Original: English

CI-2003/WS/3
March 2003

GUIDELINES FOR THE PRESERVATION OF DIGITAL HERITAGE

Prepared by the National Library of Australia

Information Society Division

United Nations Educational, Scientific and Cultural Organization

Table of contents

Website: www.unesco.org/webworld/mdm
Contact: a.abid@unesco.org

Acknowledgments

The Guidelines were principally prepared by Colin Webb with the assistance and input of a number of other staff of the National Library Australia, including Kevin Bradley, Debbie Campbell, Gerard Clifton, Mark Corbould, Maura O'Connor, Margaret Phillips, and Julie Whiting, who provided ideas and in some cases draft text for some chapters; and a number of non-NLA people who provided ideas and comments, including Prof. Arnaldo Coro Antich of Havana, Dr T. Matthew Ciolek of the Australian National University; Mr Simon Davis of National Archives of Australia; Mr Ian Gilmour of ScreenSound Australia; Dr Henry Gladney of California; Mr Roger Harris of Hong Kong; Ms Justine Heazlewood of VERS; Dr. Graeme Johanson of Monash University; Ms Maggie Jones of the (UK) Digital Preservation Coalition; Ms Anne Kenney of Cornell University; Mr Stephen Knight of the National Library of New Zealand; Dr Simon Pockley of the Australian Centre for the Moving Image; Dr Johan Steenbakkers of the Koninklijke Bibliotheek; Mr Hiroyuki Taya of the National Diet Library, Japan; Mr Paul Trezise of GeoScience Australia; and Ms Deborah Woodyard of the British Library.

I must also acknowledge the interest and input from attendees at the various Regional Consultation Meetings, and patient guidance from Mr Abdelaziz Abid of the Division de la Société de l'information, UNESCO, Paris.

While not always able to reflect their comments in the Guidelines, I have learnt much from working with all of them.

Much material in the guidelines is also based on work by other insightful people working in preservation and research programmes around the world: it could hardly be otherwise. To avoid cluttering the text with citations, names and sources have generally been left to the Reading List, except where there is a direct and exclusive link between a comment in the guidelines and a specific source. However, it is most important to acknowledge the important contribution that such people have made, albeit unknowingly, to these Guidelines.

While very gratefully acknowledging all of these inputs, any misinterpretations, misconceptions, ambiguities or errors are almost certainly my own.

Colin Webb
Director of Preservation
National Library of Australia
March 2003

Preface

A large part of the vast amounts of information produced in the world is born digital, and comes in a wide variety of formats: text, database, audio, film, image. For cultural institutions traditionally entrusted with collecting and preserving cultural heritage, the question has become extremely pressing as to which of these materials should be kept for future generations, and how to go about selecting and preserving them. This enormous trove of digital information produced today in practically all areas of human activity and designed to be accessed on computers may well be lost unless specific techniques and policies are developed to conserve it.

Preserving valuable scientific information, research data, media output, digital art, to name but a few areas, poses new problems. If such material is to be accessed in its original form, technical equipment – original or compatible hardware and software - must be maintained alongside the digital files that make up the data concerned. In many cases, the multimedia components of websites, including Internet links, represents additional difficulty in terms of copyright and geography, sometimes making it difficult to determine which country a website belongs to.

UNESCO has been examining these issues with a view to defining a standard to guide governments' preservation endeavours in the digital age. The General Conference adopted Resolution 34 at its 31st session, drawing attention to the ever growing digital heritage in the world and the need for an international campaign to safeguard endangered digital memory. The General Conference also invited the Director-General to prepare a discussion paper for the 2001 Spring session of the Executive Board containing elements of a draft charter on the preservation of born-digital documents, as well as to encourage the governmental and non-governmental organizations and international, national and private institutions to ensure that preservation of the digital heritage be given high priority at the national policy level.

During the meeting of the Organization's Executive Board in May 2001, Member States agreed on the need for rapid action to safeguard digital heritage. The debate was largely inspired by a discussion paper compiled for UNESCO by the European Commission on Preservation and Access (ECPA)[1], an Amsterdam-based non-profit foundation, which outlined the issues involved in digital preservation.

Traditional preservation methods, such as the "legal deposit" used by national libraries to ensure that copies of all printed materials are kept, cannot be applied as such to digital material for a variety of reasons, notably because Web "publications", often draw on data stored on servers in different parts of the world. The sheer volume of data concerned also poses a problem. It is estimated that the Internet features one billion pages whose average lifespan is extremely short, estimated at 44 days to two years.

Considered as the most democratic publishing medium ever, some argue that the ever growing Internet deserves to be preserved as a whole as its pages and discussion forums can be considered a priceless mirror of society.

There are technical problems in ensuring that the digital material that is saved in archives remains accessible in its original form. While the share of total information and art produced around the world on traditional media such as the printed page, analogue tape or film, is declining yearly as compared to material designed for computer access, software and hardware are constantly replaced by more powerful new generations which ultimately become incompatible with their predecessors. This means that within just a few years, material - which often includes sound and moving graphics or pictures, as well as links to Internet sites and, or, databases - becomes inaccessible.

The sheer volume of data to be sifted in order to select what is worthy of preservation is staggering. "The world's total yearly production of print, film, optical, and magnetic content would require roughly 1.5 billion gigabytes of storage. This is the equivalent of 250 megabytes per person for each man, woman, and child on earth," according to a recent study by the School of Information Management and Systems at the University of California at Berkeley.[2]

Another complex issue concerns copyright, including copyright of software required to access digital files. A dazzling array of rights may be associated with websites combining mixed materials from various sources and agreement on the principle of "the right to copy for preservation" still has to be developed worldwide.

While valuable initiatives have been undertaken in many countries to preserve digital heritage, including websites, the ECPA study points to the limits of these efforts, arguing in favour of international standards.

The complexity of the problems involved means that the task of preservation must involve producers of digital information, including software, who should take conservation into consideration as they design their products. Obviously the days are gone when preservation was the sole responsibility of archival institutions.

Co-operation, guidance, leadership and sharing of tasks are all key elements for preservation of digital heritage. Cultural institutions need the co-operation of creators of information and of software producers. Adequate resources and support at policy level are indispensable to ensure that future generations continue to have access to the wealth of digital resources in whose creation we have invested so much over the past decades.

Based on the above findings, UNESCO has developed a strategy for the promotion of digital preservation. This strategy is centred on: a) a wide consultation process with governments, policy makers, producers of information, heritage institutions and experts, the software industry as well as standard-setting organisations; b) dissemination of technical guidelines; c) implementation of pilot projects and; d) and preparation of a draft charter on the preservation of digital heritage for adoption by the General Conference at its 32nd session.

The present document, prepared for UNESCO under contract with the National Library of Australia, introduces general and technical guidelines for the preservation and continuing accessibility of the ever growing digital heritage of the world. This document is intended to be a companion volume of the Draft Charter on the Preservation of the Digital Heritage.

Thanks are due to Colin Webb and the National Library of Australia for preparing the Guidelines and holding the Regional Consultation Meeting on the Preservation of Digital Heritage for Asia and the Pacific, held in Canberra, Australia, 4-6 November 2002. This was the first of a series of similar regional consultation meetings held in Managua, Nicaragua, 18-20 November 2002; Addis Ababa, Ethiopia, 9-11 December 2002; Riga, Latvia, 18-20 December 2002; and Budapest, Hungary, 17-18 March 2003.

These regional meetings were attended by a total of some 175 experts from 86 countries, representing a wide range of stakeholders and disciplines including libraries and archives, Internet service providers, national standardization agencies, software and hardware industry representatives, journalists, lawyers, universities and government authorities. They all contributed useful comments on the draft Guidelines and the Preliminary Draft Charter on the Preservation of the Digital Heritage.

We hope that these Guidelines will prove useful in helping managers and preservation specialists in addressing the complex technical issues facing the preservation and continuing accessibility of the world's digital heritage.

Abdelaziz Abid
Information Society Division
UNESCO

[1] http://unesdoc.unesco.org/images/0012/001255/125523e.pdf
[2] http://www.sims.berkeley.edu/how-much-info

SECTION 1

INTRODUCTORY MATERIALS

Chapter 1. Introduction

Our cultural, scientific and information heritage exists increasingly in digital forms, and increasingly only in digital forms. The technologies we use to create and enjoy the digital heritage have many advantages that explain their extraordinarily rapid take up in many parts of the world.

But there are very serious challenges in keeping our emerging, but already burgeoning, digital heritage usable and available. The media we use to carry and store it are unstable, and the technology needed for access is quickly superseded by newer technologies, wave after wave. As technologies lose support, access to the digital heritage that they fostered is also lost.

These challenges are not only technical in nature; they have organisational and societal dimensions as we struggle with the responsibility of keeping access lines open over extended periods of time, often with insufficient resources and uncertain strategies.

The interest of UNESCO in this situation comes as no surprise. UNESCO exists in part to encourage and enable the preservation and enjoyment of the cultural, scientific and information heritage of the world's peoples. The growth of digital heritage and its vulnerability could hardly go unnoticed.

These Guidelines form a small part of a far-seeing campaign by UNESCO to improve access to digital heritage for all the world's peoples, and to ensure that the means of preserving their digital heritage are in the hands of every community.

The scope and ambition of the Guidelines are constrained. In such a rapidly evolving, but already extensive and complex field, they can only present a small amount of information. In the interests of offering guidance to individuals and organisations who are contemplating a responsibility for preserving digital heritage – frequently from a position of few resources and a plethora of information – it was decided to adopt a principles approach that might serve as a (rather extended) checklist of issues and possibilities that programmes need to take into account.

It is impossible to provide answers to every technical and practical question that will arise in managing digital preservation programmes, so the Guidelines will perhaps be most usefully seen as a guide to the questions that programme managers need to find answers to. However, they are based on a firm conviction that it is time to ask questions that can lead to positive action, rather than continuing to ask questions that merely highlight difficulties.

It is to be hoped that the Guidelines, in conjunction with a wealth of technical information already available from sources listed in the Reading List, will help preservation programme managers identify the decisions they need to make, the actions they need to take, the principles they should take into account, and the practical considerations they need to address.

It is expected that the audience will include cultural and research organisations such as

libraries, archives, museums, research institutes, data archives, publishers, community groups, and others with an interest in and a potential responsibility for preserving digital heritage. Such an audience will include many with a long history of collecting and preserving the world's 'memory heritage' of documents, records, publications, maps, manuscripts, artworks, images, sound recordings, moving imagery, cultural objects, and scientific, research and statistical information. It will also include many coming to digital preservation for a different background, less familiar with the preservation perspectives developed in 'memory' organisations.

These Guidelines were prepared by the National Library of Australia under contract with UNESCO, and are based on extensive review of literature, the Library's own experience, and UNESCO-organised consultations in various regional centres. For more information on inputs and responsibilities, readers should consult the Acknowledgments page; for help on how to use the Guidelines, readers should consult chapter 3: A guide to the Guidelines.

Chapter 2. The UNESCO Draft Charter on the Preservation of the Digital Heritage

INTRODUCTION

The UNESCO Draft Charter on the Preservation of the Digital Heritage presents a compelling case for digital preservation. It is included in the Guidelines to provide a clear link between the two documents, and to present those advocacy and public policy issues that are outside the scope of technical and practical guidelines.

REVISED DRAFT CHARTER ON THE PRESERVATION OF THE DIGITAL HERITAGE

PREAMBLE

The General Conference,

Considering that the disappearance of heritage in whatever form constitutes an impoverishment of the heritage of all nations,

Recalling that the Constitution of UNESCO provides that the Organization will maintain, increase and diffuse knowledge, by assuring the conservation and protection of the world's inheritance of books, works of art and monuments of history and science, that its "Information for All" Programme provides a platform for discussions and action on information policies and the safeguarding of recorded knowledge, and that its "Memory of the World" Programme aims to ensure the preservation and universal accessibility of the world's documentary heritage,

Recognizing that such resources of information and creative expression are increasingly produced, distributed, accessed and maintained in digital form, creating a new legacy – the digital heritage,

Aware that permanent access to this heritage will offer broadened opportunities for creation, communication and sharing of knowledge among all peoples, as well as protection of rights and entitlements and support of accountability,

Understanding that this digital heritage is at risk of being lost and that its preservation for the benefit of present and future generations is an urgent issue of worldwide concern,

Bearing in mind the UNESCO Universal Declaration on Cultural Diversity,

Proclaims the following principles and *adopts* the present Charter.

THE DIGITAL HERITAGE AS A COMMON HERITAGE

Article 1 – Digital heritage

Resources of human knowledge or expression, whether cultural, educational, scientific and administrative, or embracing technical, legal, medical and other kinds of information, are increasingly created digitally, or converted into digital form from existing analogue resources. Where resources are "born digital", there is no other format but the digital original.

Digital materials include texts, databases, still and moving images, audio, graphics, software, and web pages, among a wide and growing range of formats. They are frequently ephemeral, and require purposeful production, maintenance and management to be retained.

Many of these resources have lasting value and significance, and therefore constitute a heritage that should be protected and preserved for current and future generations. This heritage may exist in any language, in any part of the world, and in any area of human knowledge or expression.

Article 2 – Access to the digital heritage

The purpose of preserving the digital heritage is to ensure that it remains permanently accessible. Accordingly, access to digital heritage materials, especially those in the public domain, should be equitable and free of unreasonable restrictions. At the same time, the security of sensitive and personal information should be protected from any form of intrusion.

It is for each Member State to cooperate with relevant organizations and institutions in encouraging a legal and practical environment which would maximise accessibility of the digital heritage. A fair balance between the legitimate rights of creators and other rights holders and those of the public to access digital heritage materials should be reaffirmed and promoted.

GUARDING AGAINST LOSS OF HERITAGE

Article 3 – The threat of loss

The world's digital heritage is at risk of being lost to posterity. Contributing factors include the rapid obsolescence of the hardware and software which brings it to life, uncertainties about resources, responsibility and methods for maintenance and preservation, and the lack of supportive legislation.

Attitudinal change has fallen behind technological change. Digital evolution has been too rapid and costly for governments and institutions to develop timely and informed preservation strategies. The threat to the economic, social, intellectual and cultural potential of the heritage – the building blocks of the future – has not been fully grasped.

Article 4 – Need for Action

Unless the prevailing threats are addressed, the loss of the digital heritage will be rapid and inevitable. Awareness-raising and advocacy is urgent, alerting policy makers and sensitizing the general public to both the potential of the digital media and the practicalities of preservation. Member States will benefit by encouraging legal, economic and technical measures to safeguard the heritage.

Article 5 – Continuity of digital information

The digital heritage is part of the wider continuum of digital information. To preserve digital heritage, measures will need to be taken throughout the information's life cycle. Preservation of digital heritage begins with the design of reliable systems which will produce authentic and stable digital objects.

MEASURES REQUIRED

Article 6 – Developing strategies and policies

Strategies and policies to preserve the digital heritage can be developed, taking into account the level of urgency, local circumstances, available means and future projections. The cooperation of creators, holders of copyright and related rights, and relevant institutions in setting common standards and compatibilities, and resource sharing, will facilitate this.

Article 7 – Defining what should be kept

As with all documentary heritage, selection principles may vary between countries, although the main criteria for deciding what digital materials to keep would be their significance and lasting cultural, scientific, evidential or other value. Selection decisions and any subsequent reviews need to be carried out in an accountable manner, and be based on defined principles, policies, procedures and standards.

Article 8 – Protecting the digital heritage

Member States need appropriate frameworks to secure the protection of their digital heritage. Market forces alone will not achieve this.

As a key element of national preservation policy, archive legislation and legal or voluntary deposit in libraries, archives, museums and other public repositories should embrace the digital heritage. Copyright and related rights legislation should allow preservation processes to be undertaken legally by such institutions.

The right to permanent access to legally deposited digital heritage materials, within reasonable restrictions, should be guaranteed without causing prejudice to their normal exploitation.

Legal and practical frameworks for authenticity are crucial to prevent manipulation or intentional alteration of digital heritage. Both require that the content, functionality of files and documentation be maintained to the extent necessary to secure an authentic record.

Article 9 – Promoting cultural diversity

The digital heritage is inherently unlimited by time, geography, culture or format. It is culture-specific, but potentially accessible to every person in the world. Minorities may speak to majorities, the individual to a global audience.

The digital heritage of all regions, countries and communities should be preserved and made accessible, creating over time a balanced and equitable representation of all peoples, nations, cultures and languages.

RESPONSIBILITIES

Article 10 – Roles and responsibilities

It is for each Member State to designate one or more agencies to take coordinating responsibility for the preservation of digital heritage, and to provide the necessary staff and resources. The sharing of tasks and responsibilities may be based on existing roles and expertise.

Measures should be taken to:

(a) urge hardware and software developers, creators, publishers, producers, and distributors of digital materials as well as other private sector partners to cooperate with national libraries, archives, museums and other public heritage organizations in preserving the digital heritage;

(b) develop training and research, and share experience and knowledge among the institutions and professional associations concerned;

(c) encourage universities and other research organizations to ensure preservation of research data.

Article 11 – Partnerships and cooperation

Preservation of the digital heritage requires sustained efforts on the part of governments, creators, publishers, relevant industries and heritage institutions.

In the face of the current digital divide, it is necessary to reinforce international cooperation and solidarity to enable all countries to ensure creation, dissemination, preservation and continued accessibility of their digital heritage.

Industries, publishers and mass communication media are urged to promote and share knowledge and technical expertise.

The stimulation of education and training Programmes, resource-sharing arrangements, and dissemination of research results and best practices will democratize access to digital preservation techniques.

Article 12 – The role of UNESCO

UNESCO, by virtue of its mandate and functions, has the responsibility to:

(a) take the principles set forth in this Charter into account in the functioning of its Programmes and promote their implementation within the United Nations system and by intergovernmental and international non-governmental organizations concerned with the preservation of the digital heritage;

(b) serve as a reference point and a forum where Member States, intergovernmental and international non-governmental organizations, civil society and the private sector may join together in elaborating objectives, policies and projects in favour of the preservation of the digital heritage;

(c) foster cooperation, awareness-raising and capacity-building, and establish

standard ethical, legal and technical guidelines, as a companion sourcebook to this Charter;

(d) determine, on the basis of the experience gained over the next six years in implementing the present Charter and the Guidelines, the need for further standard-setting instruments for the promotion and preservation of the digital heritage.

Chapter 3. Guide to the guidelines

INTRODUCTION

3.1 Aim

The Guidelines have been prepared to address a number of different audiences, and to cover a very large territory of information. This chapter aims to serve as a road map, helping readers find ways of using the Guidelines that will suit them best. (The table of contents, index, and cross referencing at the end of each chapter all have similar aims.)

3.2 Audiences

The consultation process revealed at least four audiences who can be expected to use the Guidelines, each with different but overlapping needs.

Policy makers requiring very high level information regarding the case for digital preservation, and sufficient outline to inform their policy commitment.

The Guidelines address these needs through:

- The inclusion of the UNESCO Draft Charter on the Preservation of the Digital Heritage in chapter 2
- The summary of principles in chapter 5
- The *In a nutshell* summaries at the start of most chapters.

High level managers seeking to understand the conceptual basis for digital preservation and the management issues their programmes will face.

The Guidelines address these needs through:

- The chapters in Section 2, which all have a management focus
- The *key management challenges* and *principles* sections of the more detailed chapters on processes found in Section 3
- The summary of principles in chapter 5

Line managers involved in day-to-day decisions, who need both a good conceptual understanding, and insight into the detailed issues they will have to manage.

The Guidelines address these needs through:

- The conceptual overview chapters in Section 2, (especially chapters 7, 8 and 10)
- The detailed chapters in Section 3, each looking at issues associated with particular processes

Technical practitioners, needing detailed technical guidance as well as a good perspective on how the various technical issues and processes fit together to make an integrated programme with coherent preservation objectives.

The Guidelines do not attempt to address the need for detailed technical information, which was both too situation-specific and too quickly outdated to fit easily into the Guidelines. However, it is recommended that UNESCO create a Technical Information section of the Web version of these Guidelines, where technical standards, manuals, and useful tips can be sourced.

The Guidelines should, however, provide technical practitioners with an integrating perspective through the arrangement of the chapters. The Reading List should also provide a useful guide to further study.

3.3 Content

The arrangement of chapters is significant.

Section 1 contains introductory materials, including the case for digital preservation, argued by the UNESCO Draft Charter (chapter 2); a note on terminology needing to be understood before embarking on reading the Guidelines (chapter 4); and a summary of principles (chapter 5).

Section 2 presents a management perspective. It begins with an explanation of *digital heritage* and why it is threatened (chapter 6), then introduces *digital preservation* (chapter 7), the nature of *digital preservation programmes* (chapter 8), the basis for deciding what *preservation responsibility* to accept (chapter 9), the *management of preservation programmes* (chapter 10), and opportunities to *work cooperatively* (chapter 11).

Section 3 presents a more detailed and process-focused view, discussing each of the major areas of responsibility in managing digital heritage preservation, beginning with *selection of what is important enough to keep* (chapter 12), *working with the producers* of digital heritage (chapter 13), *taking control* of materials – transferring, identifying and describing them (chapter 14), *managing rights issues* (chapter 15), looking after *authenticity and data protection* (chapter 16), and finding ways to *maintain the means of providing access* (chapter 17) – the core area of preservation uncertainty. This chapter is structured differently from the others as it seeks to compare a range of options.

The section finishes with some suggested *starting points for programmes* as a stimulus for discussion and thinking, and a proposed set of *minimum expectations* for programmes seeking to undertake some kind of digital preservation programme (chapter 18).

Section 4 contains a selective glossary of terms and an extensive reading list, as well as references to good resources for keeping up to date.

3.4 For programmes with few resources

The Guidelines accept a special responsibility to offer some guidance for those people

struggling to set up programmes with extremely limited resources. Each chapter in section 3 includes some suggestions specifically aimed at this need.

3.4 Case studies

A number of chapters in section 3 include brief case studies. These are almost all fictionalised cases, based on real experiences. Fictionalising allows certain aspects to be emphasised to illustrate a particular issue without misrepresenting the actual programmes on which they may have been based.

Chapter 4. A note on terminology

INTRODUCTION

4.1 Aim

A few terms have been used in the Guidelines in ways that may fall outside normal usage. Because they are core terms used repeatedly in these Guidelines, it is important to explain their use from the beginning.

