this electronic version available on the web, which will be updated on an annual basis to ensure currency of web references and cited projects.The print publication provides a convenient reference work which will be complemented and supported by the web version. All urls cited in the handbook are correct as of May 2002 but may be subject to future changes. Users of the print and CD versions should consult the website regularly for current urls and future updates.

Future development and support

The development of the handbook to date has been the result of collaboration and input from a wide range of institutions and individuals. We hope the momentum and focus for future development and maintenance of the handbook has been created by the establishment of the Digital Preservation Coalition in 2001.This will provide the opportunity to link the handbook to supporting materials and training workshops and to add further case studies and exemplars as experience and practice in digital preservation grows.

The Digital Preservation Coalition welcomes feedback and constructive suggestions for improving the online edition of the handbook. Please send comments to:

Email: info@dpconline.org

Mail:

Digital Preservation Coalition
Innovation Centre
York Science Park

The Handbook was first compiled by Neil Beagrie and Maggie Jones and is now maintained and updated by the DPC.

York
YO10 5DG

1.2 How to Use the Handbook

The needs of institutions regarding the digital materials they create and acquire vary considerably. This handbook is intended to provide a bridge between broad, high level overviews and explicit, detailed guidelines applicable to the needs of a specific institution. The strategic overviews are intended to link to operational activities in order to reinforce the need to develop practical procedures grounded firmly in the business mission of the institution. The handbook provides pointers to where to find further guidance and to assist in developing policies and practices which are most applicable to the individual institution.

Ideally, the handbook should be used as a mechanism to help focus thoughts, increase overall understanding, promote training, and act as a catalyst for further action. Nothing will preclude the need for each organisation ultimately to commit the necessary resources to an action plan but this handbook is intended to oil the wheels of that process. While the challenges may seem daunting initially, the overall message of the handbook is that it is possible to take action now, and to make significant progress towards developing the infrastructure that will greatly improve future prospects for quality digital materials being created at this time. It is neither possible nor necessary to wait until all challenges have been overcome before action is taken. Nor is it necessary to wait until the long-term costs are clearly known. It is often never known precisely how much it costs to preserve traditional collections nor has it been necessary to resolve every obstacle definitively before developing effective preservation programmes.

If, as seems clear, society is to rely on assured access to digital as well as non-digital sources of information, then the ways and means of providing assured and continued access must be found.

A range of tools has been used in the handbook, both because of the inconsistent level of existing guidance and also the range of user needs. For some aspects, such as digital imaging, there are numerous excellent sources of practical guidance, while other operational requirements are less fully developed at this stage. All can be expected to change and evolve fairly rapidly so the handbook is advocating an overall approach to preserving digital resources based on sound principles and policies rather than prescriptive formulae. As the crucial importance of digital preservation becomes more widely recognised, an increasing number of valuable sources of guidance are appearing at a rapid rate. While potentially incredibly valuable, their proliferation can make it bewildering to decide which ones are likely to be most applicable for a given situation.

By selecting key See **Exemplars and Further Reading** in each section, the handbook should make it easier to navigate through existing sources of advice, guidance and options. In addition to pointing to existing sources of guidance, a combination of decision tree, summary checklists, selected case studies, and commentary have been used. These are intended to stimulate and promote further thought and discussion but above all, to stimulate action by

institutions to develop digital preservation management policies and strategies appropriate to their needs.

The handbook is intended for a wide and diverse audience, from those who are only beginning to consider managing digital materials to practitioners who have already accumulated considerable theoretical and/or practical experience. It has been written with the intention of allowing quick and easy access to the most appropriate chapters.

Each chapter is preceded by an 'at a glance' guide to its intended primary audience, their assumed level of knowledge, and the purpose of the chapter. The table below is intended to help in deciding which chapters are likely to be of most relevance. It is however not intended to be rigidly prescriptive and anyone wishing to, can of course read the handbook in its entirety!

All readers are encouraged to read Introduction, which provides the background to the development of the handbook and guidance on how it is to be used. A set of definitions and concise explanations of key concepts is provided in **Definitions and Concepts** and a glossary of acronyms and initials use by organisations and projects throughout the handbook is provided in **Acronyms and Initials** of the **Introduction**. Finally a topic index at the back of the handbook allows readers to identify and locate key subjects covered.