A number of other terms, used less idiosyncratically, are explained in the Glossary in section 4.

4.2 Terms

Digital preservation is used to describe the processes involved in maintaining information and other kinds of heritage that exist in a digital form. In these Guidelines, it does *not* refer to the use of digital imaging or capture techniques to make copies of non-digital items, even if that is done for preservation purposes. Of course, digital copying (also known as digitisation, or digitalisation), may well produce digital heritage materials needing to be preserved.

Digital materials is generally used here as a preferred term covering items of digital heritage at a general level. In some places, ***digital object*** or ***digital resource*** have also been used. The terms have been used interchangeably and generically: they do not imply a particular kind of item unless that is clearly stated.

Preservation programme is used to refer to any set of coherent arrangements aimed at preserving digital materials. More commonly used terms such as *digital archive* and *digital repository* have been avoided because of their potential ambiguities: *archive* has different meanings for the records management community and the ICT community, whereas both *archive* and *repository* may imply a single storage site – not an appropriate implication where very distributed arrangements may be in place.

Of course, the term *programme* also comes with some baggage. It should be understood to cover all the aspects of preservation responsibility, including policy and strategy as well as implementation.

Presentation, re-presentation are used to describe the processes of providing access to digital materials. The second term has been consistently (and idiosyncratically) hyphenised to emphasise the understanding that digital preservation seeks to re-present what was previously presented.

Chapter 5. A summary of principles

INTRODUCTION

5.1 Aims

The purpose of this chapter is to bring together the main statements of principles from throughout the Guidelines, as a summary for managers.

5.2 Principles

5.2.1 Heritage

1. Not all digital materials need to be kept, only those that are judged to have ongoing value: these form the digital heritage.
2. For those materials that warrant keeping, continuity of survival and accessibility is critical. The chances of recovering lost access to large amounts of data are very slim. Continuity requires sustained, direct action (called *digital preservation*) rather than passive 'benign neglect.'

5.2.2 Digital preservation

3. Digital materials cannot be said to be preserved if access is lost. The purpose of preservation is to maintain the ability to present the essential elements of authentic digital materials.
4. Digital preservation must address threats to all layers of the digital object: physical, logical, conceptual and essential.

5.2.3 Responsibility

5. Digital preservation will only happen if organisations and individuals accept responsibility for it. The starting point for action is a decision about responsibility.
6. Everyone does not have to do everything; everything does not have to be done all at once.
7. Comprehensive and reliable preservation programmes are highly desirable, but they may not be achievable in all circumstances of need. Where necessary, it is usually better for non-comprehensive and non-reliable action to be taken than for no action at all. Small steps are usually better than no steps.
8. In taking action, managers should recognise that there are complex issues involved. It is important to do no harm. Managers should seek to understand the whole process and the objectives that eventually need to be achieved, and avoid steps that will jeopardise later preservation action.
9. Acceptance of responsibility should be explicitly and responsibly declared, taking account

of the likely implications for other preservation programmes and for other stakeholders.

5.2.4 Deciding what to keep

10. Selection decisions should be informed, consistent and accountable.

11. A decision to preserve can be made subject to later review; a decision not to preserve is usually final.

5.2.5 Working with producers

12. Currently, preservation efforts have to work against the prevailing trend of digital technology and how it is developed and used.

13. Digital materials are very often created without long-term preservation intentions in mind.

14. Working with producers to influence the standards and practices they use, and to increase their awareness of preservation needs, are important investments.

5.2.6 Rights

15. Preservation programmes must clarify their legal right to collect, copy, name, modify, preserve and provide access to the digital materials for which they take responsibility.

5.2.7 Control

16. Digital heritage materials must be moved to a safe place where they can be controlled, protected and managed for preservation.

17. Digital heritage materials must be uniquely identified, and described using appropriate metadata for resource discovery, management and preservation.

18. Taking the right action later depends on adequate documentation. It is easier to document the characteristics of digital resources close to their source than it is to build that documentation later.

19. Preservation programmes should use standardised metadata schemas as they become available, for interoperability between programmes.

20. The links between digital objects and their metadata must be securely maintained, and the metadata must be preserved.

5.2.8 Authenticity and data protection

21. Authenticity is a critical issue where digital objects are used as evidence. It may also be important for other kinds of digital heritage.

22. Data that underlies digital objects must be safely stored and managed if there is to be any chance of re-presenting authentic objects to users.

23. Digital preservation programmes are subjected to increased authenticity concerns because they so frequently have to use processes that involve change.

24. Authenticity is best protected by measures that ensure the integrity of data is not compromised, and by documentation that maintains the clear identity of the material.

25. Data protection is built on the principles of system security and redundancy. For preservation programmes, redundancy must include securely stored backups designed around the long-term maintenance of data rather than a cycle of overwriting old data with new.

5.2.9 Maintaining accessibility

26. The goal of maintaining accessibility is to find cost-effective ways of guaranteeing access whenever it is needed, both in the short-term and the long-term.

27. Standards are an important foundation for digital preservation, but many programmes must find ways to preserve access to poorly standardised materials, in an environment of changing standards.

28. Preservation action should not be delayed until a single 'digital preservation standard' appears.

29. Digital data is always dependent on some combination of software and hardware tools for access, but the degree of dependency on specific tools determines the range of preservation options.

30. It is reasonable for programmes to choose multiple strategies for preserving access, especially to diverse collections. They should consider the potential benefits of maintaining the original data streams of materials as well as any modified versions, as an insurance against the failure of still uncertain strategies.

31. Strategies for preserving accessibility do not stand alone: they are supported by other responsibilities, and many strategies can be combined to good effect.

32. Preservation programmes are often required to judge acceptable and unacceptable levels of loss, in terms of items, elements, and user needs.

5.2.10 Management

33. Waiting for comprehensive, reliable solutions to appear before taking responsible action will probably mean material is lost.

34. Preservation programmes require good management that consists largely of generic management skills combined with enough knowledge of digital preservation issues to make good decisions at the right time.

35. Digital preservation incorporates the assessment and management of risks.

36. programmes are usually faced with more material and more issues than they can cope with, so they must set priorities.

37. The costs of preservation programmes are hard to estimate because they encompass so much uncertainty, including evolving techniques, changing technologies and very long timeframes. Costs may well be lower per unit of information than for non-digital materials, but the amount of information to be managed in digital form is very large so total costs are also likely to remain high, including set-up costs and significant recurrent costs.

38. Preservation programmes may start as pilot projects but they eventually need to establish

sustainable business models.

39. While suitable service providers may be found to carry out some functions, ultimately responsible for achieving preservation objectives rests with preservation programmes, and with those who oversee and resource them.

5.2.11 Working together

40. Working collaboratively is often a cost effective way to build preservation programmes with wide coverage, mutual support and the required expertise.

41. Collaboration involves costs and choices as well as potential benefits.

SECTION 2

MANAGEMENT PERSPECTIVES

Chapter 6. Understanding digital heritage

INTRODUCTION

6.1 Aims

The purpose of this chapter is to introduce the concepts of *digital heritage* and *digital continuity*. The chapter aims to help readers understand the value and range of digital heritage materials and the threats to their survival. These are important understandings for managers as well as those designing and implementing programmes.

6.2 In a nutshell

Digital heritage is made up of computer-based materials of enduring value that should be kept for future generations. Digital heritage emanates from different communities, industries, sectors and regions. Not all digital materials are of enduring value, but those that are require active preservation approaches if continuity of digital heritage is to be maintained.

MANAGEMENT PERSPECTIVE

6.3 Heritage and digital heritage

Heritage is explained in UNESCO documents as "our legacy from the past, what we live with today, and what we pass on to future generations." A heritage is something that is, or should be, passed from generation to generation because it is valued.

The idea of cultural heritage is a familiar one: those sites, objects and intangible things that have cultural, historical, aesthetic, archaeological, scientific, ethnological or anthropological value to groups and individuals. The concept of natural heritage is also very familiar: physical, biological, and geological features; habitats of plants or animal species and areas of value on scientific or aesthetic grounds or from the point of view of conservation.

Is there an emerging *digital heritage*?

According to the Draft Charter for the Preservation of Digital Heritage:

> Resources of human knowledge or expression, whether cultural, educational, scientific and administrative, or embracing technical, legal, medical and other kinds of information, are increasingly created digitally, or converted into digital form from existing analogue resources. Where resources are "born digital", there is no other format but the digital original.
>
> Digital materials include texts, databases, still and moving images, audio, graphics, software, and web pages, among a wide and growing range of formats. They are frequently ephemeral, and require purposeful production, maintenance and management to be retained.

> Many of these resources have lasting value and significance, and therefore constitute a heritage that should be protected and preserved for current and future generations. This heritage may exist in any language, in any part of the world, and in any area of human knowledge or expression.

Using computers and related tools, humans are creating and sharing digital resources – information, creative expression, ideas, and knowledge encoded for computer processing - that they value and want to share with others over time as well as across space. This is evidence of a digital heritage. It is a heritage made of many parts, sharing many common characteristics, and subject to many common threats.

Definitions of heritage need to be seen in context. For example, UNESCO defines a *world heritage* made up of globally outstanding sites of cultural and natural value that should be preserved; many national and state legislatures also define their own national, regional or state heritage. However, heritage value may also be based on what is important at a group or community level. Heritage materials can exist well beyond the limits suggested by national legislation or international conventions. Anything that is considered important enough to be passed to the future can be considered to have heritage value of some kind.

This digital heritage is likely to become more important and more widespread over time. Increasingly, individuals, organisations and communities are using digital technologies to document and express what they value and what they want to pass on to future generations. New forms of expression and communication have emerged that did not exist previously. The Internet is one vast example of this phenomenon.

It is also likely that the development of tools to support greater multi-lingual and multi-script use of the Internet will lead to further rapid growth in digital heritage in parts of the world that are currently disadvantaged by the predominant use of English on the Internet.

Making sure this burgeoning digital heritage remains available is thus a global issue relevant to all countries and communities.

6.4 Types of digital heritage

Over time, new types of digital heritage can be expected to emerge: we have already seen the enabling power of the technology in forms as diverse as word processing, email, websites, relational databases, computer models and simulations, digital audio and video, space imagery and computer games. At the time of writing, digital heritage includes a great range of materials including (by no means exhaustively):

- Electronic publications, being information that is made available for wide readership. Publications are distributed in various ways including online via the World Wide Web, or on portable carriers such as CDs, DVDs, floppy disks and various electronic book devices. Some publications manage to combine both online and portable carrier access to different parts of the publication. As well as their means of distribution, digital publications may be classified by genres, some familiar from traditional publishing formats like monographs and serials, and others less easily defined like websites and e-zines. Some publications are released as complete items, but others evolve over time, their creators taking advantage of the interactive potential of the Internet. Print

publishing continues to grow, but increasingly publications are appearing in digital versions, increasingly in digital-only versions. Both commercial and non-commercial publishers produce digital publications, as do millions of other people who would not see themselves as publishers at all

- 'Semi-published' materials such as pre-print papers and theses held in e-print and other archives available for restricted use within specific communities such as universities and scholarly societies
- Organisational and personal records of activity, transactions, correspondence, etc. A very large part of the world's business and government records now exist in electronic record keeping systems. Email, messages to discussion lists and bulletin boards, web diaries, 'blogs' and 'cams' – dynamic, informal interactions enabled by digital technology - may also include important digital records amongst a tidal wave of data
- Datasets collected to record and analyse scientific, geospatial, spatial, sociological, demographic, educational, health, environmental and other phenomena
- Learning objects used in technology-assisted education
- Software tools such as databases, models, simulations, and software applications
- Unique unpublished materials that may include research reports, oral history and folklore recordings
- Electronic 'manuscripts' such as drafts of works and personal correspondence
- Entertainment products from the film, music, broadcasting and games industries, both commercial and non-commercial
- Digitally generated artworks and documentary photographs
- Digital copies of images, sound, text and three-dimensional objects from non-digital originals.

Many of these materials exist only in a digital form (even if carried on a physical carrier of some kind). With no equivalent non-digital version, their content is especially vulnerable to the influences that threaten digital materials.

There are also rapidly growing collections of digital copies. Having originally been generated from non-digital sources, these might appear to be less vulnerable, but many of them are the only surviving version of originals that have since been damaged, lost or dispersed.

6.5 Digital continuity

Continuity of the digital heritage is profoundly important. Increasingly this is a heritage that documents the actions of governments, the results of scientific research, the debate of ideas, the aspirations and imagination of communities, the histories of the current and coming world.

If these are not to be lost or distorted, continuity is required: continuity of production, continuity of survival, and continuity of access. This must be achieved in the face of many threats:

- The carriers used to store these digital materials are usually unstable and deteriorate

within a few years or decades at most

- Use of digital materials depends on means of access that work in particular ways: often complex combinations of tools including hardware and software, which typically become obsolete within a few years and are replaced with new tools that work differently
- Materials may be lost in the event of disasters such as fire, flood, equipment failure, or virus or direct attack that disables stored data and operating systems
- Access barriers such as password protection, encryption, security devices, or hard-coded access paths may prevent ongoing access beyond the very limited circumstances for which they were designed
- The value of the material may not be recognised before it is lost or changed
- No one may take responsibility for the material even though its value is recognised
- Those taking responsibility may not have adequate knowledge or facilities
- There may be insufficient resources available to sustain preservation action over the required period
- It may not be possible to negotiate legal permissions needed for preservation
- There may not be the time or skills available to respond quickly enough to a sudden and large change in technology
- The digital materials may be well protected but so poorly identified and described that potential users cannot find them
- So much contextual information may be lost that the materials themselves are unintelligible or not trusted even when they can be accessed
- Critical aspects of functionality, such as formatting of documents or the rules by which databases operate, may not be recognised and may be discarded or damaged in preservation processing.

The steps taken to maintain continuity in the face of these and other threats have come to be called *digital preservation* – a new form of preservation specifically concerned with digital heritage materials.

SPECIAL CONSIDERATIONS

6.6 The stability of the Internet as a specific risk scenario

The Internet is an interesting scenario in which many of these threats are played out. In assessing the risks associated with the Internet, it is necessary to distinguish between two overlapping but different concerns.

One concern is to look at the Internet as a whole. There is no central agency that can decide what happens to material made available through the Internet. Typically, users connect with a kaleidoscope of information for which no single creator, publisher, or any other agency is responsible. Is this smorgasbord of information, and the experience of using it, to be lost

because it is no one's responsibility?

A different viewpoint is to look at individual resources published through the Internet. These may also be quite unstable, but their instability reflects the decisions and actions of their owners in deleting, changing, moving, or renaming them. The loss of materials in this environment is attributable to how they are managed at a local level, exposed to many of the same threats that other kinds of digital materials experience.

While it is largely beyond the power of Internet users to control whether information remains available, it is very much within the power of those owning and managing digital objects and sites. If they are committed to maintaining access, it is generally within their power to do so.

However, the Internet does present some special risks. For example:

- There is a strong novelty factor, so some publishers choose to change things frequently – sometimes the way information looks, sometimes also the content
- Many Internet resources are a virtual composite drawn from a number of sources, which may not be stored together anywhere. Changes in one part may destroy the whole
- The sense of global access may lead some information managers to assume they will be able to rebuild their information if it is lost, ignoring the fact that their information exists on a local system and is vulnerable to damage or loss associated with that system. There is a danger that information managers may fail to take the normal backup and security measures that they would automatically take in a stand-alone system
- It is possible to publish digital materials on the Internet quite easily and cheaply, so many 'publishers' have no plans for maintaining their publications or the means to do so: their works are truly ephemeral.

CASE STUDIES

(These fictionalised case studies have been chosen to illustrate just a few examples of digital heritage material – not necessarily to endorse the way they are managed.)

- A government department has recently issued personal computers to all staff so they can produce their own letters, internal memos, reports, and send messages by email. The department issues a directive that all final documents and important drafts, as well as business emails, must be filed for long-term retention. (They are part of the digital heritage.) However, personal emails and rough drafts do not need to be kept.
- An isolated rural community has long been concerned that its traditional cohesion is being lost along with respect for its way of life. Community elders decide to record everything they can about the community's traditions, and use a computer network to record and share it. This becomes a focus of renewed interest and pride in community life among almost all members of the community, and a source of shared income as selected aspects of the database are made available to authorised outsiders. The community agrees that the growing database must be kept, and should be managed by

the community.

- A recording studio has been making digital recordings for the past 10 years. The masters are stored on a variety of tapes and CDs that are sometimes used to put together new products for local record companies, but mostly just left in storage. Every 12 months the manager sorts through and throws out any old tapes that don't look interesting or sound like they are on the way out.
- A university teacher sets up a website to encourage discussion within her discipline. While she regularly goes to conferences and publishes papers in scholarly journals, she finds the best debate in her field happens on her web diary attached to her web site. She worries that it will be lost without trace, and that future students and researchers will have no idea how certain concepts were first discussed.
- A research institute studies the water flows and flood levels in a major river system, recording comprehensive data over many decades and using various computer simulations to model the effects of different rainfall events in catchment areas. As land use patterns change, they notice changes in their data.

REFERENCES – **where to look for more information**

Cross references

Threats see Risk management, chapter 10

Chapter 7. Understanding digital preservation

INTRODUCTION

7.1 Aims

The purpose of this chapter is to help those who may be responsible for preserving digital heritage materials understand the basic nature, objectives and strategies of digital preservation. These are important understandings for managers as well as those designing and implementing programmes.

7.2 In a nutshell

Digital preservation consists of the processes aimed at ensuring the continued accessibility of digital materials. To do this involves finding ways to re-present what was originally presented to users by a combination of software and hardware tools acting on data. To achieve this requires digital objects to be understood and managed at four levels: as physical phenomena; as logical encodings; as conceptual objects that have meaning to humans; and as sets of essential elements that must be preserved in order to offer future users the essence of the object.

MANAGEMENT PERSPECTIVE

7.3 Digital preservation

Digital preservation can be seen as all those processes aimed at ensuring the continuity of digital heritage materials for as long as they are needed.

The most significant threats to digital continuity concern loss of the means of access. Digital materials cannot be said to be preserved if the means of access have been lost and access becomes impossible. The purpose of preserving digital materials is to maintain accessibility: the ability to access their essential, authentic message or purpose.

7.4 A 'performance' approach to digital preservation

There is an underlying similarity in the way digital objects are accessed in the present, and how they will be accessed in any future use. In both cases, access can be seen as a *performance*. [1]

Digital objects are made accessible by applying software and hardware tools to data in order to create a presentation or performance that has meaning to a user. It may be the presentation

[1] This concept is well discussed in Heslop H, Davis S (2002) (unpublished). *An Approach to the Preservation of Digital Records*. National Archives of Australia, Canberra

of a word processing document, or a piece of recorded audio, or a Web page, or results from a database query, or any other kind of digital object depending on the way the data is encoded and on the actions the tools are programmed to perform. We expect that if we apply the same tools to the same data we will get a repeat performance each time.

Digital preservation must work in the same way, somehow re-presenting what are judged to be the essential elements of the original performance when required at some later time.

Conceptualised in this way, digital preservation can be seen as straightforward. Indeed, copying data from carrier to carrier, and providing the right tools to recreate the intended performance will preserve continuity of access to most digital objects.

However, this simple model encompasses great complexities: it may be hard to define the performance that must be re-presented; it is usually difficult to work out what tools are needed once the original ones have been lost; the tools themselves typically rely on other tools that also may have been superseded; and it may be difficult to find tools that will create the required performance in a reliable, cost-effective and timely way, especially in the context of many thousands, millions or more of digital objects.

Despite such underlying complexities, the performance model helps in recognising what digital preservation programmes must aim for: the best means of re-presenting what users needs to access.

7.5 Understanding the materials being preserved

Preservation programmes must deal with digital objects in four guises:

- As **physical objects**, consisting of 'inscriptions' (usually binary states of 'on-ness' or 'off-ness') on carrier media such as computer disks or tapes. (Despite the impression of that they exist in 'cyberspace', even online resources must exist on physical carriers somewhere)
- As **logical objects** consisting of computer readable code, whose existence at any particular time depends on the physical inscriptions but is not tied to any particular carrier
- As **conceptual objects** that have meaning to humans, unlike the logical or physical objects that encode them at any particular time. (This is recognisable as the performance presented to a user)
- As bundles of **essential elements** that embody the message, purpose, or features for which the material was chosen for preservation.

This multi-layered nature of digital objects has profound implications for digital preservation. Preservation means different things for each layer. [2]

Preservation programmes for non-digital heritage have traditionally worried about preserving the physical object as the embodiment of the object's meaning. However, individual physical manifestations of a digital object are almost inevitably lost, one after another, because the

[2] This concept is adapted from Thibodeau K (2002).

media used for physical storage are typically unstable and liable to short-term deterioration. Preservation requires a succession of data transfers from one physical carrier to another.

Despite this shift in focus from the physical object to a conceptual object inherent in digital preservation, it must never be forgotten that digital objects cannot survive without some kind of appropriate physical form.

The logical encoding normally has a much longer life than any particular physical inscription, but it is by no means sacrosanct. As the layers of technology used for access – hardware such as computer processors, disk drives and peripheral equipment, and many layers of software such as operating systems, specific applications, and presentation tools - become obsolete, it may be necessary to change the logical encoding so that it can present the same conceptual object using different technology.

The conceptual object is the ultimate focus of preservation concern; as noted above, it is at this level that digital objects convey meaning to human users.

However, for most digital objects there is a further layer that must be considered. Many objects consist of several elements, some of which are more important than others in carrying the object's essential message. Preservation programmes have to decide which sub-set of elements should be preserved for re-presentation to users.

7.6 Strategies for preserving digital materials

Digital preservation involves choosing and implementing an evolving range of strategies to achieve the kind of accessibility discussed above, addressing the preservation needs of the different layers of digital objects. The strategies include:

- Working with producers (creators and distributors) to apply standards that will prolong the effective life of the available means of access and reduce the range of unknown problems that must be managed
- Recognising that it is not practical to try to preserve everything, selecting what material should be preserved
- Placing the material in a safe place
- Controlling material, using structured metadata and other documentation to facilitate access and to support all preservation process
- Protecting the integrity and identity of data
- Choosing appropriate means of providing access in the face of technological change
- Managing preservation programmes to achieve their goals in cost-effective, timely, holistic, proactive and accountable ways.