Figure 1 Recommended Sections and Audiences	
Audience	**Recommended Sections**
Anyone requiring an introduction to the subject	Digital Preservation
Creators and publishers	Digital Preservation, Organisational Activities and Media and Formats
Funding agencies	Digital Preservation
Operational managers	Institutional Strategies, Organisational Activities and Media and Formats
Operational staff	Organisational Activities and Media and Formats
Senior administrators	Institutional Strategies
Third party service providers	Institutional Strategies, Organisational Activities and Media and Formats

1.3 Definitions and Concepts

Introduction

A major difficulty in any newly emerging discipline, such as digital preservation, is the lack of a precise and definitive taxonomy of terms. Different communities use the same terms in different ways which can make effective communication problematic. The following working set of definitions are those used throughout the handbook and are intended to assist in its use as a practical tool. These definitions will not necessarily achieve widespread consensus among the wide ranging communities the handbook is aiming at, they are offered here as a mechanism to avoid potential ambiguities in the body of the handbook rather than as a definitive gloss. Where they have been taken from existing glossaries, this has been acknowledged.

Access As defined in the handbook, access is assumed to mean continued, ongoing usability of a digital resource, retaining all qualities of authenticity, accuracy and functionality deemed to be essential for the purposes the digital material was created and/or acquired for.

Authentication A mechanism which attempts to establish the authenticity of digital materials at a particular point in time. For example, digital signatures.

Authenticity The digital material is what it purports to be. In the case of electronic records, it refers to the trustworthiness of the electronic record as a record. In the case of "born digital" and digitised materials, it refers to the fact that whatever is being cited is the same as it was when it was first created unless the accompanying metadata indicates any changes. Confidence in the authenticity of digital materials over time is particularly crucial owing to the ease with which alterations can be made.

"Born Digital" Digital materials which are not intended to have an analogue equivalent, either as the originating source or as a result of conversion to analogue form.This term has been used in the handbook to differentiate them from 1) digital materials which have been created as a result of converting analogue originals; and 2) digital materials, which may have originated from a digital source but have been printed to paper, e.g. some electronic records.

Digital Archiving This term is used very differently within sectors.The library and archiving communities often use it interchangeably with digital preservation. Computing professionals tend to use digital archiving to mean the process of backup and ongoing maintenance as opposed to strategies for long-term digital preservation. It is this latter richer definition, as defined under digital preservation which has been used throughout this handbook.

Digital Materials A broad term encompassing digital surrogates created as a result of converting analogue materials to digital form (digitisation), and "born digital" for which there has never been and is never intended to be an analogue equivalent, and digital records.

Digital Preservation Refers to the series of managed activities necessary to ensure continued access to digital materials for as long as necessary. Digital preservation is defined very broadly for the purposes of this study and refers to all of the actions required to maintain access to digital materials beyond the limits of media failure or technological change. Those materials may be records created during the day-to-day business of an organisation;"born-digital" materials created for a specific purpose (e.g. teaching resources); or the products of digitisation projects. This handbook specifically excludes the potential use of digital

24

technology to preserve the original artefacts through digitisation. See also **Digitisation** definition below.

- Long-term preservation - Continued access to digital materials, or at least to the information contained in them, indefinitely.

- Medium-term preservation - Continued access to digital materials beyond changes in technology for a defined period of time but not indefinitely.

- Short-term preservation - Access to digital materials either for a defined period of time while use is predicted but which does not extend beyond the foreseeable future and/or until it becomes inaccessible because of changes in technology.

Digital Publications "Born digital" objects which have been released for public access and either made available or distributed free of charge or for a fee. They may consist of networked publications, available over a communications network or physical format publications which are distributed on formats such as floppy or optical disks.They may also be either static or dynamic.

Digital Records See **Electronic Records**

Digital Resources See **Digital Materials**

Digitisation The process of creating digital files by scanning or otherwise converting analogue materials.The resulting digital copy, or digital surrogate, would then be classed as digital material and then subject to the same broad challenges involved in preserving access to it, as "born digital" materials.