REFERENCES – **where to look for more information**

Cross references

Some issues and concepts in this chapter are discussed further elsewhere in these Guidelines:

Essential elements see chapters 12 and 17

Strategies for preservation see appropriate chapters in section 3

Offsite references (all links viewed March 2003)

- Heslop H., Davis S. (2002) (unpublished). *An Approach to the Preservation of Digital Records*. National Archives of Australia, Canberra
- Thibodeau K. (2002). Overview of Technologic al Approaches to Digital Preservation and Challenges in Coming Years. In: *The State of Digital Preservation: An International Perspective – Conference Proceedings*, Documentation Abstracts, Inc., Institutes for Information Science, Washington, D.C., April 24025, 2002. Council on Library and Information Resources, Washington, D.C. http://www.clir.org/pubs/reports/pub107/thibodeau.html

Chapter 8. Understanding digital preservation programmes

INTRODUCTION

8.1 Aims

In this chapter the reader will find high-level information on the responsibilities, functions and characteristics of comprehensive and reliable digital preservation programmes. This information is important for managers as well as those designing programmes.

8.2 In a nutshell

Preservation programmes have certain responsibilities and functions that have been defined, at least at a conceptual level. Comprehensive programmes must take control of appropriate digital materials and ensure they remain understandable and usable as authentic copies. This generally involves taking the materials, properly prepared, along with associated documentation or metadata, into an archival digital storage system of some kind, where they can be managed to deal with the threats of data loss an technological change. The characteristics or attributes of programmes that can be relied upon to deliver ongoing digital preservation have also been described, in terms of responsibility, viability, sustainability, technical suitability, security, and accountability.

MANAGEMENT PERSPECTIVE

8.3 Some concepts

8.3.1 Preservation programmes

In these Guidelines, the sets of arrangements put in place to give effect to digital preservation are called *preservation programmes*. This is a broad concept that includes policy as well as practical aspects of implementation.

8.3.2 Safe places

These Guidelines assume that digital heritage materials must be moved from an operating environment to a *safe place* or archive where they can be protected from the influences that threaten them at the physical and logical levels, and where they can be managed for ongoing accessibility.

(There is a counter argument that says digital materials are much more likely to survive if they remain in frequent use, because someone will then make the effort to keep them accessible. Material in a 'dead' archive is more likely to be neglected and to miss out on required preservation action until it is too late. This argument points to two important truths:

preservation action should not be neglected; and material that is in demand is more likely to survive than material that is not used. However, this argument cannot be sustained for heritage materials, which must be preserved even though they will often receive low levels of use. Whether material is frequently used or not, there must be a copy that is stored and managed securely if it is to survive, even if this involves creating a safe place within a working environment.)

8.3.3 Information packages

Digital information objects are generally not understandable or re-presentable by themselves: users need help to use them. Preservation depends on maintaining digital objects **and** any information and tools that would be needed in order to access and understand them. Together, these can be considered to form an *information package* that must be managed either as a single object or as a virtual package (with the object and associated information tools linked but stored separately).

8.4 Responsibilities of comprehensive preservation programmes

Preservation programmes that aim to be comprehensive are responsible for:

- Negotiating for and accepting appropriate digital materials from producers
- Controlling the material sufficiently to support its long term preservation
- Working out for whom the material is being kept and who will need to be able to understand it
- Ensuring that the material will remain understandable to this defined community of expected users
- Ensuring that the material is protected against all likely threats, and enabling the material to be accessed and its authenticity trusted
- Making the preserved material available to the designated community of users as appropriate
- Advocating good practice in the creation of digital resources.

8.5 Functions of comprehensive preservation programmes

To fulfil these responsibilities, preservation programmes that seek to be comprehensive must carry out the following functions:

8.5.1 Creating or finding a safe place

Preservation programmes must identify a safe place where digital materials can be stored and managed. Because the concept of the preservation programme allows for distributed arrangements and shared responsibilities, it is quite conceivable that while some programmes will create their own repositories, others may decide to look for a suitable 'safe place' operated and managed by someone else. By definition, a decision to manage heritage materials through someone else's repository does not obviate the ultimate responsibility of the preservation programme concerned.

These Guidelines, including the notes on functions below and the notes on protecting data in Chapter 16, can be used to suggest the criteria by which potential 'safe places' may be identified and assessed.

8.5.2 Ingest

The processes of receiving, preparing and transferring digital materials into the archival system are usually referred to as *ingest*.

Preparing material for entry into the archival information system is critical to the way the whole system is managed. It involves a number of important steps that may determine how easy or difficult it is to maintain the archived information packages in the system. These steps include:

- Applying collection policies and selection criteria to assess whether material is in scope so that it can be sought, or accepted if submitted
- Clarifying or negotiating rights issues with rights owners
- Checking the quality of the submitted information package, including its completeness, the functionality of its component parts, its authenticity, and whether it contains unwanted material such as viruses
- Labelling the material with unique identifiers
- Assessing the elements that must be maintained, and assigning preservation objectives
- Setting retention and review periods for the material, if appropriate
- Checking and if necessary upgrading the documentation that describes the material, including the technical and preservation metadata
- Assessing the file format(s) and deciding if they need to be changed to comply with the Programme's policy on what formats it will manage (which may be restrictive or unrestricted)
- If necessary, changing the file formats to comply with the policy
- Adjusting the documentation to reflect any changes.

Once the digital object and its metadata have been prepared and associated with each other to form an information package, they are saved to the archival storage system.

8.5.3 Archival storage

A preservation programme must provide archival storage that maintains, protects and verifies the integrity of the stored information packages, both the digital object and the metadata, whether stored as a single data stream or as separate but linked data streams.

To achieve this, the storage function must include practices that protect the data stream from unintended change, damage or loss: this will usually require regular copying of the data stream to fresh media, and when necessary copying to new media types. Storage practices must also include checking that the data stream has not been corrupted; system security; backup regimes that place copies at remote sites; and disaster recovery plans that address

contingencies such as complete loss of the system's operating infrastructure.

Obviously, this requires technical capabilities to provide a suitably secure and reliable storage service. Such a capability can be achieved with modest equipment so long as the equipment and the whole system are well managed. The more material there is, and the more diverse and complex it is, the more sophisticated the storage system needs to be.

8.5.4 Preservation planning

For most digital materials, preserving accessibility requires more than the data protection offered in the Archival Storage function. Only material being kept for very short periods could be stored without further attention to the means of providing access.

The purpose of the Preservation Planning function is to monitor threats to accessibility, and to specify action to pre-empt or respond to them.

The relevant threats mostly relate to changes in the technology that underlies access, so this function looks for such changes and takes action to maintain accessibility despite these changes. Frequently, the action will involve changing the information package: transforming the digital object itself to a different coding (as happens in migration), or changing the metadata that describes the means of access and links to current access tools.

8.5.5 Data management

Managing information packages in the archive generates its own data about what material is stored, what can be accessed, and about the management of the archive. This data must be managed to support use of the archive, and to support its effective administration.

8.5.6 Access

This function provides a user interface to the archive, allowing users to discover what is held, to request material and if appropriate to receive copies.

For many archives, access will be subject to restrictions for some or all potential users. The Access function may well require mechanisms to control access.

8.5.7 Liaison and advocacy

The preservation programme must find ways to advocate good practices among producers, aimed at facilitating preservation of the material for which the programme will be responsible. There is also a need to understand who the likely users of the material will be, so that preservation and access arrangements can be tailored to their needs and expectations.

8.5.8 Management, administration and support functions

The overall operation of the programme must be managed. In part this responsibility involves the development of policy frameworks and standards covering all areas of operations; in part it involves the ongoing supply of appropriate resources and infrastructure including suitable technical systems; and in part management processes such as monitoring and reporting on the programme's operations.

[The responsibilities and functions set out above (in simplified and slightly modified form), are described in much greater depth and detail in the Reference Model for Open Archival Information Systems (OAIS), released in 2002 as a draft international standard by the International Standards Organisation. The OAIS Reference Model is the most successful attempt to define both a conceptual model for managing digital materials of enduring value, and a vocabulary with which to discuss it.

Anyone contemplating a responsibility for managing digital materials should seek to understand the concepts articulated in the Reference Model itself.

The Reference Model is a high level conceptual framework that can be used as a reference point for those designing, using and evaluating real implementations. It is important to realise that it is not an implementation specification: it does not provide a set of instructions on how to preserve digital information. Its value lies in explaining what is required at a highly conceptual level, regardless of the means chosen to achieve it.]

8.6 Characteristics of reliable preservation programmes

The reliability and trustworthiness of digital preservation programmes are very important issues to many stakeholders. Producers, users, investors and the broad community have a strong interest in ensuring that digital heritage materials are managed by arrangements that can be trusted. Those potentially responsible for the programmes also have an interest in assessing what they can offer and the risks of accepting responsibility.

Preservation programmes offering long-term reliability are expected to have the following characteristics:

- Responsibility: a fundamental commitment to preservation of the digital materials in question
- Organisational viability, including the prospect of an ongoing mandate; a legal status as an organisation that would support an ongoing preservation role; and a demonstrated ability to put together the resources, infrastructure and work teams that could manage the complexity of digital preservation
- Financial sustainability: a likely prospect of the organisation being able to continue to provide the required resources well into the future, with a sustainable business model to support its digital preservation mandate
- Technological and procedural suitability: the use of appropriate systems and procedures to do what is required to manage and preserve digital resources
- System security of a very high order
- Procedural accountability, with clear allocation of responsibilities and mechanisms for reporting and assessing performance.

Arrangements that are able to demonstrate these attributes should be trustworthy. Development of trust may be a matter of demonstrating these characteristics over time. In the long term, certification programmes will probably be needed but at the time of writing no certification programmes for digital preservation arrangements have appeared. It remains very

much the responsibility of those proposing digital preservation programmes to show why their arrangements should be trusted, and very much the responsibility of other stakeholders to determine that any arrangements on offer can provide an acceptable level of reliability.

REFERENCES – **where to look for more information**

Cross references

Ingest also see Taking control: chapter 14

Archival storage also see Protecting data: chapter 16

Preservation planning also see Maintaining accessibility: chapter 17

Comprehensive and reliable preservation programmes are contrasted with other possibilities in Taking responsibility: chapter 9

Offsite references

- Consultative Committee for Space Data Systems (2002). *Reference Model for an Open Archival Information System (OAIS)*. CCSDS 650.0-B-1. Blue Book. Issue 1. January 2002. Washington D.C., CCSDS Secretariat, 2002. http://wwwclassic.ccsds.org/documents/pdf/CCSDS-650.0-B-1.pdf
- Research Libraries Group (RLG), (2002), *Trusted Digital Repositories: Attributes and Responsibilities - An RLG/OCLC Report*. Mountain View, California, 2002. http://www.rlg.org/longterm/repositories.pdf

Chapter 9. Accepting responsibility

INTRODUCTION

9.1 Aims

The purpose of this chapter is to help programme managers decide on the preservation responsibilities they will accept.

9.2 In a nutshell

Digital preservation will only happen if organisations and individuals accept responsibility for it. Accepting responsibility includes putting arrangements in place to take the kind of preservation steps outlined in these Guidelines, and appropriate emerging processes as they become available. While comprehensive and fully reliable preservation arrangements are necessary, in many instances they may be currently unobtainable, and more limited, responsible programmes may offer valuable contributions.

MANAGEMENT PERSPECTIVE

9.3 Deciding on a preservation responsibility

Responsibility is a crucial issue in the preservation of digital heritage. The starting point for action is a decision about responsibility.

Because the cost implications for an organisation can be significant, and because the requirements may be complex and uncertain, it is no small thing to accept a responsible role. The responsibilities and functions of comprehensive programmes, and the characteristics of reliable programmes, as described in the previous chapter, are not undertaken lightly. They imply investment of resources, energy and vision.

On the other hand, the problem is pressing: large parts of the digital heritage will be lost within a short time unless organisations and individuals agree to take action.

The approach suggested by these Guidelines is for interested organisations to break their responsibility decisions into two sets of considerations:

- Whether there is a basis for accepting responsibility
- If so, what kind of responsibility should be accepted?

In all cases, the quality of the decisions will be influenced by knowledge and insight about the materials being considered, the tasks that will be required, the expectations of stakeholders, and the resources that may be available.

9.3.1 Consideration one: Is there a basis for accepting a preservation responsibility?

Table 9-1 presents a series of questions that may help as a starting point in considering this decision.

Key decision: Whether there is a basis for accepting a preservation responsibility		
Key questions:	Contributing questions:	Such as:
1. Does the business of the organisation imply an existing or potential preservation obligation for any kinds of digital heritage materials? (Is the organisation required to take responsibility?)	Are there any existing legal requirements?	– legal deposit or other statutory obligation – organisational rules – contractual obligations
	Do current obligations imply a possible extension to digital materials? (eg A deposit library lacking legislation for digital materials)	– responsibility for parallel material – serving parallel client needs
	Has the organisation accepted a custodial role for digital materials with an implied preservation expectation?	– donated data – data stored for depositors – data transferred from another preservation Programme
2. Does the organisation have an interest in accepting a preservation responsibility? (Does it want to have a role?)	Does it have a 'natural interest' in identifying materials and keeping them accessible?	– for users – for future research – for re-use – for community pride – for profit
	Does it have an indirect interest based on a valued relationship with a particular community?	– the producer community – other stakeholders
3. Does the organisation have, or could it acquire, the capacity to take on a preservation responsibility?	Does it have what will be required to fulfil a responsibility?	– commitment and vision – resources – knowledge and skills – contacts – credibility
4. Is this really someone else's responsibility?	Is there someone else with this responsibility already, or who could take it on?	Someone – who already does it – who is already required to do it – with a natural interest in doing it – with the capacity to do it

Table 9-1 *Considerations in deciding if a preservation responsibility exists*

9.3.2 Consideration two: What kind of responsibility should be accepted?

These Guidelines strongly recommend that preservation programmes should seek to comply with the criteria for comprehensiveness and reliability described in chapter 8. These are important benchmarks for all programmes.

However, many organisations may decide that they have a preservation responsibility but find they are unable to comply with these criteria, and ask whether there is a place for more modest preservation programmes. Are there no alternatives other than action for those who can comply and inaction for those who cannot?

In many environments, there may be no one who can offer a full and reliable preservation responsibility. The only chance of survival for some digital heritage may depend on someone taking limited, unreliable but informed action while they can. This may buy at least enough time for more reliable arrangements to be put in place.

Even where comprehensive and reliable programmes are available, there may still be an important role for programmes that can take responsibility for some processes although they are unable to take responsibility for all processes. In fact, most large preservation programmes may only be sustainable if they can find partners who are willing and able to contribute a limited but complementary role.

As well as the degree of comprehensiveness and reliability they can offer, preservation programmes can also be distinguished by the range of materials they seek to preserve, and by the length of time their responsibility extends. There is definitely a role for programmes that can offer comprehensive and reliable preservation for quite restricted ranges of material, over quite limited periods of time.

This is most important, as the often-repeated claim that digital preservation involves very long-term commitment may well act as a barrier to preservation by discouraging agencies that are well placed to take short-term action when it is needed. With good succession planning, agencies that are able to play an effective but time-limited role can assist the smaller number of agencies that are able to commit to really long-term, sustained custody.

Table 9-2 suggests a range of levels of responsibility against four key continuums: the scope of material preserved; the time frame for which responsibility is accepted; the extent of core functions and responsibilities undertaken; and the presence of reliability characteristics.

Communities concerned with digital preservation may need to develop their own responsibility levels and criteria for programmes entrusted with their heritage.

Organisations accepting a preservation responsibility, and their stakeholders, may find it helpful to chart their level of accepted responsibility on such continuums.

Any proposals to take preservation responsibility should be informed and well considered, based on a clear view of what must be achieved – even if it is not completely clear how all obstacles will be overcome or all challenges met.

1. Scope of material	**Restricted Programme**		**Selective Programme**	**Broad Programme**
	very restricted		wide range, comprehensively collected	
2. Scope of time	**Initial Programme**	**Caretaker Programme**		**Long-term Programme**
	only until technology changes	only until use ceases	for a limited number of years	"forever"
3. Scope of functions and responsibilities	**Partial, non-comprehensive Programme**			**Comprehensive Programme**
	restricted functions		comprehensive functions	
4. Level of reliability	**Non-reliable Programme**			**Fully reliable Programme**
	limited characteristics of reliability		all characteristics of reliability	

Table 9-2 *Levels of responsibility – some possible continuums*

9.4 Planning for long-term preservation

All programmes, but especially those unable to offer reliable long-term commitments, must seek to put I place some kind of *fail-safe mechanism*. The purpose of such arrangements is to provide a good prospect of preservation continuing beyond their own involvement, should that become necessary.

Fail-safe arrangements extend along a continuum, from a commitment to find someone else to take over responsibility before discarding digital materials, through to legally mandated arrangements for one agency to take over the management of data held by another agency if it fails to meet its preservation responsibilities.

Responsibility for initiating and maintaining succession plans obviously lies with whomever is managing the material, but there may be other players who should accept some responsibility. Agencies that believe they will be asked to pick up the preservation task, and those that wish to see the materials remain accessible, also have an interest and may need to take an active role in initiating negotiations with the current custodians.

9.5 Some pragmatic responsibility principles

In facing the daunting challenges of preservation responsibility, it may help to consider some pragmatic principles:

- Someone has to take responsibility: if no one does, the chances of any particular materials surviving are very small
- Everyone does not have to do everything: responsibility can be shared. As these guidelines indicate, there are more than enough responsibilities for one preservation Programme. Many tasks, such as deciding what should be preserved, are best managed in partnership with others. If there is no one suitable to share responsibilities, preservation programmes should make realistic judgments about the responsibilities they can carry alone
- Everything does not have to be done at once: developing all the components of a large-scale, comprehensive preservation programme takes time. It is good to approach the task with a sense of urgency, but it may also be necessary to look for ways of buying time. This might require prioritising the issues that need to be addressed, or the material that needs attention. It may involve looking for easily managed materials ('low hanging fruit'). Some problems must be addressed without delay; some can be addressed in stages; and some can wait
- Responsibility does not have to be forever: there is definitely a place for time-limited contributions to an overall preservation programme, so long as the time limits are explicitly understood
- Limited responsibility should not mean causing harm: preservation programmes may need to take steps before all problems are solved and all techniques settled, but they must also try to minimise the harm of making later preservation efforts more difficult
- Someone must take a leading role: even when responsibility is shared, progress usually depends on at least one partner accepting the responsibility to lead.

9.6 Who might take responsibility

Who might take responsibility for establishing and managing preservation programmes for digital heritage materials? Possibilities include extension of the role of established 'memory' institutions such as libraries, archives and museums; or establishment of a new kind of institution focused solely on preserving digital materials; the extension of a preservation role to a range of other potential 'keepers' already involved in managing digital materials; or some combination of these.

9.6.1 The role of established heritage institutions

In early discussion of how digital materials would be preserved, it was frequently claimed that digital technology changed the picture entirely: existing institutions would find they had no role in managing digital materials.

It is still much too early to judge whether the predicted demise of traditional cultural and information institutions will be realised, but experience suggests the obituaries are premature. When one looks for agencies that might offer what is required, institutions that already manage non-digital heritage materials appear to have many advantages. Many of them offer:

- Expertise in recognising important heritage materials
- Experience in working with user communities
- Experience in working with rights owners
- Expertise and international networks dedicated to organising and describing heritage materials so they can be found and understood
- Commitment to their long term preservation
- At least some relevant expertise and infrastructure that might be brought to bear on digital asset management, and
- At least some prospect of an ongoing mandate from their communities to manage and preserve digital heritage.

To this promising foundation, some institutions have been able to add a leadership role in looking for practical ways of preserving digital heritage.

This does not necessarily mean that all institutions with a traditional heritage role should try to become digital heritage managers: in some cases the resources and expertise required are just not available; while in others their existing role is so important and so demanding that they should not sacrifice what is already in their care, for the sake of what may be much less important digital material.

It also does not mean that existing heritage institutions are the only organisations that need to manage digital heritage materials.

But it does suggest that existing heritage institutions are good crystallisation points around which digital heritage preservation programmes can grow. Such institutions should not disregard the existing strengths that they bring to the management of digital heritage materials, often in partnership with others who can bring a range of new skills and understandings.

Among existing heritage institutions, national libraries, national archival agencies and other lead institutions in various sectors may have a particularly important role in initiating preservation programmes. This has already emerged in many countries.

9.6.2 The roles of new kinds of digital preservation agencies

Some people believe that new institutions will be needed to take on the task of preserving digital heritage. Presumably these agencies would offer specialist expertise and facilities dedicated to digital materials, and possibly dedicated to preservation rather than a wider range of functions which existing collecting institutions perform, like arranging and interpreting materials and promoting their use.

Many data archives fit into such a field already, as they exist solely to manage and preserve digital materials. They often have the paradoxical advantages of being able to focus on a limited range of materials and management tasks, while being able to offer services to a wider range of data-producing communities.

9.6.3 The roles of other trusted keepers

Who else might take on the role of trusted keepers? Just as it is too early to judge the long-term role of libraries, archives and museums, it is too early to attempt a definitive list of others who may play an important role. However, some possibilities are already obvious:

- Universities and other institutes of research and learning have a natural interest in ensuring ongoing access to certain kinds of digital materials, and may have both the long-term viability and the technical infrastructure to play a preservation role
- Publishers and creators of digital content have a range of interests in ongoing management and accessibility. In many cases this extends beyond immediate commercial considerations to a longer-term investment in the exchange of ideas and the intellectual and cultural capital that ongoing accessibility encourages. Some publishers and creators may be willing and able to provide the infrastructure needed to maintain digital heritage materials in which they have an interest.

9.7 Declaring responsibility

When preservation programmes have decided what kind of responsibility they will take, it is very important that they declare their intentions. This makes it easier for others to work with them, reduces the likelihood of effort being unnecessarily duplicated, and provides a clearer picture of what material is likely to survive and what is not.

Explicit statements of responsibility must also be realistic: overly optimistic claims may suggest a level of preservation security that does not exist, and other programmes may not be able to step in at the last moment to save material they thought was someone else's responsibility.

REFERENCES **– where to look for more information**

Cross references

Comprehensive and reliable preservation programmes also see Understanding digital preservation programmes: chapter 8

Rights issues in accepting responsibility also see Managing rights: chapter 15

Chapter 10. Managing digital preservation programmes

INTRODUCTION

10.1 Aims

The purpose of this chapter is to suggest some key areas of management attention for preservation programmes. Experienced managers will find much of this discussion is already familiar.

10.2 In a nutshell

Preservation programmes require good management, often drawing on generic management skills such as tailoring programmes to the priorities and circumstances of the case, and making the right decisions at the right time. Digital preservation programmes present some particular management issues associated with their developing nature, the range of stakeholders, and the long-term impacts of current decisions.

MANAGEMENT PERSPECTIVE

10.3 The need to manage

Like all programmes, preservation programmes need to be managed in a coherent way. Digital preservation management should not be seen as a mysterious art: it draws on good general management skills, allied with enough subject knowledge and understanding of technical issues to see realistic possibilities and to make good decisions. The discussion that follows focuses on those issues specifically relevant to digital preservation.

10.4 What programme management must deal with

10.4.1 Decisions

Good management often comes down to knowing what decisions have to be made and making them at the right time. The important decisions do not have to be made all at once, but eventually they will be required on issues including:

- Whether to get involved in preserving digital materials at all
- The mission of the programme
- The scope of involvement: what kind of materials should be included, how big the programme should be, whether the programme will aim to be comprehensive and reliable or something more modest

- Where to obtain the services the programme will need
- What resources will be made available and how the programme will be sustained
- What organisational structures are required to support the programme
- Whom the programme will work with
- What issues should be given priority attention
- How accessibility will be maintained
- What succession arrangements should be put in place as a fail-safe mechanism.

10.4.2 Risks and risk management

Preservation programmes must seek to understand and respond to threats that would jeopardise ongoing accessibility and other aspects of the programme's mission. A risk management approach provides an appropriate basis for deciding what risks warrant attention, and for planning action that will lower the level of risk.

There are many suitable risk management models available. A reasonably simple but effective one is suggested in Table 10-1.

Risk analysis, even undertaken informally, helps in a number of ways:

- Recognising:
  - The most pressing threats (such as web publications disappearing; media failure of magnetic carriers; impending replacement of equipment or software; a change in government agency arrangements that will threaten record keeping systems)
  - Threats that may not require immediate action (such as the eventual obsolescence of a standardised, ubiquitous file format such as TIFF, the impact of which should be manageable when a replacement standard appears)
  - Threats over which the programme may have no influence (such as the business imperatives of producers)
  - Threats that are so pressing but so intractable that the programme may decide to withdraw from responsibility (such as a refusal by rights owners to allow any form of access or preservation copying at any future time).
- Deciding:
  - Where to allocate resources
  - What steps to take as a priority
  - When action may be needed
  - What supporting action is needed to address priority risks.
- Planning ahead
- Justifying decisions.

Risk assessment is especially helpful if it is extended beyond the immediately apparent risks to include:

- The risks associated with the action that is proposed to deal with the threat. For example, the programme may not have the required skills, resources, permissions, etc. This in turn might lead to action that is a prerequisite for dealing with the priority

threat

- The causes of the original threat. For example, web publishers may not be aware of steps they could take; producers may be using standards incorrectly; or producers may be at risk of business failure. This analysis might lead to action that addresses the causes, such as education campaigns, the development of standards in conjunction with producers, or development of indicators of impending business failure including signs that web sits are not being maintained or personal knowledge that projects are coming to a close.

Steps	**Worked example**
1. *Asset identification*: identify what needs to be protected, as specifically as possible	Online publication stored on a web site managed by someone else
2. *Threat identification*: identify the threats that appear to pose a risk to the programme's objectives	Access to a particular version of the online publication will be lost because the owner overwrites old versions with new versions
3. *Probability assessment*: estimate the likelihood of each threat happening	Very likely to happen, based on previous history of the site
4. *Consequence assessment*: estimate the likely impact if the threat did eventuate	Likely to result in complete loss of the old version, as the owner does not appear to maintain an archive of overwritten versions
5. *Risk level assessment*: Calculate the level of risk by combining the probability and consequence	High risk – likely to happen and would result in complete loss
6. *Mitigation*: propose action that could reduce the likelihood or the impact of the threat, or both	Options – contact site owner and suggest owner makes archived copies; negotiate permission to take copies now; or to take copies before versions are overwritten
7. *Risk threshold*: decide whether the level of risk is acceptable with or without mitigation action	Material is considered important so level of risk warrants taking mitigation measures
8. *Allocation of ownership*: determine who is responsible for taking action, and any constraints	The owner could be responsible but might not be willing to take action; programme could take responsibility but may need permission
9. *Priority setting*: compare risk levels for identified threats and decide what risks should be given priority	High priority compared with other risks
10. *Reality checking*: decide whether the risk and priority assessments tally reasonably well with expectations	No lingering doubts - the analysis 'makes sense'
11. *Action triggers*: decide whether action is needed immediately; if not, identify some signs that will indicate when action is required	Owner approves copying by the preservation programme immediately before each version is overwritten, but cannot guarantee a schedule. The programme decides to contact the owner regularly for information on planned updates, and will assess whether this proves to be an adequate indicator.

***Table 10-1** A simple risk management model*

10.4.3 Stakeholder relationships

A broad range of stakeholders have an interest in digital heritage materials and how they are managed. They may include those creating or distributing the materials, those who need to use them now or in the future, and those who cite the materials, assuming that they will remain available. Some stakeholders may be hidden but play a very influential role behind the scenes, such as hardware and software manufacturers, providers of funds, and bodies authorised to control access to certain materials. Some stakeholders may be of direct interest to the preservation programme itself, such as potential collaborative partners, standards bodies, and researchers developing new digital preservation methods.

Preservation programmes must recognise those who have an interest in their objectives or who could exert an influence on them. Risk assessment is likely to indicate which stakeholders will be most important, as well as suggesting the kind of relationship the programme needs to develop with them. (Some relevant issues are discussed in later chapters regarding producers, rights owners and other preservation programmes.)

Because the windows of opportunity for preservation are much narrower than for non-digital heritage, preservation programmes may have to be much more active in pursuing relationships with producers of heritage materials. They may also have to look for ways of influencing the 'hidden' stakeholders whose decisions may have a critical impact.

10.4.4 Sustainability and business models

Reliable preservation programmes must be sustained over long periods, so they require business models that guarantee adequate resources will continue to be available. Unfortunately, such guarantees are rare in the real world. Most programmes have to survive with less certainty.

Especially in the initial stages, sustainability presents a dilemma: programmes must eventually have it, but they often cannot tell what resources will be needed to fulfil their responsibilities, nor what resources may become available once a successful programme has been established. Programme managers have the challenge of finding long-term business models, but they may also have to find short-term funding arrangements, and recognise the appropriate time to switch from one to the other. (In fact, some managers may find they can only build a long-term programme on the back of a series of short-term funding arrangements.)

Some of the business models commonly considered for preservation programmes include:

- General community funding, often through taxes or the allocation of grants from special sources such as lottery funds
- Funding by a specific community with an interest in the programme, such as a local community project, or an industry peak body levying members
- Central funding by a parent organisation as a normal cost of business, such as a university library or a business archive
- Payments by users of the material
- Payments by producers who deposit material with the programme

- Sponsorship or philanthropic assistance
- Cross subsidisation by other activities within an organisation
- A combination of approaches.

Each of these models may be more suited or feasible for particular kinds of programmes. There may also be other models that can be established to provide sustainability.

10.5 What programme managers need to do their job

10.5.1 Information

Standards and practices in digital preservation are evolving rapidly (though perhaps not as rapidly as they are needed), and programme managers must find ways to keep their knowledge up to date.

There are a number of international initiatives dedicated to just such a purpose, including the journals, websites and subject gateways included in the Reading List. Despite some overlap, these initiatives complement each other; they offer excellent ways of keeping abreast of developments.

There also appears to be an opportunity to explore solidarity relationships between programmes with different levels of experience and expertise. Most existing programmes are willing to share information and ideas but often find themselves overwhelmed with requests. Formal information sharing arrangements may ease that burden while offering real benefits to those developing new programmes.

10.5.2 Corporate support

Digital preservation programmes often start as experiments and projects with a speculative mission and an uncertain future. They can be easily dismissed as important but bothersome add-ons to already over-stretched organisations. It often takes some time for a definite set of workflows to emerge and for the programme to take shape.

Fledgling programmes particularly need in-house mentors or sponsors to champion their cause and speak for them in corporate forums. They also need ways of connecting with others in the organisation, such as management committees, that can keep the programme aligned with corporate directions while also providing feedback on progress and problems. This corporate support must take account of the fact that preservation programmes are likely to be resource intensive and to involve complex technical and organisational issues.

10.5.3 Resource costs

Availability of resources is always a critical constraint. It is important to tailor the ambitions of the programme to a reasonably realistic idea of what can be achieved: it may not be possible to do everything!

It is difficult to estimate the long-term costs of digital preservation. While unsatisfactory from a planning point of view, it may be necessary to estimate expected costs over a short- to

medium-term period, such as five years, and to use the knowledge gained in that timeframe to inform more accurate estimates of long-term costs. However, there is always likely to be a problem in estimating the costs of long-term actions that are still unclear, such as migration costs.

Some costs do become much easier to predict following a few years' experience in collecting material, preparing it for storage, and protecting it.

Some cost assumptions can be reasonably made:

- Development costs are likely to be high, depending on how ambitious the programme is from the start. Systems design is a necessary investment in the long run but it can be a significant set up cost
- There are obvious recurrent costs associated with staff, accommodation, energy supplies, network use, telecommunications costs, storage media such as disks and tapes, and consumables. Although often funded as capital expenditure, equipment and software should be seen as recurrent costs because they will have to be replaced on a regular basis
- The staff costs of working with producers can be high because of the need to address new issues with each change in technology. The costs of negotiating rights may vary depending on the complexity of rights ownership and on whether or not rights need to be purchased
- The costs of identifying and selecting materials for preservation are likely to be low per unit, but there may be many units. A non-selective approach may reduce costs, although adding to long-term preservation costs overall, as more material must be stored, processed, preserved and organised for access. Human judgments about selection are expensive where labour costs are high; automating decisions may reduce costs if the high set-up costs can be spread over a large amount of material, and if it is feasible to automate what are often complex human judgments
- The costs of collecting or transferring materials are likely to be low per unit but large programmes may generate significant transfer costs overall. This may incorporate high transmission costs where automated gathering searches and downloads large amounts of data. The cost of quality control checking is likely to be high unless it can be automated
- Converting material to a restricted range of standard formats may be inexpensive if the conversion is easy, but expensive if individual handcrafting or correction is needed. However, there may be considerable long-term cost benefits in being able to deal with a restricted range of formats
- The costs of describing material and adding metadata are likely to be high because of the amount of information to be recorded, and the difficulty of finding it in some cases. Costs could be greatly reduced by producers using more standardised structures and creating good metadata and documentation themselves. For heavily standardised formats such as those widely used for archival versions of images and audio, costs will be reduced by automated gathering of metadata from files and during production processes
- Costs associated with storage are theoretically low and decreasing, but in total they

reflect the amount of data to be stored, which may be large. Estimates of storage and processing costs should take account of backups and multiple versions of material. The cost of regularly copying data to new media is generally discounted by regular increases in storage densities

- The costs of providing means of access – such as analysing data structures, writing new code for migration, writing or finding emulators, quality checking, and so on – may be high or low, depending on how difficult the process turns out to be. For example, if a conversion software tool can be developed that works automatically for millions of similar files, the unit costs may be extremely low, although the total cost of computer time could still be significant. On the other hand, strategies that must be managed item by item will be very expensive
- Computer costs to store and serve large numbers of access copies, or to generate them on request, should not be ignored. These would generally not be counted as preservation costs, but preservation programmes may still have to bear them as part of their business function.

10.5.4 Resource issues

10.5.4.1 Staffing

Staff numbers required will depend on the size and type of programme. Where the material is very similar and well organised, and where the preservation tasks are straightforward, it may be possible to automate most processes and reduce staffing requirements to a minimum. On the other hand, where repeated human intervention is needed, the programme will need staff who are adequately equipped to provide it. It is very hard to automate processes that require subjective judgments, although not impossible. Even highly automated programmes, however, will eventually need staff to deal with changes in the operating environment and with the need to replace systems.

Finding staff with the right skills is often a challenge. There are few training programmes for 'digital conservators' or even for digital collection managers. However, digital preservation draws on a range of existing skill areas: information technology, preservation management, collection management, and information curatorship. It may well be possible to identify staff with skills in one or more of these areas with a capability and interest in extending their skills. It may also be possible to complement in-house skills with those available from suitable service providers.

Programme managers need to have:

- Good problem solving skills, and an ability to deal with complex problems for which there may be no current long-term answers
- A pro-active approach that considers short-term, medium-term and long-term issues
- Adequate awareness of the relevant technical, preservation, corporate, legal and political aspects
- An ability to think critically, but also to receive new ideas and to adapt to change
- Collaborative abilities and in interest in sharing information and looking for ways to work with others.

10.5.4.2 Equipment

Preservation programmes require sophisticated systems and tools, although small programmes dealing with largely un-automated processes and with carriers that can be stored on shelves may be able to get by using quite modest computer equipment that is already available. Where services have been contracted from an outside supplier, the programme itself may not need to provide much of the technological infrastructure.

The systems and tools likely to be required include those for:

- Storing and managing the collection material
- Storing and managing metadata
- Managing the transfer of material to the collection, possibly including gathering
- Managing information about rights and managing access in accord with rights
- Storing and managing accessibility tools such as original software, plug-ins, conversion and emulation programmes
- Searching by users
- Making appropriate copies available for users
- Managing many of the processes described in these guidelines, especially those that generate management information, copy material from one place or format to another, or require automated process checking.

Procurement of appropriate systems is a significant task, requiring special attention to specifications and the evaluation of options. Some existing programmes are willing to share information on the specifications they have used and their experience in procuring suitable systems.

10.6 Useful tools for preservation programmes

10.6.1 Standards

Standards lie behind almost everything that computers do, so they form a crucial foundation to the creation and use of digital materials. However, they have not yet been great enablers of preservation. They can be expected to make several important contributions if they are well chosen and well used. Those contributions can be seen in:

- The creation of digital materials that should be relatively straightforward to preserve. Some file formats (themselves standardised) have proven to be so widely useful that they have been adopted by creator communities as a best practice, raising the prospect that they will continue to be used for a long time. Examples include the Tagged Interchange File Format (TIFF) for images, and the Structured General Markup Language (SGML) family of formats including HTML and XML, for structured documents
- Where such widely adopted formats are based on open source, non-proprietary specifications (as is the case with TIFF and SGML), it should be relatively easy to find or develop tools when they are needed to provide ongoing access across changes in

technologies

- Agreed approaches to a number of preservation processes, including the recording of metadata, migration processes, data protection and item identification. Standards are still evolving in these areas
- Defining the responsibilities and functions of certain kinds of preservation programmes. The outstanding example of this to date is the Reference Model for an Open Archival Information System (OAIS), accepted by the International Standards Organisation in 2002.

Standards should not be seen as a preservation panacea. Even where they exist, they are subject to change, versioning and non-standard use by producers. Many format standards are in fact proprietary specifications that may not be publicly available, so it may be impossible to know whether tools for future access will be available.

While increased standardisation of processes can only help preservation programmes, it is important not to wait for a single 'digital preservation standard' to emerge before taking sensible preservation action.

10.6.2 Organisational structures

There are many different models that have been used as organisational structures to manage digital preservation. Some possibilities include:

- Setting up a single separate digital preservation unit to look after all aspects
- A series of specialist units looking after different aspects
- A matrix of people working in different areas, responsible to an overall programme manager
- Mainstreaming of the work through existing work areas so that it becomes part of the normal work integrated with other operations
- Embedding the programme in a particular existing work area such as IT, preservation, collection development, or collection management sections.

These different models tend to produce different emphases, reflecting various levels of comfort with IT tools, collecting objectives, preservation thinking, etc. Any model can be made to work effectively, so long as it draws on the perspectives and skills that are needed and has strong management support.

10.6.3 Preservation policy and planning

Preservation programmes should be guided by a policy framework that says what the programme is trying to do and how it will try to achieve it. In a field of such complexity and evolving understandings, a policy document needs to provide clear, long-term direction as well as regularly reviewed guidance.

In implementing policy and developing action plans, it is almost always necessary to decide what issues, actions and materials should be given priority for attention, and to understand what work is critical before other work can be attempted.

Some commonly used questions that help in setting priorities include:

- What is most important to support or fulfil the responsibility accepted by the programme (including any legal requirements)?
- What is most at risk?
- What material is most likely to be in demand but likely to become unusable?
- What risks will be easiest to do something about?
- What action would make life easier if attended to now, and what will make life harder if not attended to?

10.6.4 Service providers

There may be alternative paths for achieving the preservation commitments of the programme. Organisations should consider whether they will get a better result from doing the work in-house, or by contracting with someone else to do it, or by a combination of approaches. Many processes such as storage and documentation may be amenable to contracting, which may offer access to specialised expertise and facilities that would not be available otherwise. In many cases, programmes may not be able to afford the initial investments involved in setting up and maintaining infrastructure, so paying for someone else to provide services may be attractive.

Contracting may present some potential risks, including:

- Creating a distance between programme objectives and service provision
- The possibility of being locked into the service provider's services because the cost of seeking alternatives is greater than the cost of continuing
- Less control over what happens on a day-to-day basis
- Higher costs (they may also be lower) than in-house arrangements over the long term
- Fewer learning opportunities that might suggest better ways of doing things.

The success of contracting may depend on the programme manager's ability to define what is needed; finding a reliable supplier who can offer the services required at a suitable price; being able to negotiate a suitable contract with adequate safeguards; and skills in managing the contract.

In working with service providers, preservation managers must ensure they:

- Understand their business and what they need to achieve
- Communicate their needs
- Rigorously assess the capabilities, motivations, and understandings of potential suppliers
- Prepare and negotiate an appropriate contract
- Ensure communication channels are open so that any problems that be reported early
- Monitor performance closely and evaluate arrangements regularly

- Ensure there are responsible exit strategies and succession plans in place for when arrangements come to an end.

REFERENCES – **where to look for more information**

Cross references

Scoping of programmes also see Accepting responsibility: chapter 9

Keeping material accessible also see Maintaining accessibility: chapter 17

Equipment also see Protecting data: chapter 16

Standards also see Working with producers: chapter 13, and Maintaining accessibility: chapter 17

Offsite references (all online documents viewed March 2003)

- Committee on International Cooperation, University Archivists Group (UAG), (2001). *Standards for an Electronic Records Policy*. http://www-personal.umich.edu/~deromedi/CIC/cic4.htm
- International Council for Scientific and Technical Information (ICSTI), (2002). *Digital Archiving: Bringing Issues and Stakeholders Together – An Interactive Workshop Sponsored by ICSTI and ICSU Press, 30-31 January 2002, UNESCO House, Paris*. http://www.icsti.org/2000workshop/index.html
- Jones M, Beagrie N, (2001). *Preservation Management of Digital Materials: A Handbook*. The British Library, London.
- Lawrence GW, Kehoe WR, Rieger OY, Walters WH, Kenney AR, (2000). *Risk Management of Digital Information: A File Format Investigation*. Council on Library and Information Resources, Washington, D.C. http://www.clir.org/pubs/reports/pub93/contents.html
- National Library Australia, (2002). *A Digital Preservation Policy for the National Library of Australia*. http://www.nla.gov.au/policy/digpres.html
- Price L, Smith A, (2000). *Managing Cultural Assets from a Business Perspective.* Council on Library and Information Resources. http://www.clir.org/pubs/reports/pub90/contents.html

Chapter 11. Working together

INTRODUCTION

11.1 Aims

The purpose of this chapter is to encourage programme managers to consider collaboration as a means to achieving their preservation objectives, and to provide some basic information on options that may help in deciding on the most suitable models to pursue.

11.2 In a nutshell

There are good technological, economic and political reasons for preservation programmes to cooperate. Decisions about collaboration should be based on assessment of the benefits expected and the costs involved. There are a number of possibilities regarding where to look for partners, what the focus of the relationship should be, and structural frameworks that would be suitable. Successful cooperation is usually the result of careful attention to these choices, and to putting in the effort required to manage the collaboration in practice.

MANAGEMENT PERSPECTIVE

11.3 The need to collaborate

Digital heritage and collaboration go together very well. The technology of digital materials supports collaboration: such materials are easy to duplicate, and many are designed for networked access, so remote management is not difficult.

It is also expensive – often too expensive – to set up the whole infrastructure of digital preservation for every preservation programme, so there is a strong incentive to look for ways of sharing facilities.

There may even be a political imperative to work cooperatively: the community may reasonably expect that programmes will collaborate to ensure as much digital heritage as possible can be preserved, as coherently as possible.

11.4 Potential benefits of collaboration

Collaboration costs. It takes time and energy to negotiate agreements, to work with remote partners, and to maintain momentum. Organisational priorities can be sidetracked by problems in the collaborative relationship itself, taking attention away from the real mission of preserving digital materials. In the face of such potential costs and difficulties, it is important to identify the benefits that any collaboration is meant to deliver.

Benefits driving a cooperative effort could include:

- Access to a wider range of expertise
- Shared development costs
- Access to tools and systems that might otherwise be unavailable
- Shared learning opportunities
- Increased coverage of preserved materials
- Better planning to reduce wasted effort
- Encouragement for other influential stakeholders to take preservation seriously
- Shared influence on agreements with producers
- Shared influence on research and development of standards and practices
- Attraction of resources and other support for well-coordinated programmes at a regional, national or sectoral level.

11.5 Ways to achieve the benefits of working together

The benefits of collaboration usually do not happen by accident, but result from careful attention to choices. Programmes need to consider their potential partners, ways of working together, structural models, and the steps involved in making the proposed collaboration work. (Ultimately, they may also have to consider what opportunities are on offer.)

11.5.1 Partners

Potential partners are likely to be others working in the same sector. Examples could include a consortium of university libraries, or networks of data archives, or government agencies agreeing to use the same application software, or a group of recording studios sharing storage facilities for masters. It may be possible to join an existing collaboration, or to form new partnerships.

However, there may also be benefits to be gained from looking beyond sectoral boundaries, especially as digital technologies and users expectations are increasingly blurring the edges between sectors. For example, a number of libraries, archives, research institutes, data archives and producers in a regional area may consider joining forces to develop a local programme that meets all their needs.