Documentation The information provided by a creator and the repository which provides enough information to establish provenance, history and context and to enable its use by others. See also **Metadata**. "At a minimum, documentation should provide information about a data collection's contents, provenance and structure, and the terms and conditions that apply to its use. It needs to be sufficiently detailed to allow the data creator to use the material in the future, when the data creation process has started to fade from memory. It also needs to be comprehensive enough to enable others to explore the resource fully, and detailed enough to allow someone who has not been involved in the data creation process to understand the data collection and the process by which it was created." (History Data Service)

Electronic Records Records created digitally in the day-to-day business of the organisation and assigned formal status by the organisation.They may include for example, word processing documents, emails, databases, or intranet web pages.

Emulation A means of overcoming technological obsolescence of hardware and software by developing techniques for imitating obsolete systems on future generations of computers.

Life-cycle Management Records management practices have established life-cycle management for many years, for both paper and electronic records. The major implications

25

for life-cycle management of digital resources, whatever their form or function, is the need actively to manage the resource at each stage of its life-cycle and to recognise the inter-dependencies between each stage and commence preservation activities as early as practicable.This represents a major difference with most traditional preservation, where management is largely passive until detailed conservation work is required, typically, many years after creation and rarely, if ever, involving the creator. There is an active and inter-linked life-cycle to digital resources which has prompted many to promote the term "continuum" to distinguish it from the more traditional and linear flow of the life-cycle for traditional analogue materials.We have used the term life-cycle to apply to this pro-active concept of preservation management for digital materials. The rationale for this approach is summed up in the following quotations:"...the prospects for and the costs involved in preserving digital resources over the longer term rest heavily upon decisions taken about those resources at different stages of their life cycle. Decisions taken in the design and creation of a digital resource, and those taken when a digital resource is accessioned into a collection, are particularly influential."(Beagrie and Greenstein 1998)"At each phase of the cycle, electronic records need to be actively managed, according to established procedures, to ensure that they retain qualities of integrity, authenticity and reliability."(PRO 1999)

Metadata Information which describes significant aspects of a resource. Most discussion to date has tended to emphasise metadata for the purposes of resource discovery. The emphasis in this handbook is on what metadata are required successfully to manage and preserve digital materials over time and which will assist in ensuring essential contextual, historical, and technical information are preserved along with the digital object.

Migration A means of overcoming technological obsolescence by transferring digital resources from one hardware/software generation to the next. The purpose of migration is to preserve the intellectual content of digital objects and to retain the ability for clients to retrieve, display, and otherwise use them in the face of constantly changing technology. Migration differs from the refreshing of storage media in that it is not always possible to make an exact digital copy or replicate original features and appearance and still maintain the compatibility of the resource with the new generation of technology.

Reformatting Copying information content from one storage medium to a different storage medium (media reformatting) or converting from one file format to a different file format (file **re-formatting).**

Refreshing Copying information content from one storage media to the same storage media.

1.4 Acronyms and Initials

ADS Archaeology Data Service

> http://ads.ahds.ac.uk

AHDS Arts and Humanities Data Service

> http://ahds.ac.uk

CAMiLEON Creative Archiving at Michigan & Leeds: Emulating the Old on the New

http://newweb2.si.umich.edu/CAMILEON/

Cedars CURL Exemplars in Digital Archiving

http://www.leeds.ac.uk/cedars

CLIR Council on Library and Information Resources

http://www.clir.org

CNI Coalition for Networked Information

http://www.cni.org

CURL Consortium of Research Libraries

http://www.curl.ac.uk

DDI Data Documentation Initiative

http://www.icpsr.umich.edu/DDI/users/dtd/index.html

Update 27 Jan 2006
New link http://www.icpsr.umich.edu/DDI/

DLF Digital Library Federation

http://www.diglib.org

Update 07 Mar 2007

ECUP European Copyright User Platform removed

EPIC European Preservation Information Centre

http://www.knaw.nl/ecpa

EROS Electronic Records in Office Systems

http://www.pro.gov.uk/recordsmanagement/eros/default.htm

Update 10 Aug 2006
Project completed - the EROS project has now ended. This redirected page contains links to relevant guidance: http://www.nationalarchives.gov.uk/recordsmanagement/

HDS History Data Service

http://hds.essex.ac.uk

HEDS Higher Education Digitisation Service

http://heds.herts.ac.uk

InterPARES project International Research on Permanent Authentic Records in Electronic Systems

http://www.interpares.org

27

The Handbook was first compiled by Neil Beagrie and Maggie Jones and is now maintained and updated by the DPC.