There may also be opportunities for formal cooperative arrangements between preservation programmes and interested stakeholders including groups of producers, industry peak bodies, user groups, IT industry groups, or government bodies interested in fostering good practice.

11.5.2 Ways of working together

There are many ways that preservation programmes can work together, depending on the benefits they want to achieve and what each partner has to offer. They include:

- *Shared standards*: agreements to do things in the same ways, either with a view to interoperability between programmes, or simply based on a shared understanding of

what practices will best support preservation objectives

- *Information sharing*: agreements to share information either at a general level or on specific issues, such as procurement specifications or research results
- *Speaking with a common voice*: agreements to develop and present a common message in advocacy campaigns, or in publicity aimed at raising the profile of digital heritage preservation
- *Division of labour*: agreements to work together at an operational level, taking preservation steps in a coordinated way, with responsibilities either carried out in parallel or divided between programmes
- *Shared resources*: agreements to share resources such as systems, staff or funds to work on a common programme.

11.5.3 Structural models

Most collaborations can be seen to fit into one of four categories of structural models, each offering different strengths and weaknesses:

- *Centralised distributed models*, consisting of one partner that leads on policy, sets directions and provides most of the infrastructure, working with a number of others who have clearly specified but limited roles, such as identifying material to be preserved and adding metadata, possibly with limited responsibility for long-term maintenance. (For example: a central records authority working with government business agencies, setting standards and providing guidance.)

  Like all distributed models, this offers some cost sharing and creates a pool of ideas and perspectives. It allows economies of scale if functions like storage are centralised. It may offer more reliable preservation because processes can be better controlled and more specialised expertise used than in some other models. Decision making, largely in the hands of the central agency, may be more efficient than in more equally distributed models.

  On the other hand, this model may not encourage ownership of the programme among the peripheral partners, so it may not be effective in encouraging transfer of skills from the central agency.

  Such a model is probably good for beginning programmes seeking to collaborate with large, advanced programmes. It is also suitable where there may be one programme willing to take ongoing responsibility and a number of others who can help but are not sure about their long-term commitment.

- *More equally distributed models*, consisting of a number of partners with similar levels of commitment and responsibility. (For example: a group of data archives that decide to agree on standards and share specifications for purchasing computer equipment.)

  This model also offers cost sharing and the input of ideas, but it may have the

advantage of encouraging shared levels of ownership, without one partner having to bear the pressure of making decisions alone. On the other hand, it may be difficult to establish effective leadership, and consultation and decision making may be time consuming. Economies of scale may be lost if large centralised systems are replaced by a number of small parallel systems.

Such a model is probably suitable where there are a number of players willing to share responsibility but none wanting to lead a programme.

- *Very highly distributed collaborations*, consisting of a large number of partners, each playing a very restricted role, perhaps limited to self-archiving. (For example: networks of local community projects that decide that they will all keep their material for posterity.)

  Such a model may be a useful starting point for a preservation programme, raising awareness and allowing some steps to be taken. However, it is unlikely to offer much reliability without a large investment in specifications, training and checking. This can lead to high costs overall, although the model is attractive because of the low costs for each partner. Such models may have trouble addressing long-term preservation issues in a coordinated way.

  Such a model may be indicated where there are a number of small sites capable of taking some limited responsibility, especially if there is one partner able to play a coordinating role. It may also work for material for which preservation is desirable rather than essential.

- *Standalone arrangements* may contribute to later collaboration by allowing programmes to develop expertise, strategies and systems before looking for suitable partners. programmes operating in an environment where there are no suitable potential partners can make good progress on their own, and look for collaborative opportunities as they arise.

  (For example: a small research facility operating in a new discipline in an isolated location may decide that its data must be preserved and set up a modest programme to document, back up and migrate its data, hoping to eventually find a national or international programme that will take responsibility for it.)

11.5.4 Setting up a collaboration

Experience suggests that organizations often work together successfully when they:

- Understand what they want to achieve collaboratively
- Choose appropriate partners who can contribute
- Share interests and commitment, established through discussions and demonstrated in action
- Allocate enough resources to meet commitments: it is difficult to sustain cooperation

in an environment of frustration and failure

- Communicate often and effectively, both at an appropriate operational level and through some kind of management board for the joint programme
- Set realistic targets and regularly evaluate the arrangements.

REFERENCES – **where to look for more information**

Cross references

Costs also see Resource costs: chapter 10

Standards also see Useful tools for managers: chapter 10

Offsite references – case studies (all links viewed March 2003)

There are few available analyses of collaboration in digital preservation (though many on collaboration in other aspects of managing digital information). The following list therefore focuses on a spread of projects and programmes that may be worth studying.

Many well-known digital preservation collaborations have been basically research projects and feasibility studies, without an ongoing remit to manage material. Well-documented ones include:

- CEDARS, a collaboration of three UK university libraries, developed a Distributed Digital Archiving Prototype System of particular relevance, along with important reports. http://www.leeds.ac.uk/cedars/
- CAMiLEON, a collaborative research project at the Universities of Michigan (US) and Leeds (UK), examining methods of maintaining accessibility. http://www.si.umich.edu/CAMILEON/
- NEDLIB, a European collaboration of nine national libraries, a national archive and three large publishers, which produced a number of tools relevant to distributed programmes. http://www.kb.nl/coop/nedlib/

A few of the many active preservation programmes built on various collaborative models include:

- The Austrian On-Line Archive (AOLA), a joint initiative of the Austrian National Library and the Technical University of Vienna's Department of Software Technology. AOLA is an archive of snapshots of Austrian web space. http://www.ifs.tuwien.ac.at/~aola/
- Academic Research in the Netherlands Online (ARNO) which links the document servers of the University of Amsterdam, Tilburg University and the University of Twente, to make and keep their academic output electronically accessible. http://www.uba.uva.nl/en/projects/arno/
- Australian Digital Theses Project which aims to establish a distributed and maintained database of digital versions of theses produced by postgraduate research students at the participating institutions. http://adt.caul.edu.au/
- The China Digital Library Project which plans to establish a digital data storage centre coordinated by the National Library of China
- 'Purge Alert', an international initiative of the Committee on Earth Observation Satellites (CEOS), to encourage members of the global spatial data community to transfer responsibility for still-valued datasets before they are deleted by their original custodians.

http://edc.usgs.gov/archive/ceos/data_purge_alert.html

- Digital Image Archive of Medieval Music (DIAMM), a collaborative project of the University of Oxford, Royal Holloway, University of London, in consultation with the Arts and Humanities Data Service, established as a permanent electronic archive of European medieval polyphonic music. http://www.diamm.ac.uk/
- Digital Imaging Project of South Africa (DISA), a national collaborative of the major research institutions in South Africa, operating as a trusted digital repository on the OAIS model, and developing within a framework of formal agreements between the participants. http://disa.nu.ac.za/nu.ac.za
- European Visual Archive (EVA) focused on provision of easy and preserved access to the integrated collections and information held in European archives. http://www.eva-eu.org/
- JERRI: Ohio's Joint Electronic Records Repository Initiative, a joint project between the Ohio Historical Society, the State Library of Ohio, the Ohio Supercomputer Centre and the Ohio Department of Administrative Services to maintain public access to state electronic records and web-based publications of enduring historical value via an electronic archive. The project is partnered with OCLC's digital collection management and preservation project. http://www.ohiojunction.net/jerri/
- National Digital Archives programme (NDAP) launched in Taiwan in 2002 as a collaboration among nine national organizations including museums, libraries, archives, academic institutions and government.
- Norwegian Digital Radio Archive, a collaboration between the National Library of Norway and the Norwegian Broadcasting Corporation to build a common archive to handle and preserve large numbers of digital audio files.
- PANDORA, a programme initiated by the National Library of Australia in partnership with Australian State and Territory libraries and the national screen and sound archive, ScreenSound Australia, to capture, preserve and provide access to online Australian publications. http://pandora.nla.gov.au/
- The Victorian Electronic Records Strategy (VERS) Project undertaken by the Public Record Office Victoria (PROV) in conjunction with the Australian Commonwealth Scientific and Industrial Research Organisation (CSIRO) and Ernst & Young in 1998. The project produced a strategy to be used by Victorian government agencies for the long-term preservation of electronic records. http://www.prov.vic.gov.au/vers/

SECTION 3

TECHNICAL & PRACTICAL PERSPECTIVES

Chapter 12. Deciding what to keep

INTRODUCTION

12.1 Aims

From this chapter the reader should understand the key challenges in deciding what digital materials should be selected for preservation, and some guiding principles. The chapter also offers some technical and practical advice including suggestions on identifying the essential elements that must be preserved.

12.2 In a nutshell

It is usually necessary to decide what digital materials are worth keeping, as has been the case with non-digital materials. Many of the same approaches – selection based on criteria embodied in collection development policies, and good knowledge of the materials and their context – are fundamental for digital heritage selection. Preservation programmes also need to define the elements or characteristics of the materials they select that give them value, so that those elements can be maintained.

12.3 Terminology

S*election* has been used as a generic term in this chapter. It should be understood to encompass concepts like *appraisal* that have particular meaning in the records archive community.

KEY MANAGEMENT ISSUES

12.4 What should be kept

Deciding what should be preserved, by whom, and for how long, have been fundamental decisions in managing all kinds of tangible heritage. Such decisions are necessary because there are usually more things – more information, more records, more publications, more data – than we have the means to keep. Every choice to preserve is at the expense of something else.

12.5 Building on existing concepts

The selection of digital heritage is conceptually the same as selection of non-digital materials. Any existing programme with well-established procedures for assessing and selecting material for preservation will already have policies, skills and tools that can help in selecting digital materials, even though they may need some adjustment.

12.6 The challenge of digital materials

However, digital materials do present some new challenges that programme managers must take into account in making the best selection decisions they can.

- There is often a large amount of material to be assessed
- The means of producing and disseminating digital materials are widely available, so their quality is often inconsistent
- At the same time, there may be pressure to preserve the entire traffic carried by new communication channels such as the World Wide Web, regardless of quality
- Timing of selection is usually critical, as digital materials quickly de-select themselves by becoming unusable. It may not be possible to wait for evidence of enduring value to emerge before making selection decisions
- Some digital objects may be hard to pin down. New genres may not fit into existing classifications; some digital resources consist of linked or overlapping parts; many also exist in parallel versions. The selection process must find a way through these complications to clear, unambiguous decisions about what is to be preserved
- Even when external boundaries have been defined, it may be hard to tell which elements need to be maintained if the digital object is to fulfil its essential purpose
- It may even be difficult to tell where digital materials come from, making it hard to decide who is responsible for their preservation and with whom to negotiate the rights required by preservation programmes.

PRINCIPLES IN ADDRESSING THESE CHALLENGES

12.7 Informed, consistent and accountable decisions

Selection processes often have to deal with uncertainties and they involve judgments that are subjective and speculative; however, they should be informed, consistent and accountable:

- Decisions should be well informed about the material, its context, and the needs of stakeholders who will be affected
- Decisions should be consistently based on a selection policy reflecting the objectives of the organisation that accepts preservation responsibility. For collecting institutions such as libraries, museums and archives, an existing collection development policy may provide good direction
- For accountability, selection processes should be visible, based on publicly available policy documents, and produce clear and explicit statements about what has been selected and what has been excluded.

12.8 A basis for selection criteria

It is not possible to suggest specific criteria for selecting digital heritage materials because they are judged to be worth keeping for such diverse reasons. However, in principle:

- Decisions should be based primarily on the value of material in supporting the mission

of the organisation taking preservation responsibility

- This value must be weighed against the likely costs and difficulties of preservation, and the expected availability of resources. There is much to be said for starting with material that can be saved easily. However, the future costs and capabilities of digital preservation programmes are still unclear so it would probably be irresponsible to refuse valuable material just because it may appear difficult to preserve
- Where preservation programmes are unable to manage material they believe should be chosen for preservation, they should to indicate this in their selection policies
- It is desirable that the total effect of all collecting and preservation efforts will preserve at least a sample of all kinds of digital materials, including samples of the clearly ephemeral.

12.9 Recognising the elements that give material its value

Deciding to select an item, or a class of items, for preservation may not be enough:

- Preservation involves maintaining the elements and characteristics that give the material its value. The selection process should consider what those elements and characteristics are
- The process should document that reasons why the material was chosen so that preservation managers can understand what they are required to maintain. (Some more detailed notes are provided later in this chapter.)

12.10 A cautious approach

A decision not to preserve is usually a final one for digital materials. A cautious approach would be to decide what materials definitely must be preserved and for how long; what definitely does not need to be preserved; and what should be accepted for interim preservation action while a more definitive selection decision can be made.

TECHNICAL AND PRACTICAL ISSUES

12.11 Assumptions about value

There are dangers in assuming that current value assessments are a completely reliable guide to future values. For example, remote sensing data collected in previous decades has become unexpectedly important for assessing environmental change. This experience suggests it is probably better to err on the side of collecting more material rather than less, if the preservation programme can manage it.

12.12 Documentation

Where digital materials can only be understood by reference to a set of rules such as a record keeping system, database or data generation system, or other contextual information, selection processes must identify the documentation that will also need to be preserved.

12.13 Role of producers

Producers of digital materials may have a significant role to play in selecting what should be preserved. They are often well placed to understand why digital objects were brought into being, their essential 'message', and the relationships between objects and their context. If that information is not captured from the producer it may be too difficult to reconstruct it later.

12.14 Selective or comprehensive collecting

There may be a question of whether comprehensive or selective collecting is preferred. (This issue frequently arises in discussions about materials made available through the World Wide Web, for instance.)

Both comprehensive and selective approaches are supported by strong arguments. Advocates of a comprehensive approach argue that any information may turn out to have long-term value, and that the costs of detailed selection are greater than the costs of collecting and storing everything. Advocates of a more selective approach argue that it allows them to create collections of high value resources, with some assurance of technical quality and an opportunity to negotiate access rights with producers.

There may well be a place for both approaches, as they are likely to produce quite different collections of digital heritage that are valued for different purposes.

12.15 Collecting agreements

To minimise the risk of important materials being missed, and to avoid unnecessary duplication of effort, it may be necessary to seek agreements with other potential collecting and preservation agencies about respective roles and responsibilities.

12.16 Defining items

Selection policies may have to decide whether to select whole items and whole collections, or samples only. It is generally preferable to preserve whole items to retain their integrity, but it may be necessary to restrict collecting to representative samples as a way of at least evidencing the existence of some kinds of materials.

Selection policy may also have to consider whether the re-use of material constitutes a new item that should also be preserved.

12.17 Rights issues

Rights issues may influence selection decisions. Preservation programmes often select materials that are still subject to rights, but generally would not select material if rights were so restrictive that arrangements for giving access at some future stage cannot be negotiated. If the material can never be made available for use, or if necessary preservation steps cannot be taken, there is little point in selecting it as heritage material.

12.18 Recurrent selection

Should selection decisions be final? Reviewing selection decisions in line with specified retention periods is long-standing practice in the archival community. This approach may make sense for other kinds of digital materials as well as records, to check that the value of the material still warrants the expense of keeping it. On the other hand, the selection process itself is expensive and should be repeated as infrequently as possible. Even more importantly, any selection decisions that are subject to review should be explicit in order to avoid any inference of a permanent preservation responsibility.

12.19 Supporting the selection process

Selection requires the allocation of resources: people with knowledge, time, facilities and equipment to examine material. Managing selection also requires the development of criteria for appraisal. Where the amount of material is so large that it is not feasible to assess items individually, it may be necessary to establish classes of material that can be assessed on the basis of representative samples.

SPECIAL CONSIDERATIONS

12.20 Selecting the essential elements and characteristics that must be preserved

Preservation programmes often act as agents for other stakeholders: they take preservation action on behalf of someone who wants material kept for a reason. The 'someone else' may be as broad and many-faceted as 'the nation' or 'the general community', who may mandate a programme to collect and preserve a very broad range of materials; or it may be as narrow as the members of an organisation or researchers working in a particular discipline who want their own research output preserved for later use.

The needs of the 'community' – however defined – for whom the material is being kept will drive many decisions, from what material is selected, to the kind and level of documentation that is recorded, to the level of concern with authenticity, to the strategies that are used. For example, some programmes must offer users the option of interrogating old data to produce new results, whereas others have a brief to present material in a read-only form to that it cannot be changed or manipulated. Some programmes may even have to ensure users can run old simulations, play old computer games, or view digital art in ways that reproduce the original experience rather than a speeded up experience that later technologies may provide.

Defining the essential elements or characteristics (also referred to as *significant properties* by some programmes) is not conceptually difficult, as the examples above illustrate. In some circumstances – such as clearly defined and constrained user expectations, and easily characterised materials that are all similar – defining and encoding the essential elements should be straightforward. For instance, a programme may decide that users of a large collection of electronic mail messages only need to see elements that can be characterised as 'content information', such as the name and address of the sender, subject, date and time, recipients, and the message, in a standardised structure with only the most simple of formatting. A government archive with this approach could expect to apply this essential elements template to very large numbers of email records.

On the other hand, some materials are much more difficult to characterise, and expectations about how they will be re-presented for use, especially to an open-ended community of potential users, may be so hard to define in advance that it becomes almost impossible.

Approaches to this issue are developing as more people encounter the problems of describing, storing and planning to re-present digital objects in growing collections over long periods of time.

While more sophisticated methods of defining and describing essential elements are evolving, the following questions may provide some help in the selection process. (It will be seen that this is really part of the appraisal process that records managers go through in order to understand the records they are considering for selection.)

- For whom should this material be kept? Do they have specific expectations about what they will be able to do with the material when it is re-presented?
- Why are the materials worth keeping? What gives them the value that warrants the trouble of preserving them? Is that value associated with:
  - Evidence
  - Information
  - Artistic or aesthetic factors
  - Significant innovation
  - Historic or cultural association
  - What a user can make the material do, or do with the material
  - Culturally significant characteristics?
- Is the value tied to the way the material looks? (Would it be lost or significantly degraded if the material looked different?)
- Is the value tied to the way the object works? (Would it be lost if particular functions were removed? Or if particular functions happened at a different speed or required different keystrokes?)
- Is the value tied to the context of the material? (Would it be lost if links embedded in the material did not work? Or if a user could no longer see evidence that connected the material with its original context?)
- Is it possible to distinguish between elements within each of these areas? For example, would advertising banners be considered an essential part of the way the material looked? Would some navigation elements or display functions be needed but not others?
- If it is difficult to define what needs to be maintained, it may be easier to consider the impact of an element not being maintained, and to look for functions or elements that are definitely not needed.

FOR PRESERVATION PROGRAMMES WITH FEW RESOURCES

12.21 Selectivity

Preservation programmes with few resources must still make decisions about the materials for which they accept responsibility. Because the costs of preservation are related to the amount of material to be managed, such programmes may need to be highly selective, limiting their ambitions to a small amount of highly valued material.

Preservation costs are also related to the range of problems and formats that need to be managed, so it may also make sense to severely limit the kinds of materials selected to a very few formats.

12.22 Cooperation

Collecting agreements with other programmes may shift some of the burden of selecting materials. While these would normally be negotiated with other preservation programmes, there may be potential for agreements with producers that would lead them to make decisions about what should be collected and preserved. The preservation programme would still need to take some responsibility for quality control, as well as articulating the criteria by which material should be chosen.

12.23 Starting simple

Selection processes can evolve over time, starting with some simple decisions to select easily collectable and preservable materials ("low hanging fruit"), and aiming to move over time to more sophisticated decisions about a wider range of more difficult materials.

REFERENCES – **where to look for more information**

Cross references

Essential elements also see Understanding digital preservation: chapter 7, and Maintaining accessibility: chapter 17

Rights issues also see Managing rights: chapter 15

Offsite references (all links viewed March 2003)

- Cedars Project, (1999). *Why Can't We Preserve Everything? Selection Issues for the Preservation of Digital Materials*. http://www.leeds.ac.uk/cedars/documents/ABS01.htm
- InterPARES Project (2002). *The Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project.* http://www.interpares.org/book/index.htm
- Marcum Deanna B. (2001). 'Scholars as Partners in Digital Preservation', in *CLIR Issues, no. 20*, Council on Library and Information Resources. http://www.clir.org/pubs/issues/issues20.html - scholars
- National Library of Australia, (updated 2002). *Guidelines for the Selection of Online Australian Publications Intended for Preservation by the National Library of Australia*.

http://pandora.nla.gov.au/selectionguidelines.html

- Royal Statistical Society and UK Data Archive (2002). *Preserving and Sharing Statistical Information.* http://www.data-archive.ac.uk/home/PreservingSharing.pdf

Chapter 13. Working with producers

INTRODUCTION

13.1 Aims

This chapter aims to encourage programme managers to consider ways of working with the producers of digital heritage, and to provide some guidance on practices and standards that will make the preservation task easier.

13.2 In a nutshell

Digital heritage is often created without consideration of ongoing use and accessibility. However, there are definitely standards and practices that producers can use that either help or hinder preservation. Programme managers need to look for ways of exerting a positive influence from as early in the digital heritage life cycle as possible. This often requires a willingness to work with producers.

13.3 Terminology

Producers has been used in this chapter to refer to all those involved in design, authoring, creation and dissemination of digital materials before they enter a preservation programme. Digitisation programmes fit very squarely in the category of 'producers' whose digital output must be managed for ongoing accessibility by preservation programmes.

KEY MANAGEMENT ISSUES

13.4 The 'prehistory' of digital heritage

Digital materials are created by producers who are not necessarily concerned with long-term availability: creation of 'digital heritage' may not be part of their intention. Even those hoping to make something of enduring value may not have the knowledge or the means to do so, or be constrained by other impediments in their working environment.

Without some kind of intervention, it is unlikely that digital heritage materials will automatically be made in ways that will minimise costs and remove barriers to preservation. Many practices in fact make preservation much harder.

13.5 Difficulties in dealing with producers

In seeking to work with producers to overcome preservation barriers, programmes are like to encounter challenges:

- In many cases, the 'producer' is a layered concept, made up of a number of agents performing quite different functions, such as software developers, creators (often

multiple), editors, publishers and service providers

- Some producers may be diffident or even hostile to the idea that a third party is interested in somehow 'managing as digital heritage' the materials they have created.

PRINCIPLES IN ADDRESSING THESE CHALLENGES

13.6 The need to work with producers

Preservation efforts that wait until problems start to appear are likely to be more costly, more difficult, and less effective than efforts that start early.

Organisations that have both heritage-creating and heritage-preserving functions have learnt from experience that care invested from the start in the use of standards, documentation, good file management and other practices, pays dividends later in lower preservation and maintenance costs, as well as more easily accessed, re-used and managed collections.

While all preservation programmes do not have the same opportunities to influence production practices, all programmes should seek to influence the way materials are created, and managed, from as early in their life cycle as possible.

13.7 What 'working with producers' means

In broad terms, working with producers is likely to include some or all of the following:

- Making them aware of the preservation programme's existence, mission and operations
- Discussing ways in which the production process can help or hinder the preservation process
- Identifying benefits for both parties in minimising any hindrances to preservation
- Looking for mutually acceptable ways of facilitating the preservation process
- Identifying concerns of producers and looking for mutually acceptable ways to address them
- If appropriate, providing detailed advice on good practices such as the use of standards, formats, file management and metadata
- Negotiating arrangements for transfers and rights management
- Establishing agreements to take specific action, often based on working through pilot projects and joint evaluations.

13.8 Effective collaboration

The effectiveness of collaboration between a preservation programme and producers may depend on a range of factors such as:

- The nature of the relationship between them. For example, consider the difference in potential leverage for:

  - An organisational records archive with legal jurisdiction over the creation of records within their organisation
  - A nationally recognised data archive negotiating with independent researchers who produce datasets within a broad academic discipline
  - A government audio-visual archive seeking to convince independent record producers that their 'backyard' recordings are part of the national heritage
  - A small, specialised collection trying to preserve commercially produced, internationally marketed computer games.
- The readiness of producers to participate
- The technical expertise and insight the preservation programme can offer
- The skill of the preservation programme in negotiating mutually beneficial arrangements.

The preservation programme should seek to maximise its effective influence within realistic constraints.

13.9 Benefits

There are many potential benefits for the preservation programme in working with producers to overcome preservation barriers; there are also potentially benefits for the producer. Some of these are presented in Table 13-1.

Short term benefits to preservation programme	Long term benefits to preservation programme	Benefits to producer
established points of contact may make communication easier	improved choice of formats and how they are used, and opportunities to negotiate arrangements for bypassing security devices that block preservation copying	improved representation of output in archived collections
transfers may be easier, especially where automated 'gathering' does not work (see Ch 14)	improved transfer of documentation	more efficient workflows, less 'redo' work to meet archiving requirements
producer involvement in deciding what should be preserved	better understanding of roles and responsibilities	enhanced recognition of the value of their work
understanding what material is available and how it is viewed by the producer community	insight into future trends in production of digital heritage materials	bringing their work to a wider audience which may create new markets and foster wider interest
identification of otherwise 'invisible' materials	basis for identifying priority issues with specific communities	may help establish credibility for new forms of producing and distributing information

less costly transfers	less costly long term preservation	increased interest in using open source software (for preservation purposes) may encourage new collaborative production models

Table 13-1 *Some potential benefits of collaboration between preservation programmes and producers of digital heritage materials*

TECHNICAL AND PRACTICAL ISSUES

13.10 Recognising differences

It is important to recognise that creators of digital materials work in different environments and are likely to be quite diverse in many ways: how they approach their work, their size of operations, the organisational and technical support at their disposal, and their interest in long-term access issues. For example, scientists collecting data are likely to have an overriding interest in how accurately and securely their data is protected; how well any proposed formats and standards fit with their working needs; the convenience of transfer arrangements; and maintenance of moral rights and access controls over their data. On the other hand, commercial publishers of CD-ROM packages are likely to be more interested in controls on unauthorised copying; the costs and risks associated with providing 'unprotected' versions to a preservation programme; the potential re-use of their content; and their licence obligations to software owners whose products they have used.

13.11 Approaches to working with producers

There are many ways in which those responsible for preserving digital heritage may approach working with those who create and disseminate it.

- An obvious first step is to identify who is involved. Some action can be usefully undertaken with industry representative groups, but some action may require individual contact and negotiation
- Creators also need to know who to deal with. Preservation programmes should proactively promote awareness of their own role
- It may be advantageous to identify particular groups of producers and work with them, addressing specific issues, rather than trying to resolve everyone's concerns at a generic level
- At a broader industry level, it may be helpful to develop a code of practice that sets out agreed understandings about roles and responsibilities, and defines the scope and terms of ongoing cooperation
- Many sectors have active industry groups that provide forums for discussing issues. As well as offering opportunities for dialogue with industry leaders, such forums may help in establishing new norms of thinking that incorporate a longer term perspective
- It is important for preservation programmes to offer positive encouragement and feedback for the steps producers are willing to take, and to provide a level of

accountability for the way preservation programmes deal with their materials. Evidence that cooperation is leading to effective preservation action may well encourage producers to accept and support further collaboration.

13.12 A 'two way street'

In many situations, working with producers means a real input by the preservation programme, not just the producer. Possible areas of input may include:

- Providing written guidelines and specifications
- Providing training for staff
- Help in designing systems and workflows
- Exchange of information and working tips
- Succession arrangements for material in a producer-managed preservation programme.

SPECIAL CONSIDERATIONS

13.13 Specifications and best practice guidance

Guidance on good practice is likely to include advice on the following:

- Organisational issues that will make it easier to manage digital materials
- Project planning, emphasising system design prior to the creation of any records or publications
- Choice of carrier. Producers should be encouraged to use 'industrial strength' products that will survive long enough for the data to be transferred to other carriers, either by the producer or on transfer to the preservation programme
- Choice of appropriate file formats and data standards. Unless there are very good reasons to do otherwise, creators should be encouraged to use very widely adopted, well-standardised file formats that fit their purposes. Generally speaking, data in simpler formats using open source, non-proprietary software are easier to preserve (although some proprietary applications achieve such widespread use that they may be accepted as an industry standard, especially if their specifications are openly published). Online materials published for public access should be readable by commonly used browsers. Structuring documents in a standard, easily recognised and durable format such as XML (Extensible Markup Language) should be considered for material of enduring value
- Validation of formats. It is not enough simply to choose a standard format and then to use it in non-standard ways: formats should be implemented in compliance with their standard and if necessary validated to remove any idiosyncrasies likely to complicate preservation. (There are many online tools available for validating a range of file formats)

- File names should be consistent and unambiguous
- Online files should be managed for persistent access through the use of a persistent identifier and resolver service, or re-direct messages if files are moved. A number of PI schemes are in use internationally in different sectors, although none are in universal use. The DOI (Digital Object Identifier) scheme used by commercial publishers to manage rights has the widest acceptance
- Creators should create good quality metadata for the resources they create, using a widely accepted schema such as MARC, the Dublin Core metadata elements or one of its many sector-based enhancements. The metadata will help users find and use their resources. Metadata should also be recorded that describes the technical nature of the digital objects, what is required to access them, and any changes in these details over their life cycle: this information will be needed in managing them. The metadata can be either embedded in the resources or stored in a linked metadata file
- File management. Preservation master files should be stored and managed separately from dissemination copies. Database management procedures should ensure that data is not overwritten before it is captured
- System security. Files and systems should be fully protected from damage or loss by adopting best practice security measures and by appropriate backup arrangements even for short-term storage
- Authenticity. All files should be identified and their provenance and history documented to provide continuous evidence of authenticity
- Training. Staff, contractors and others coming into contact with the digital materials should be guided by appropriate procedures and manuals, and be adequately trained, motivated and equipped to use them
- If access or copying barriers are considered necessary to protect intellectual property, they may well make preservation impossible. Arrangements will be needed to allow preservation processes such as copying to take place
- Initial steps in maintaining access may include keeping all the software required for access, as well as any specialised hardware. This will not be an effective long-term strategy but may well be necessary in the short term
- There may be a need to evaluate digital materials, decide how long they should be kept and by whom, in accordance with an approved policy such as an archival disposal authority.

FOR PRESERVATION PROGRAMMES WITH FEW RESOURCES

13.14 Reducing the load

Digital preservation programmes with few resources may find they are unable to spare any resources to work with producers. However, it may be possible to make a worthwhile investment in reducing future costs by taking limited, targeted actions aimed at influencing the material they have to manage. For example:

- Engaging with just one or two producers to explore what can be achieved may reveal some easy steps that can be agreed on
- Restricting the range of materials selected to a few well-standardised formats may make it easy to provide specifications that producers can follow without needing individual input
- Making use of existing guidelines prepared for other programmes may achieve the same aim, so long as the guidelines are appropriate. (Many such guidelines are available online from organizations such as the Library of Congress, data archives within the UK Arts and Humanities Data Service. Various organizations have also negotiated licence agreements with commercial publishers, for example, that may provide good models for discussion with local producers.)

13.15 Spreading the load

Preservation programmes may also find partners willing to share the load of liaising with producers:

- It may be possible to find a partner institution with a better resourced programme who has already established good working arrangements with a producer community. Under a development agreement, producers may be willing to include other partners in the agreement so long as there are adequate safeguards for their interests
- It may also be possible for a number of smaller programmes working in the same region to form a consortium to negotiate arrangements with producers on behalf of all.

CASE STUDIES

Table 13-3 presents some possible scenarios in a variety of environments.

REFERENCES – **where to look for more information**

Cross references

Rights issues also see Managing rights: chapter 15

Offsite references (all links viewed March 2003)

- Arts and Humanities Data Service, (nd). *Digitisation: a Project Planning Checklist.* http://ahds.ac.uk/checklist.htm
- Canadian Heritage Information Network (CHIN) (2002). *Creating and Managing Digital Content.* http://www.chin.gc.ca/English/Digital_Content/index.html

- Cornell University Library (2003). *Moving Theory into Practice: Digital Imaging Tutorial.* http://www.library.cornell.edu/preservation/tutorial/contents.html
- Digital Library Federation, (2002). *Benchmarks for Digital Reproductions of Monographs and Serials.* http://www.diglib.org/standards/bmarkfin.htm - bench
- Humanities Advanced Technology and Information Institute (HATII), University of Glasgow, and the National Initiative for a Networked Cultural Heritage (NINCH) (2002). *The NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials.* http://www.nyu.edu/its/humanities/ninchguide/
- Institute of Museum and Library Services (IMLS) (2001). *A Framework of Guidance for Building Good Digital Collections.* http://www.imls.gov/pubs/forumframework.htm
- Library of Congress (2002). *Building Digital Collections: Technical Information and Background Papers [for American Memory Programme].* http://lcweb2.loc.gov/ammem/ftpfiles.html
- MATRIX: (The Center for Humane Arts, Letters and Social Sciences Online at Michigan State University), (nd). *Working Paper on Digitizing Audio for the Nation Gallery of the Spoken Word and the African Online Digital Library*. http://africandl.org/bestprac/audio/audio.html
- National Library of Australia (2002). *Safeguarding Australia's web resources: guidelines for creators and publishers.* http://www.nla.gov.au/guidelines/webresources.html
- Natural Environment Research Council (NERC) (UK), (2002). *NERC Data Policy Handbook*. http://www.nerc.ac.uk/data/documents/datahandbook.pdf
- Pockley, Simon (1998). *Cinemedia: Good Practice Guide.* http://www.acmi.net.au/FOD/DuckDigital/GoodP.html
- Rowe, J (2002). 'Developing a 3D Digital Library for Spatial Data: Issues Identified and Description of Prototype' in *RLG DigiNews, 6(5).* http://www.rlg.org/preserv/diginews/diginews6-5.html - feature1
- Townsend, S, Chappell, C, Struijve, O, (1999). *Digitising History: a Guide to Creating Digital Resources from Historical Documents.* (An AHDS Guide to Good Practice). http://hds.essex.ac.uk/g2gp/digitising_history/index.asp

	Library with preservation rights	**Library with licensed access**	**Institutional archive**	**Specialist A/V archive**	**Data archive**	**Community project archive**
Likely level of influence or control:	Poor: may be hard to identify producers or contact them may be extremely diverse set of producers (eg Web publishers) issues with commercial rights unable to specify formats to be used	Good control over formats but may be poor control over preservation	Potentially good if able to establish specifications and standard procedures may have legislated or organisational sanctions to enforce compliance	May be good for commissioned depositors but poor for others may be diverse formats and standards used	Likely to be good if has accepted role and credibility with producer and user communities	Likely to be good if involved early in project planning but poor if left to end
Possible influence strategies:	Identify and work with representative producers try to establish a code of practice education programme, seminars, guidelines emphasise benefits for producers may require individual negotiation may need to seek legal deposit law	May require protracted negotiations to secure ongoing accessibility	Education programme and technical support to encourage compliance influence the specification, design and procurement of record keeping systems and practices	Establish standard formats for acceptance into programme develop code of practice with industry groups or industry leaders develop close relationships with the producer community	Promulgate requirements help producers to design their projects encourage the deposit of contextual information	Work in close partnership with producers provide tools that make it easy for community participants to comply integrate preservation programme into the community project objectives

Table 13-3 *Some opportunities to work with producers in various sectors*

Chapter 14. Taking control: transfer and metadata

INTRODUCTION

14.1 Aims

This chapter aims to provide both management and technical advice on issues to do with the control of digital heritage by preservation programmes.

14.2 In a nutshell

Controlling what happens to digital materials is a key preservation step. In most cases this requires the safe transfer of data and documentation to the care of a preservation programme, where they are given unique identification, and described using various kinds of metadata. Metadata enables digital materials to be found and, crucially from a preservation point of view, to be managed and re-presented accurately. Although preservation metadata standards are still developing, programmes must describe the technical characteristics, provenance and preservation objectives of the digital materials in their care.

KEY MANAGEMENT ISSUES

14.3 Transferring data to a safe place

These Guidelines recommend the transfer of digital heritage materials from an operating environment to a safe place to avoid the risks of damage or loss associated with day-to-day use of digital files. In most cases this requires the transfer of data into the care of a responsible preservation agency.

The transfer process itself is not without risks as it provides opportunities for data to be lost, changed, misidentified, or divorced from the context that gives it meaning.

14.4 Rights issues

Because the producers of digital materials generally have some rights in the materials they produce, transfer raises a number of legal and moral rights issues.

14.5 Imposing control

Once digital materials have been transferred, they must be controlled and organised in effective and efficient ways. This generally includes requirements that materials can be easily located, accessed, used, managed and preserved, in accord with permissions.

PRINCIPLES IN ADDRESSING THESE CHALLENGES

14.6 Building on past practices

Transfer and control are long-established practices in managing non-digital heritage. When applied to digital materials, these processes must be modified.

- An appropriate legal basis for transfer is required. It must address concerns over the ease with which materials can be re-used, as well as the need to copy data for its preservation
- Transfer must affected without loss of data, often using quite different methods from those used for transfer of non-digital materials
- The transfer of accompanying documentation is particularly critical for digital data that may not be understandable without it.

14.7 Two approaches to effecting transfers

Most transfer strategies are variations of two basic concepts: producers *pushing* digital materials to the preservation programme, or the preservation programme *pulling* materials from the producer.

Programme managers must decide which approach will be most suitable for the materials being transferred and for the workflows of the parties involved.

14.8 Controlling formats and standards

Many programmes impose controls at the point of transfer on the formats of the material they receive. The purpose of this is to simplify preservation by reducing the variations that have to be managed in storing the material and in keeping it accessible. Not all programmes are able to restrict the formats they accept, but they should seek to verify that formats have been used in a standard way.

14.9 Controlling material by identification

Digital files must be given suitable file identifiers so they can be retrieved. Each file within a storage system must be identified with a unique file name so that it cannot be confused with any other file.

It is also most important for preservation programmes to ensure that the materials they keep can be reliably found, whatever their location. The Universal Resource Locator (URL) used to identify Web-based resources, for example, does not allow users to find material once it has been moved. Thus, items can be effectively lost even though it may still exist and be well protected. Overcoming this problem requires some form of *persistent identification*, built around an identifier and a means of *resolving* or linking to the file in its current location. There are a number of schemes proposed or in place including the Digital Object Identifier (DOI) used by publishers, and various schemes being investigated by libraries and archives, but none has yet found universal acceptance.

14.10 Controlling material by description

Preservation programmes use metadata – structured information about data resources – to describe the digital materials in their care. There are at least three compelling reasons for describing digital heritage materials in detail:

- So they can be found, assessed, made available and understood. This need has led to the development of *resource discovery* metadata ranging from simple listings of file names to extensive descriptions encapsulating rich contextual information. Resource discovery metadata schemes such as Dublin Core, MARC, archival description standards and museum catalogues, are important tools for preservation programmes to consider and use as appropriate to their needs
- So that workflows can be managed. Preservation programmes generate large amounts of information about way material is created, transferred and used; about rights and who is authorised to do what; and other management processes. One example of a very extensive *resource management metadata* set is the US NISO *Data Dictionary-Technical metadata for digital still images* published as a draft standard in 2002 (available online at <http://www.niso.org/standards/resources/Z39_87_trial_use.pdf>
- So that preservation programmes can understand how to re-present digital materials when they are needed for access. *Preservation metadata* describes the means of providing access, along with those elements of resource management metadata required to manage preservation processes. It is critical for any preservation programme; its careful design and management is especially important for large collections that must be processed with as much automation as possible.

14.11 Metadata as an information resource

Metadata is itself an information resource that must be managed and preserved, along with the material that it describes.

14.12 A standards approach to metadata

Individually developed metadata schemes can be successful in describing collections of digital materials, but there are increasingly good reasons to use a standardised approach in line with other widely adopted schemes:

- To reduce the considerable costs of developing individual schemes
- To take advantage of available software tools that automatically recognise and record standard metadata elements from digital materials, greatly reducing the cost of metadata capture
- To allow preservation programmes to share information, making their collections visible and searchable to a much wider audience
- To allow collection materials to be moved from one repository to another without the need for wholesale rewriting of metadata
- To encourage the standardisation of preservation processes that are described and controlled by the metadata.

Apart from preservation metadata, which is discussed in more detail below, further

information about metadata is beyond the scope of these Guidelines. As a principle, managers of preservation programmes should make themselves aware of standardised metadata schemes that are widely used in their sector of interest, and adopt those that will best meet their needs. They should also pay attention to the evolution of metadata standards by various international communities interested in managing digital resources.

TECHNICAL AND PRACTICAL ISSUES

14.13 Initiating data transfer

- The reproducibility of digital materials means that transfer no longer requires removal of the material from one site in order to move it to another. Perfectly authentic copies can be transferred for preservation while 'live' copies remain with the creator
- The timing of transfer may be critical. Even though material may have been selected for preservation, selection of itself does nothing to slow down processes like media deterioration or obsolescence of technology. Transfer needs to happen quickly enough to pre-empt these threats
- The transfer process may need to include, in addition to the selected files:
  - Transfer of documentation (including packaging for published physical format carriers such as CDs and diskettes), data rules, and information about provenance and original context
  - Transfer of existing metadata
  - Information about rights including any licence agreements
  - Information about the means of providing access, and possibly the means themselves - any special software and even hardware – that is needed for current access.

14.14 Specifying media and file formats

There is no standard way of effecting the physical transfer of digital materials. Data can be transferred on a wide range of physical carriers such as various forms of diskettes, CDs, tapes, cartridges, and disk drives; or through communication networks using means such as email attachments, file transfer protocol (FTP), and downloading from Web sites. The choice of transfer media depends on the needs of the parties involved.

Whatever means are chosen, the data must remain secure. Some transfer environments may present particular risks for specific media; for example, physical carriers may be easily lost or stolen, while communication networks may be unreliable and it may be safer to hand deliver a physical carrier.

The transfer medium must allow the data to be loaded and retrieved. When both sender and receiver use the same technologies, transfer should be relatively straightforward. When technologies are mismatched, one or both parties will need to bear the cost of using different technologies.

Preservation programmes may have facilities to handle a wide range of media, or their

facilities may be more restricted. Physical carriers require specific hardware which preservation agencies may not be able to provide. In such cases they will need to decide whether it is reasonable to require transfers via specific media that they can process, or to invest in facilities to handle a wider range of media.

Some considerations in deciding on transfer media are included in Table 14-1 below.

If data must remain on transfer media for medium-term storage …	avoid short-term carriers such as diskettes or DAT tape
If data will be immediately loaded to another carrier for storage …	short-term media may be suitable for transfer
If the costs of accommodating a wide range of media are prohibitive …	specify a narrower range of media
If workflows are built around specific media …	specify media that suits workflows, or specify media that producers will find easy to supply and adjust workflows

Table 14-1 *Decision factors in choosing transfer media*

14.15 Transfer strategies

Transfer of data usually involves the preservation programme either receiving files from the producer ('push' approaches), or actively taking files from the producer's site ('pull' approaches).

There are many push or deposit approaches that are used, such as sending files loaded onto a physical carrier through the mail or by courier; attaching files to email messages; or transmitting them by FTP directly to the preservation programme's server. Push approaches have many advantages, as they allow the producer to deposit more easily preserved versions of their work than may be publicly available, and give producers more opportunity to influence selection of what will be preserved.

On the other hand, preservation programmes relying on deposit may find that transfers depend on production factors beyond their control, including changes in personnel, changes in priority, or declining levels of interest, all leading to inconsistent transfers.

Pull approaches place more control in the hands of the preservation programme regarding timing and content of transfers. Some producers consider this an infringement of their rights and either block the software used to copy their files or demand rights agreements, so the control offered by pull approaches is not absolute. (On the other hand, many producers are happy to have their material captured, preserved and made available at no cost to themselves.)

Gathering or automated *harvesting* of material from producers' sites is made possible by communications networks. Using software programmed to search the network for files that satisfy specified criteria, preservation programmes can copy and download files to their own computer systems. Such an approach is widely used by Internet search engines and by most preservation programmes capturing networked material. Various indexing and 'search and retrieval' software programmes are available, with varying capabilities for defining what should or shouldn't be retrieved.

Gathering can be a highly efficient means of capturing data, but it can also present problems. Some files may be invisible to the software, being accessed only via a user interface that interacts with underlying data. Many producers also store higher quality versions of their work, such as images and audio files, separately from derivative versions suitable for network delivery: gathering misses the versions which should be preserved and captures versions intended only for short-term access.

A solution to these dilemmas is often found in mixed arrangements whereby producers agree to place a suitable version of their work where the preservation programme can gather it.

14.16 Quality control

Regardless of the means of transfer, preservation programmes should check material as it is received to confirm that all the required files have been received, that they work as intended, and that metadata and any other documentation is in order.