Digital Preservation Coalition
November 2008
Website: www.dpconline.org
Email: info@dpconline.org

JISC Joint Information Systems Committee of the Higher and Further Education Councils

http://www.jisc.ac.uk

NDAD UK National Digital Archive of Datasets

http://www.pro.gov.uk/recordsmanagement/uknda

Update 10 Aug 2006
http://www.ndad.nationalarchives.gov.uk/

NEDLIB Networked European Deposit Library

http://www.kb.nl/nedlib

NESLI National Electronic Site Licensing Initiative

http://www.nesli2.ac.uk

NGfL National Grid for Learning

http://www.ngfl.gov.uk

NML National Media Laboratory (United States)

http://www.imation.com/en_US/main.jhtml?Id=64_04

NOF New Opportunities Fund

http://www.nof.org.uk and

http://www.peoplesnetwork.gov.uk

NPO National Preservation Office

http://www.bl.uk/services/preservation

OAIS Open Archival Information System

http://ssdoo.gsfc.nasa.gov/nost/isoas/overview.html

OCLC Online Computer Library Center

http://www.oclc.org/home

PADI Preserving Access to Digital Information

http://www.nla.gov.au/padi

PADS Performing Arts Data Service

http://www.pads.ahds.ac.uk

Update 12 Feb 2007

New location http://ahds.ac.uk/performingarts/

PANDORA Preserving and Accessing Networked Documentary Resources of Australia

http://pandora.nla.gov.au

28

PRISM Project Preservation, Reliability, Interoperability, Security, Metadata

> http://prism.cornell.edu/main.htm

PRO Public Record Office

> http://www.pro.gov.uk

> **Update 10 Aug 2006**
> Now TNA - The National Archives
> http://www.nationalarchives.gov.uk

PURL Persistent Uniform Resource Locator

> http://purl.nla.gov.au

RLG Research Libraries Group

> http://www.rlg.ac.uk

> **Update 04 Aug 2006**
> New link http://www.rlg.org

TASI Technical Advisory Service for Images

> http://www.tasi.ac.uk

TNA The National Archives

> http://www.nationalarchives.gov.uk

UKOLN UK Office for Library and Information Networking

> http://www.ukoln.ac.uk

ULCC University of London Computer Centre

> http://www.ulcc.ac.uk

VADS Visual Arts Data Service

> http://vads.ahds.ac.uk

References

1. Modernising Government. March 1999. Cm 4310. Chapter 5.
   http://www.archive.official-documents.co.uk/document/cm43/4310/4310.htm

2. The seven commissioned reports are available at the UKOLN website
   http://www.ukoln.ac.uk/services/elib/papers/supporting

3. Beagrie, N. and Greenstein, D. (1998). A Strategic Policy Framework for Creating and Preserving Digital Collections.Version 4.0 (Final Draft). ELib Supporting Study p3. Library Information and Technology Centre, South Bank University, London.
   http://ahds.ac.uk/strategic.htm
   Update 26-11-2004

The Handbook was first compiled by Neil Beagrie and Maggie Jones and is now maintained and updated by the DPC.

Version 5 of this document is now available in either PDF or Word format at:
http://www.ahds.ac.uk/about/publications/index.htm

4. Hendley,T. (1998). Comparison of Methods and Costs of Digital Preservation. British Library Research and Innovation Report 109. London: the British Library.
http://www.ukoln.ac.uk/services/elib/papers/tavistock/hendley/hendley.html

5. Hedstrom, M. and Montgomery, S. (1998). Digital Preservation Needs and Requirements in RLG Member Institutions. Mountain View CA: RLG.
http://www.rlg.org/preserv/digpres.html

6. History Data Service. Guideline for Depositors. http://hds.essex.ac.uk/depguide.asp

7. Beagrie, N. and Greenstein, D. (1998). A Strategic Policy Framework for Creating and Preserving Digital Collections.Version 4.0 (Final Draft). ELib Supporting Study p3. Library Information and Technology Centre, South Bank University, London. p. 3.
http://ahds.ac.uk/strategic.htm
Update 26-11-2004
Version 5 of this document is now available in either PDF or Word format at:
http://www.ahds.ac.uk/about/publications/index.htm

8. PRO. (1999). Management,Appraisal and Preservation of Electronic Records.Volume 1: Principles. p. 37.
http://www.nationalarchives.gov.uk/electronicrecords/advice/guidelines.htm