14.17 File identification

Digital objects can have a number of identifiers variously used for local control, for system-wide identification, and for global access, (just as a book on a library shelf can be identified by its title, a classification number, a shelf location, an accession number, a record number in the catalogue database, an International Series Book Number, and so on).

Persistent identification of some kind is needed so that items can be found even if they are moved in a storage system. Any links embedded within objects will only continue to work if linked to persistent identifiers.

Some alternative approaches include:

- Within a small system, ensuring all users are informed of any location changes
- Automatic re-direction messages that take users to the new location
- Managing file storage to minimise file movements
- Use of a persistent identifier (PI) scheme, involving a unique file name and subscription to a resolver service that registers PI's and their current locations. (programmes should note that existing PI schemes are largely either still under-developed, or available but expensive to participate in.)

14.18 Looking after metadata

It has already been noted that metadata must be not only recorded, but also looked after. There are a number of elements to this:

- Structuring. Organising metadata into a standardised document structure such as an XML template should make it easier to preserve
- Linking. The links between metadata records and the digital objects they describe must be maintained. There is much debate about the best place to store metadata to achieve this. While some metadata must be attached so that software tools can automatically process the material, there is disagreement on whether full metadata

records should be stored separately, attached to, or even become part of, the objects they describe. Separate storage allows metadata to be accessed and updated without needing to extract the linked digital objects from storage – a great advantage. On the other hand, many programme managers worry about the potential for the essential link between object and record to be disrupted over long periods of time. Managers should assess the risks they operate with and decide which approach is more suitable

- Quality control. Ensuring the trustworthiness of metadata records is a high priority. Quality control measures are needed whenever metadata records are created or changed
- Protection. The integrity of metadata records must be ensured, requiring the same preservation attention as the objects they describe.

14.19 Preparing the archival package for storage

Once the digital material has been transferred and any necessary control and description work undertaken, it must be prepared for entry into a storage system, ensuring the various part of the information package (including the content and any metadata) are linked, and a data stream created that can be safely stored on the storage media in use and can found by the appropriate file searching programmes.

The package is then saved to storage.

Before putting the digital object in storage as a preservation master, many programmes create additional copies, for at least two very good reasons:

- In order to have a copy available for use without the need to extract the preservation master from storage. Use copies are often optimised for access with currently available communications and display technologies, (such as low resolution, compressed versions of image files that can be much more quickly delivered online). Derivative access copies generally do not need to be preserved across changes in technology, and they often do not have detailed preservation metadata records.

- In order to store objects in more than one format, opening up alternative strategies for providing access in the future. As discussed in chapter 16, it is good practice to retain copies of digital objects in their original formats, regardless of the need to create new formats as a preservation master or current access copy.

Obviously, any parallel versions must be managed as separate but related digital objects.

SPECIAL CONSIDERATIONS

14.20 Preservation metadata

Preservation metadata is structured information about a digital object, which:

- Identifies the material for which a preservation programme has responsibility
- Communicates what is needed to maintain and protect data
- Communicates what is needed to re-present the intended object (or its defined essential elements) to a user when needed, regardless of changes in storage and access technologies
- Records the history and the effects of what happens to the object
- Documents the identity and integrity of the object as a basis for authenticity
- Allows a user and the preservation programme to understand the context of the object in storage and in use.

Arrangements for recording preservation metadata must accommodate the fact that the same basic content (or conceptual object) may exist in many manifestations during its life. Some of these manifestations will co-exist as digital objects, while others may follow each other in a series of separate or overlapping generations. Some preservation programmes reflect this by creating a record for a single version identified as a Preservation Master, documenting variants and changes as part of the history of that object. Other programmes create a record for each manifestation requiring preservation action, ensuring the relationships between manifestations are explicit in their metadata records.

The information required for preservation metadata is often divided into two classes (in line with the Reference Model for an Open Archival Information System or OAIS referred to in chapter 8):

- *Content information*, consisting mainly of details about the technical nature of the object which tells the system how to re-present the data as specific data types and formats. As access technologies change, this re-presentation metadata also needs to be updated
- *Preservation description information*, consisting of other information needed for long-term management and use of the object, including identifiers and bibliographic details, information on ownership and rights, provenance, history, context including relationships to other objects, and validation information,

Obviously, some of this metadata may refer to other information objects such as software tools and format specifications that must also be managed. The interdependent nature of digital materials means that programmes often have to manage networks of linked objects and their metadata.

There are still no accepted standards for preservation metadata schemas for universal use, so programmes may have to choose between accepting (and possibly adapting) one of a number

of models being used by others, or designing their own schema (either as a complete solution or as a minimal interim one until a standard emerges).

Many national archives authorities have released metadata specifications for record keeping systems that include preservation needs. In the library field, an international working group convened by OCLC and the Research Libraries Group (RLG) released a recommended preservation metadata framework in mid-2002 (available online at <http://www.oclc.org/research/pmwg/pm_framework.pdf>). Their report is a good starting point for exploring the metadata that may be needed.

An interesting implementation by the National Library of New Zealand attempts to adapt the OLCLC/RLG work to a particular programme and its circumstances (available online at <http://www.natlib.govt.nz/files/4initiatives_metaschema.pdf>). This schema proposes the following elements (somewhat summarised here):

Describing a digital object
Name of the object
Local identifiers
Global persistent identifier
File location in storage system
Date when created as preservation master
Overarching technical composition (no of files of each MIME type)
Structural type (eg text, image)
Hardware required for object to function
Software required for object to function
Special installation instructions
Built-in access inhibitors and facilitators
Quirks (in-built anomalies)
Authentication or validation keys
Who created metadata and when

Describing any process applied to an object (including creation)
Name of process
Purpose
Agent who carried out process
Agent who approved process and when
Hardware used
Software used
Steps involved in process
Outcomes
Standards or specifications used
When completed

Describing technical characteristics of any files within the object
Specific file identifiers
Relationship to other component files
File size
When created
MIME type/format (eg image/tif)
Version
Key file that provides access
Characteristics of specific file types
(eg for image files: resolution, dimensions, tonal resolution, colour space, colour management, colour lookup table, orientation, compression)
(eg for text files: compression, character set, associated DTD for structured text, structural divisions)
(eg for audio: resolution, duration, bit rate, compression, encapsulation, track number and type)
(eg for video: frame dimensions, duration, frame rate, compression, encoding structure, sound)
(eg for datasets: common elements above only)
(eg for executable files: common elements above only)

Describing update of metadata
Agent modifying metadata
When modified
Field modified

FOR PRESERVATION PROGRAMMES WITH FEW RESOURCES

14.21 Transfer

Programmes with few resources may need to explore ways of reducing transfer costs:

- 'Push" arrangements may require less investment by the preservation programme and shift most of the cost of transfer to the producer. However, without agreements about the media, formats and quality control to be used by producers when transferring material, short-term savings may produce greater preservation costs in the long term
- Well chosen restrictions on the range of media and formats accepted by the programme may produce savings
- programmes may be able to store transferred material on their transfer media if relatively stable carriers have been chosen, and if backup copies can be made for security.

Some communities without access to separate preservation agencies may have to pursue a 'non-transfer' model, setting up the best preservation arrangements they can within an operating environment. Even in these circumstances, many of the same principles apply: ongoing accessibility is more likely with some kind of internal transfer to even a modest 'back up archive' where files can be managed outside the normal risks of operational use. Files will still need to be sufficiently well described and protected to allow later transfer to a more secure preservation programme.

14.22 Metadata

The costs of recording metadata can be a significant part of overall preservation costs. There may be potential for savings by either reducing the amount of information recorded to a minimum (and accepting that both access and preservation will be made more difficult); or by investing in software that will capture metadata automatically (which will become easier as metadata standards develop).

In choosing a minimal set of metadata, programme managers may find it helpful to consider what users will need in order to find material, and what questions will require answers in taking any foreseeable preservation action.

CASE STUDIES

14.23 Case study 1

A national library collecting online digital publications effects transfers by gathering files from publishers' sites, using programmable searching, copying and downloading software such as HTTrack, in accordance with agreements negotiated with each site owner. The gathering process involves staff in looking for potential sites that might

meet the library's selection guidelines, deciding what should be captured and how far links on the site should be followed, (the selection policy suggests that linked documents on the same site should be captured, but no other links followed). When files are downloaded by the software, staff check to see that all desired material has been downloaded and that all files work. A metadata record is created using a mixture of software-generated and manually entered data. An individual entry page is created for each title captured, using a system-generated template, so that users can understand what they are getting and how it relates to both the publisher's Web site and to other material captured in the archive. When completed, the metadata record, which includes a link to the captured objects, is saved to the metadata repository, and the captured objects are saved to the repository mass storage system.

14.24 Case study 2

A small ethnomusicology archive receives field recordings from collectors on DAT tape, which is cheap and convenient for collecting use but unsuitable for storage. After checking that the material fits within the archive's collecting policy, and that recording quality is adequate, staff accept the material, manually entering information about the consignment into a separate database. The material is accessioned and allocated a running number in the collection. The data on the tape is copied to two sets of CD-Rs: one as a preservation copy and one as a backup. The DAT original is shelved as an access copy for short-term use, and the CD copies are shelved separately. The metadata record is updated with the location of all copies.

REFERENCES – **where to look for more information**

Cross references

Liaison with producers also see Working with producers: chapter 13

Metadata and means of access also see Maintaining accessibility: chapter 17

Offsite references (all links viewed march 2003)

1. Transfer

The (UK) Arts and Humanities Data Service (AHDS) and its affiliated data archives (in the fields of literature, archaeology, visual arts, history and performing arts) have produced a number of excellent 'guides for depositors'. These include good technical information on preferred formats for a wide range of types of materials. They may serve as good models for similar data-based programmes. For example:

- History Data Service (nd). *Guidelines for Depositors*. http://hds.essex.ac.uk/depguide.asp
- Oxford Text Archive (1999). *Depositing with the OTA: the Depositors Guidelines*. http://ota.ahds.ac.uk/publications/ID_Depositing-Introduction.html
- Visual Arts Data Service (nd). *Guidelines for Depositors*. http://vads.ahds.ac.uk/depositing/depositor_guidelines.pdf

2. Persistent identification

- Corporation for National Research Initiatives (CNRI), (nd). *The Handle System.* http://www.handle.net/index.html
- Dack Diana (2001). *Persistent Identification Systems (Report on a consultancy for the National Library of Australia).* http://www.nla.gov.au/initiatives/persistence/PIcontents.html
- International DOI Foundation (nd). *The Digital Object Identifier System.* http://www.doi.org/
- Internet Engineering Task Force (IETF), (2001). *Uniform Resource Names (URN).* http://www.ietf.org/html.charters/urn-charter.html
- National Library of Australia (2001). *Managing Web Resources for Persistent Access.* http://www.nla.gov.au/guidelines/2000/persistence.html
- The PURL Team (nd). *PURL – Persistent URL Homepage*. http://purl.oclc.org/

2. Metadata

Metadata standards and initiatives abound in various fields of heritage management, with extensions or adaptations to accommodate digital materials. For examples, see:

- *Dublin Core Metadata Initiative*. http://dublincore.org/
- IFLA Universal Bibliographic Control and International MARC Core Programme (UBCIM) (2000). *UNIMARC Guidelines no 6: Electronic Resources.* http://ifla.org/VI/3/p1996-1/guid6.htm
- International Council on Archives,(1999). *General International Standard Archival Description*, 2nd edition. http://www.ica.org/biblio/cds/isad_g_2e.pdf
- Consortium for the Computer Interchange of Museum Information, (1999). *CIMI Dublin Core Metadata Testbed Project.* http://www.cimi.org/old_site/documents/meta_webliography.html
- International Association of Sound and Audiovisual Archives, (1998). *The IASA Cataloguing Rules.*http://www.iasa-web.org/icat/

Some preservation metadata sources:

- Colorado Digitization Project Metadata Workgroup, Audio Taskforce (2002). *Metadata for Digital Audio (draft).* http://coloradodigital.coalliance.org/digaudio_meta.pdf
- National Library of New Zealand, (2002). *Metadata standards framework – preservation metadata*. http://www.natlib.govt.nz/files/4initiatives_metaschema.pdf
- NISO/AIIM, (2002). *Data dictionary – technical metadata for digital still images,* released as draft standard for trial NISO Z39.87 – 2002. http://www.niso.org/standards/resources/Z39_87_trial_use.pdf
- *Preservation metadata and the OAIS Information Model: a metadata framework to support the preservation of digital objects: a report by the OCLC/RLG Working Group on Preservation Metadata,* (2002). http://www.oclc.org/research/pmwg/pm_framework.pdf
- Public Record Office (UK) (nd). *PRONOM* (concerning a database system that stores and provides information about file formats and the application software needed to open them.) http://www.pro.gov.uk/about/preservation/digital/pronom.htm
- The British Library, (nd). *Code of Practice for the Voluntary Deposit of Non-Print*

Publications. http://www.bl.uk/about/policies/codeprac.html

Chapter 15. Managing rights

INTRODUCTION

15.1 Cautionary note

These guidelines should not be interpreted as competent legal advice on rights issues.

15.2 Aims

This chapter is intended to highlight the serious responsibility of preservation programmes to be aware of rights issues, and to provide some general suggestions on how those issues may be approached.

15.3 In a nutshell

There are a range of right and expectations held by stakeholders, which preservation programmes must be aware of and, if necessary, include in their management planning. Many of these rights have legal implications, including intellectual property rights and privacy rights. Because preservation programmes must copy digital materials to preserve them, and because most programmes aim to provide some level of access, active rights management approaches are needed.

KEY MANAGEMENT ISSUES

15.4 Digital heritage and rights

Digital heritage materials are subject to a range of rights and expectations, some of which have legal force. Many, such as copyright, result from the intellectual property invested in the material. However, there may be other rights and expectations that also need to be taken into account.

15.5 A range of rights and expectations

The range of rights and expectations that preservation programmes may encounter and have to manage typically includes:

- Intellectual property rights of producers including copyright, which may exist in various layers associated with different aspects of the material; the right to set conditions of access and use; and the creator's moral right to be recognised
- Legislated rights of certain institutions to collect, preserve and provide access to some materials

- The rights and expectations of privacy, confidentiality and authorisation of use associated with some subjects of materials such as organisational records, oral history recordings, personal data and private communications
- Expectations of users regarding access and use
- Expectations of the broader community that material of enduring heritage value will be preserved and made accessible within the regime of rights established in law.

15.6 Basic rights required for preservation activities

Preservation involves many processes where rights issues are relevant. In order to achieve continuity of digital heritage, preservation programmes must:

- Obtain and hold material, usually involving making copies
- Make further copies for preservation purposes
- If necessary, bypass devices used by producers to limit access and prevent copying
- Decide what materials and what aspects of materials should be preserved
- Add metadata
- Modify file structures and file names if necessary
- Use whatever means are available at the time to preserve accessibility
- Provide managed access for authorised users.

15.7 Challenges

Obtaining permissions to cover these activities may be difficult:

- Producers and other rights owners may be unwilling to give permission
- Rights of access and rights of privacy and confidentiality are often in tension
- In an environment of fragmented or collaborative creation of digital materials, it may be hard to identify or negotiate with all rights owners
- The legal position maybe ambiguous, as many jurisdictions are still in the process of clarifying legal frameworks of rights and how they should be managed
- In dealing with globally networked materials it may be even unclear which legal jurisdiction applies: that in which material was produced, or published, or captured for preservation, or stored, or accessed – all of which may be different.

The costs of putting good rights management practices in place may be high, especially if individual negotiation is required. On the other hand, the costs associated with not managing rights issues adequately are also likely to be high.

PRINCIPLES IN ADDRESSING THESE CHALLENGES

15.8 Awareness

Preservation programmes must be aware of the legal frameworks in which they operate, including their legal rights, constraints and obligations. This may require reference to specific legal advice from a competent source. Even with good intentions to preserve important heritage materials, preservation programmes are responsible for seeking ways to achieve their mission without infringing the legitimate rights of others.

15.9 Advocacy

Preservation programmes must decide on the extent to which they should engage in advocacy on rights issues, presenting arguments for legislation that would make it easier for a wide range of digital materials to be preserved.

At a minimum, preservation programmes should ensure that interested parties are aware of the rights required for effective preservation action.

15.10 Finding workable solutions

While finding solutions to rights issues may not be easy, the problems are usually not insurmountable. Resolving them does require respect for the legitimate interests of others. Solutions can usually be developed through a cooperative approach that recognises mutual needs and benefits. Preservation programmes can make a large contribution by showing that:

- Sound management of rights is possible
- There are ways of meeting preservation objectives without jeopardising reasonable commercial interests
- Through their documentation and metadata services, preservation programmes can promote community knowledge and use of rights owners' products
- By selecting material for preservation, preservation programmes can confirm the importance of records, research results and other non-published materials.

Many preservation programmes have found satisfactory ways to approach rights issues, often in partnership with rights owners. Such models range from quite simple agreements with individual rights owners (common in data archives and in selective archives of Web publications), to long-sighted partnerships between very large commercial publishers and national libraries.

These models are usually based on a mixture of transferred, managed and retained rights. For example, the right to store and preserve material may be completely transferred, while the preservation programme is required to closely manage access and the producer retains copyright.

15.11 Legal frameworks

There may be a number of frameworks which allow preservation programmes to assume the right to collect and preserve specific digital materials. The most common include:

- Legal deposit or records management legislation
- Organisational rules governing corporate information
- Contractual requirements to deposit data
- Conditions of grants, awards, employment, or membership of organisations
- Rights inherited by one organisation from another
- Negotiated or purchased licence agreements
- Rights implied by voluntary submission of material to a preservation programme
- Some preservation agencies capture and store materials such as publicly available, free access Web sites without seeking prior approval. Some do this on the assertion of 'fair use' for material in the public domain; others rely on an 'opt out' option whereby rights owners are generally invited to express an objection.

It is the responsibility of the preservation programme to determine, on the basis of competent legal advice, whether any of these or other approaches is applicable, and what is required as an adequate legal defence.

15.12 Some common steps

Each situation requires its own set of arrangements, but preservation programmes should consider the need to take some common steps including:

- Determining the legal situation regarding rights specified by legislation, existing organisational rules, or licence agreements
- Identifying the rights that will be needed to carry out a preservation responsibility
- Identifying relevant rights owners, and other stakeholders with an influential interest in what rights are negotiated
- Preparing a clear explanation of what is needed and how it will be managed
- Approaching rights owners and negotiating a rights regime that is mutually acceptable
- Recording rights management responsibilities in metadata that is clearly and securely associated with the relevant materials

- Ensuring the responsibilities are understood by staff
- Having secure systems, procedures and tools in place to control access and copying, and to monitor compliance
- If necessary, isolating preservation actions from other kinds of access and use
- Ensuring users understand their legal rights and obligations
- Regularly evaluating systems and procedures to ensure they do what they are supposed to do
- Monitoring any triggers for a change in rights, such as the passing of a specified period of time.

15.13 Negotiating access conditions

The level of access that preservation programmes should seek will depend on their mission: it may be appropriate for some digital heritage materials to be subject to very limited access for privacy, security, or other reasons, whereas it seems reasonable to expect that published materials would be available for ongoing access through a well-managed preservation programme.

Some possibilities that may be attractive in negotiating access conditions include:

- Geographical restrictions, such as limiting access to onsite users
- Restrictions on the ability to copy, such as use of a stand alone computer without access to external networks or disk drives
- Restrictions on the number of users who can access the material at any one time
- Time thresholds allowing unrestricted access after a reasonable period for commercial exploitation
- Mutually agreeable triggers for a transfer of access rights, such as when the material is no longer available from a publisher's site
- Restricting access to authorised users who are required to meet specified conditions.

15.14 Managing rights

When rights have been negotiated, they must be managed as a core business responsibility of the preservation programme.

- Preservation programmes can expect to deal with large amounts of material, so the use of standard licence agreements covering classes of material will avoid the need to negotiate and manage rights item by item
- System tools to manage rights are available and can be expected to continue to evolve. Such tools record access conditions applying to individual items, record and filter requests for use, and report on usage. In choosing rights management tools, it is important to decide what tools are appropriate to

support a balanced approach to rights management

- It should be made easy for users to contact rights owners to negotiate their own permissions, such as the right to copy, where it is the user's responsibility to do so
- Making authorised access as easy as possible may act as a disincentive to unauthorised access and use
- Encouraging creators to use open source software should help reduce complications and costs involved in negotiating rights with proprietary software developers.

FOR PRESERVATION PROGRAMMES WITH FEW RESOURCES

15.15 Seeking efficiencies

Rights issues have the potential to add greatly to the costs of preservation programmes, so all programmes have an interest in finding efficiencies and in avoiding exposure to litigation. programmes with few resources may particularly need to look for standard agreements that reduce the costs of negotiating rights approvals. They may also have to accept that rights management is a limiting factor on the size of their operations.

Alternatively, they may need to limit their activities to materials that present minimal rights issues, for example because:

- They already have permission
- Rights have lapsed (although this is unlikely to be the case for digital materials for some decades to come)
- The producer community has a strong supporting interest in the preservation programme
- There is reliable legal advice that 'fair use' or other provisions would be a successful defence.

CASE STUDIES

15.16 Case study 1

A data archive working in an academic discipline uses a standard letter of agreement for depositors to sign, authorising the archive to make a copy of the data and to take any necessary preservation action including making further copies in whatever formats it judges to be best suited to providing reliable access. Depositors must indicate whether there are any restrictions to be placed on access for either a particular period of time, or particular classes of users, or particular kinds of use. The maximum period for closed access is 10 years. The archive manages rights manually, as the data

is not available online: user requests are checked against metadata records for the material requested before access is allowed.

15.17 Case study 2

A state library relies on legal deposit legislation that specifically authorises it to make and store copies for preservation purposes. Copyright conditions still apply, so the library informs users of the need to get permission from the copyright owner before making copies. The library negotiates access restrictions with owners of commercial publications to protect their commercial interests for an agreed period of time, usually set at 5 years during which only onsite single use is allowed. Many owners are happy with less restrictive access because it broadens the audience for their publications, while some require longer periods of restriction. A rights management metadata system is used to record restrictions and to approve or reject requests automatically.

REFERENCES – **where to look for more information**

Cross references

Working with producers also see chapter 13

Metadata also see chapter 14

Offsite references (all links viewed march 2003)

Many data archives use standard licence agreements with depositors to formalise the transfer of rights. For example, see:

- Oxford Text Archive, (2003). *Licence for depositors.* http://ota.ahds.ac.uk/, under "OTA Publications"

For significant examples of rights management negotiations producing positive results, see:

- IFLA and the International Publishers Association, (June 2002). *Preserving the Memory of the World in Perpetuity: a Joint Statement on the Archiving and Preserving of Digital Information.* http://www.ifla.org/V/press/ifla-ipa02.htm
- Koninklijke Bibliotheek (August 2002). *National Library of the Netherlands and Elsevier Science Make Digital Preservation History*. http://www.kb.nl/kb/resources/frameset_kb.html?/kb/ict/dea/ltp/ltp-en.html

Some other resources:

- CEDARS Project (2002). *CEDARS Guide to Intellectual Property Rights.* http://www.leeds.ac.uk/cedars/guideto/ipr/guidetoipr.pdf
- Kavcic-Colic, Alenka (2002). *Archiving the Web: Some Legal Aspects*, 68th IFLA Council and General Conference, Glasgow. http://www.ifla.org/IV/ifla68/papers/116-163e.pdf

Chapter 16. Protecting data

INTRODUCTION

16.1 Aims

From this chapter a programme manager should understand how important it is to maintain strict control over the integrity of the data underlying digital objects. Those involved in implementation should be able to use the information in the chapter as a basis for discussing specific requirements with IT specialists or service providers.

16.2 In a nutshell

Data protection is a fundamental of all preservation programmes. For many programmes, authenticity is also critically important. Authenticity relates to the ongoing integrity of data, and its clear and sustained identification. Data protection strategies include allocation of responsibility, technical infrastructure, maintenance, data transfer, proper storage of data carriers, backup, system security and disaster planning. Authenticity also relies on clear documentation of the origins and history of digital materials.

KEY MANAGEMENT ISSUES

16.3 Data storage and protection

Data must be stored. While it is appropriate to focus preservation attention on how best to re-present digital objects as originally intended, it must never be forgotten that the digital object has an underlying form as data. It is as data that it must be stored, managed and protected if any digital object is to be available for presentation to a user.

16.4 Authenticity

Heritage materials are often valued, at least in part, for their authenticity – the degree to which one can trust that they are indeed what they are thought to be. For archival records, scientific data, and many other kinds of digital materials, trust in their ongoing authenticity is critical, for without it they are of virtually no value.

Authenticity derives from being able to trust both the *identity* of an object – that it is what it says it is, and has not been confused with some other object – and the *integrity* of the object – that it has not been changed in ways that change its meaning.

Maintenance of both identity and integrity implies sustained and documented links

between an object as originally created and as currently presented.

Evaluating, maintaining and providing evidence of continued authenticity are key responsibilities for most preservation programmes.

16.5 Threats to authenticity

Authenticity can be jeopardised by:

- Threats to identity. Loss of certainty about how an object is distinguished from other objects damages authenticity. This may result from confusion in identifying data, changes to identifiers, or failure to document the relationships between different versions or copies
- Threats to integrity. Changes to the content of the object itself also potentially damage authenticity. Most such changes stem from threats to the object at a data level.

The nature of digital materials, and how they must be managed for preservation and access, both present challenges:

- Digital materials can be changed easily, with or without fraudulent intent, and even without any intent at all
- Changes that happen may not be obvious
- Preservation processes almost always involve making changes - transferring data from one system to another, from one carrier to another, adding or updating metadata, creating new copies that need new file names, changing the means of presentation as technologies change, and so on.

16.6 Threats to data integrity

Common threats to the ongoing integrity of data that preservation programmes are likely to encounter include:

- 'Natural' generation of errors that arise in digital storage systems
- Breakdown of carriers. Most carrier media have a reasonably short useable life before deteriorating to the point of unreliability for data storage
- Malicious attack, which may come from system hackers, viruses, staff or outside intruders interacting with the storage system
- Collateral damage from malicious acts such as terrorist attacks, acts of war or civil unrest affecting buildings or power supplies
- Inadvertent acts by staff or visitors such as turning off power, throwing out disks or tapes, or reformatting storage devices
- 'Natural' disasters such as fire, flood, or building collapse
- Business failure.

The likelihood and impact of these and other risks will vary from situation to situation. However, one can assume that all of these risks must be addressed.

PRINCIPLES IN ADDRESSING THESE CHALLENGES

16.7 How much authenticity is needed?

Where digital materials have value as records that offer evidence of some kind, authenticity is extremely important. Not all digital materials are made or selected to provide evidence: for example, they may reflect creative expression, the debate of ideas, or the desire to entertain and be entertained. Even for these materials authenticity may be an issue, as the integrity of their creators' work or ideas should be protected.

Ultimately, preservation programmes must decide how much to invest in ensuring that the authenticity of material in their care can be trusted, bearing in mind that object identity and data integrity are fundamental responsibilities.

16.8 The role of data protection

Data protection must play a key role in any preservation programme, for two reasons:

- So that there is a digital object for a user to access. This is a fundamental requirement: if data is lost or seriously corrupted it may be impossible to re-present the intended digital object at all, and the preservation process must be judged to have failed
- So that the integrity of the data can be maintained without tampering or corruption in order for users to trust the authenticity of the re-presented object.

16.9 The role of documentation

Documentation also plays a key role, for two reasons:

- By explaining the links between objects and by clearly distinguishing between them, it provides evidence of identity
- By showing what changes, if any, have taken place, with whose authority, and to what effect, it provides an audit trail to attest to authenticity.

16.10 Responsibilities for maintaining authenticity

It may not be practical to expect an entirely objective guarantee of authenticity - there may always be an element of trust or subjective judgment in deciding that authenticity has been sufficiently proven – however, it seems reasonable to expect that digital preservation programmes would accept three responsibilities:

- They must assess whether demonstrated authenticity is critical to the ongoing value of the material
- They must protect the material in their care from changes that would alter its

meaning. (This allows for external changes such as new interpretations, without allowing internal changes that would alter meaning)

- They must document the relationships on which the required level of authenticity rests. These include relationships between the object and its identifiers; between the object and its producer; between different objects; and between the object and how it has been managed.

16.11 Data protection strategies

Other kinds of heritage materials may have survived periods of neglect, but digital data will not. Digital objects require well planned, well managed, and sustained strategies to protect data as a minimum foundation of continuity. The strategies that are needed usually include:

- Clear allocation of responsibilities
- Provision of appropriate technical infrastructure, including systems, storage devices, and carriers to do the job
- Maintenance, support and asset replacement programmes for the systems
- Transfer of data to new carriers on a regular basis to ensure data is not threatened by media deterioration or changes in access hardware
- Appropriate storage and handling conditions for carriers
- A high level of redundancy as an insurance against the failure of any one copy or component; including appropriate backup regimes
- A high level of system security including controls on access to stored data
- Disaster preparedness planning.

These are covered in more detail in the Technical and Practical Issues below.

TECHNICAL AND PRACTICAL ISSUES

16.12 Using service providers

Of all the responsibilities of preservation programmes, storage and data protection may be the ones for which it is easiest to find suitable third party service providers. The considerable investments required for equipment and expertise may make this an attractive alternative to managing data in-house. However, the critical nature of data protection means that the preservation programme must still accept responsibility for ensuring that any contracted services deliver the necessary levels of care and control.

16.13 Practical aspects of data protection strategies

There is a reasonably standard suite of strategies used to manage data in long-term storage. Most are predicated on an assumption that the data carrier itself does not need

to be preserved, only the data.

- Allocation of responsibility. Someone must be given unambiguous responsibility for managing data storage and protection. This is a technical responsibility requiring a particular set of skills and knowledge as well as management expertise. Except for very small collections, data storage and protection require dedicated resources, working to an appropriate plan and accountable for these strategies
- Appropriate technical infrastructure to do the job. Data must be stored and managed with appropriate systems and on an appropriate carrier. There are digital asset management systems or digital object storage systems available that meet the requirements of digital preservation programmes. Once requirements have been determined, they should be thoroughly discussed with potential suppliers. Different systems and carriers are suited to different needs and those chosen for preservation programmes must be fit for their purpose. The overall system must have adequate capabilities including:
  - Sufficient storage capacity. Storage capacity can be build up over time, but the system must be able to manage the amount of data expected to be stored within its life cycle
  - As a fundamental capability, the system must be able to duplicate data as required without loss, and transfer data to new or 'refreshed' carriers without loss
  - Demonstrated reliability and technical support to deal with problems promptly
  - The ability to map file names into a file-naming scheme suitable for its storage architecture. Storage systems are based around named objects. Different systems use different architectures to organise objects. This may impose constraints on how objects are named within storage; for example, disk systems may impose a hierarchical directory structure on existing file names, different from those that would be used on a tape system. The system must allow, or preferably carry out, a mapping of system-imposed file names and existing identifiers
  - The ability to manage redundant storage
  - Error checking. A level of automated error checking is normal in most computer storage. Because heritage materials must be kept for long periods, often with very low human usage, the system must be able to detect changes or loss of data and take appropriate action
- Technical infrastructure must also include means of storing metadata and of reliably linking metadata to stored digital objects. Large operations often find they need to set up digital object management systems that are linked to, but separate from, their digital mass storage system, in order to cope with the range of processes involved, and to allow metadata and work interfaces to be changed without having to change the mass storage
- The currently available broad options for large scale storage carriers are discussed in Table 16-1 below:

Carrier	Access to data	Allows data modification?	Current storage capacities per unit	Speed of increase in capacity	Expected usable life of single unit	Other comments
Magnetic disk (eg hard disk)	fast random access	yes	up to 200 gigabytes	doubling every 12-18 months	around 5 years	generally fixed media
Magnetic tape	linear storage so takes longer to search and access data	generally no – 'read and write' requires data to be overwritten	up to 200 gigabytes	doubling every 12-18 months	around 5 years	portable media suitable for backup
Optical disk (CD, DVD)	fast random access, but slower than magnetic disk	yes, on some products	up to 4 gigabytes	slow because not used for very large archives or backups	wide range from say, 5 years for low quality products to several decades for high quality products	portable media. unit costs low; low cost consumer equipment widely available

Table 16-1 *Comparison of large-scale data carriers*

- Maintenance, support and replacement programmes. System components generally need to be replaced every few years. Hardware typically has a working life of around five years before technical support may become difficult to obtain. Storage carriers also need regular *refreshing* (rewriting of data) and periodic replacement by new carriers.

  The need to replace storage systems involves significant recurring costs, covering the equipment itself as well as the procurement and data transfer processes that precede and follow installation of new equipment. These costs must be built into long term budget planning.

  While the cost of replacing data carriers must be considered, replacement media typically offer increased storage capacity. Unfortunately, any savings are usually offset by growth in the amount of data to be stored.

  The market for data storage and management systems extends well beyond preservation programmes, so there are good COTS (Commercial Off The Shelf) products available. Using COTS technology is possibly the most easily managed, low risk and cost effective approach as technical support and upgrades are provided by vendors in a competitive marketplace. Standards are sufficiently widely used in the storage market to allow mixing and matching products from various vendors so that a number of upgrade and replacement paths are available when needed.

- Transfer of data to new carriers on a regular basis. Storage systems rely on safe and complete replication of data, rather than enduring carriers, for data protection. Data must be copied from carrier to carrier to avoid the impact of carrier deterioration. As new kinds of carriers prove their usefulness in storage systems, data is transferred from older kinds of carriers. This must happen before any hardware or software required to retrieve the data are discarded.

  Planning for data transfers is a management challenge, whatever the system used. For example, a small, low use archive storing data on shelved CDs, must keep track of the age and condition of the CDs as well as signs that CD technology will have to be replaced. More sophisticated mass storage systems generally automate decisions about regular transfer of data between carriers, but managers still need to decide when carriers should be replaced with new media, and when underlying technology has been superseded.

- Appropriate storage and handling conditions for carriers. Digital data carriers must be stored in conditions that do not accelerate their rate of deterioration.

  The main risks for data carriers are excessive temperature and humidity which endanger the carrier; dust or other particulates which may obscure access to the data; and in the case of optically encoded material, light, which may damage the optically inscribed data. Modern data tapes are of such a high coercivity, that accidental erasure by a magnetic field does not constitute a major risk.

  Magnetic data tapes may be integrated into a digital storage system. Typically this would be housed in a clean computer room with controlled temperature and relatively humidity set at 18°C, and 40% RH, a continuous influx of clean, dust-free air, with daily cleaning to prevent contamination. The conditions would fluctuate no more than 2°C and 10% RH in any given 24 hour period.

  Magnetic data tapes stored for optimum carrier life (away from the computer room environment) should be stored under more stringent conditions, at a temperature between 18°C and 10°C, with a daily tolerance of no more than 1°C, and between 30 and 40% RH with a tolerance of no more than 3%RH.

  Optical carriers, such as CD-Recordable, should be stored under similar conditions, in a darkened environment due to their sensitivity to light.

  There are suggestions that very low temperatures (approaching or lower than 0°C), may be detrimental to the life expectancy of certain carriers, however, this has not been proven.

- Redundancy and backup regimes. The importance of redundancy and backup regimes cannot be overemphasised: they are fundamental to all digital preservation programmes as a basic insurance against damage or loss to any single copy.

  While storing multiple copies of the same data does offer some protection

against failure, preservation programmes must also consider the risks of a disaster which damages all copies stored at the same site. Storing copies at different sites is a basic requirement; to avoid the impact of region-wide disasters such as floods, earthquakes, wildfires, and war, programmes should consider the need to store additional backup copies of important data outside their own region.

Preservation programmes may also need to adjust normal backup schedules so that preservation data, which must be kept, is refreshed (i.e. rewritten) not overwritten with new data.

- System security. Security controls are required to ensure that stored data are only exposed to controlled, authorised processes. Standard IT security measures for vital information assets are fully applicable and absolutely required.
- Disaster planning. Standard IT disaster recovery plans must be in place, and must be tested regularly. The plans may include realistic arrangements for attempting data recovery from damaged carriers, but data recovery is expensive and uncertain, and it should be seen as a very unsatisfactory alternative to proper recovery-from-back-up arrangements.

16.12 Managing risks

Table 16-2 presents a simplifed risk analysis of some of the more common threats to data in storage.

Threat	What it affects	Likelihood	Speed of onset	Impact	Prevention options
'Natural' generation of errors	data integrity	almost certain	gradual	data may not work; may prevent data recovery	error checking, error correction, data refreshing and transfer
Carrier breakdown	data integrity	certain for most carriers	gradual	severe; data may be unreadable and not recoverable	use high quality products; use more stable carriers; check condition frequently; transfer data within expected life of carrier
Malicious attack: hackers, virus, intruders	data integrity, file identity	almost certain for networked archives	likely to be sudden	likely to be severe; may include rewriting or corrupting data	security measures, logical and physical; firewalls, access controls; take data offline
Collateral damage from other attacks not directed at system	data integrity, file identity, equipment assets	varies, depending on situation	likely to be sudden and unexpected	likely to be severe, and beyond capacity of normal security measures	backup data; secure access to backups

Inadvertent acts eg turning off power, discarding carriers, reformatting storage devices	data integrity, file identity	likely unless managed	likely to be unexpected	varies from nuisance to catastrophic	backup data; staff training and physical access controls
Natural disasters eg fire, flood	data integrity, file identity, equipment assets	very likely over long term	likely to be sudden but may be warning period	may be localised and minimal or total loss	disaster preparedness; well placed storage areas; offsite backup
Business failure	access to data	varies	may be gradual or sudden	likely loss of access as carriers are dumped or re-used	business planning and management; continuity/succession arrangements; clear identification of important assets

Table 16-2 *Sample risk analysis of data protection threats*

FOR PRESERVATION PROGRAMMES WITH FEW RESOURCES

16.13 Fundamental responsibilities

Data protection is such a critical responsibility that even programmes with few resources must give it a high priority. The simplified risk analysis above may suggest areas of lower risk for some programmes. It also suggests that some risks may be reduced at the cost of reducing the level or speed of access. This may be perfectly acceptable for some collections.

16.14 Prioritisation

It may be possible to prioritise parts of the collection for additional protection, and to offer lower protection (such as less frequent backups, use of lower quality carriers, less frequent transfer of data to new carriers) to less important data.

CASE STUDIES

16.15 Case study 1

A typical arrangement that makes use of redundancy holds data on tape in three copies: one held ‘near-line’ in a tape library attached to the system, one offline but on site, and one offsite. All copies are identical and the system maintains them so. For access, a temporary copy is made to disks organised as a RAID (Redundant Array of Inexpensive Disks) in which failure of one disk is compensated for by copies on other disks. To achieve carrier redundancy, there may also be a separate copy stored offsite

on optical media.

16.16 Case study 2

A record archive documenting government business transactions goes to great effort to certify the authenticity of every record it stores. All records scanned from non-digital originals include a signed statement attesting that they are true copies; digital records captured from electronic record keeping systems include system-generated verification checks. All processes that could bring about unintended or unauthorised changes are documented in preservation metadata attached to the record.

A regional library collecting digital publications uses quality control checking to ensure that the files it captures match the copy remaining on the publisher's site. It documents the processes it applies to the material, and controls any significant threats to data integrity, but it accepts that some processes will lead to items that differ from their original appearance when re-presented in the future. It is unable to certify that the copies it presents are authentic, but claims its processes provide a reasonable basis for accepting them as archived, managed copies for research purposes.

REFERENCES – **where to look for more information**

Cross references

Relationship between data and presented digital objects also see Understanding digital preservation: chapter 7

Offsite references (all links viewed march 2003)

Some interesting views on authenticity can be found in:

- Gladney Henry M, *Digital Document Quarterly*. http://home.pacbell.net/hgladney/ddq.htm
- Graham Peter S, (2000). *Authenticity in a Digital Environment,* Council on Library and Information Resources. http://www.clir.org/pubs/reports/graham/intpres.html
- InterPARES Project (2002). *The Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project.* http://www.interpares.org/book/index.htm

Technical information on data storage devices can be located through:

- Bogart, John Van. (1995). *Magnetic Tape Storage and Handling: A Guide for Libraries and Archives* Council on Library and Information Resources, Washington, DC. http://www.clir.org/pubs/reports/pub54/index.html
- *CoOL [Conservation OnLine]: electronic storage media.* http://palimpsest.stanford.edu/bytopic/electronic-records/electronic-storage-media/
- Kodak Professional *(nd). Permanence and Handling of CDs.* http://kodak.com/global/en/professional/products/storage/pcd/techInfo/permanence.jhtml
- Library of Congress (rev ed 2002). C*ylinder, Disc and Tape Care in a Nutshell.* http://www.loc.gov/preserv/care/record.html

Chapter 17. Maintaining accessibility

INTRODUCTION

17.1 Aims

This chapter aims to explain the context of access maintenance, and what is required to support it, as well as providing a basis for comparing a range of commonly proposed strategies.

17.2 In a nutshell

Changes in software and hardware add up to a loss in the means of access to digital heritage materials. This is expected to be the core challenge for most preservation programmes. Using understandings of the relationship between digital objects and their means of access, and taking account of what has to be presented to a user in providing access, programme managers must decide on strategies that will guarantee accessibility whenever it is needed. Strategies, which are likely to vary over time and according to needs, are still evolving. The strategies discussed are grouped into those based on investment of resources from early in the life cycle of digital materials, those with short-term and with medium- to long-term effectiveness, and some alternative 'non-digital' and 'non-preservation' strategies.

KEY MANAGEMENT ISSUES

17.3 Why accessibility pathways are needed

Preserving the ability to access digital material is the key purpose of digital preservation programmes. Based on preserved data and metadata, and using access tools of software and hardware, digital objects must be re-presented to users in an understandable form. This must be done at any time in the future when they are needed, using access technologies available at that future time.

Because digital objects rely on specific combinations of technology for presentation, the ability to re-present them at a later date is typically disrupted or lost as technologies change. This phenomenon of changing access technologies is so common that it is almost a defining characteristic of stored digital materials.

17.4 Timeframes for preservation

The rate of technological change brings the horizon of loss close for many currently available digital materials. Some materials created with technologies that were common less than ten years ago are already difficult, if not impossible, to make available with the technologies of today.

While the ultimate goal is to find ways of guaranteeing access at any point in the long-term future, there is also a need to ensure accessibility in the short-term.

17.5 Defining acceptable levels of loss

Preservation programmes are likely to face scenarios requiring judgments about acceptable and unacceptable levels of loss.

Complete fidelity to the original presentation of digital materials may be difficult in any case; many currently reviewed strategies may involve losses including possible loss of content, loss of the original 'look and feel', or loss of some original functions.

These losses may be the unintended by-products of the chosen strategy (common in migrating files to a new format), or the intended result of choices to reduce preservation costs (such as discarding links or dynamic elements of Web pages). They may even be intrinsic to the preservation objectives of the programme (such as the removal of edit functions from documents saved as static records).

In these and similar scenarios, the programme requires some means of deciding what losses will be acceptable.

PRINCIPLES IN ADDRESSING THESE CHALLENGES

17.6 The responsibility of preservation programmes

Preservation programmes must find ways around the threat of changing and obsolete technologies if they are to achieve their primary objective, which is to maintain continuity of access.

17.7 Recognising which items must be preserved

Many collections contain multiple versions of the same materials, such as high quality digital images and their lower quality, derivative versions provided for easy network access. Preservation programmes must decide which version or versions should be maintained, and which can be generated anew at a later date.

17.8 Recognising which elements must be maintained

In order to define acceptable and unacceptable levels of loss, preservation programmes must define the essential elements they must maintain. As discussed previously (in chapter 12), programmes need this information in order to:

- Choose the most appropriate strategy to maintain those elements
- Choose the most cost-effective strategy
- Assess whether their strategy has been successful.

Setting preservation objectives at this level requires careful study of the material to understand why is exists, how it works, and what a user should be able to see and do

with a preserved copy.

Once the essential elements have been defined, the preservation programme's task is to find, and continue to find, combinations of data, software and hardware that will represent those elements as accurately as required.

17.9 The relationship between data and software

There is always a dependent relationship between data and software: all data require some kind of software in order to be presented in an understandable form to a user. The degree of dependency has important implications:

- Some objects are relatively independent of specific software; eg basic data sets, plain or tagged text such as ASCII could be presented using a range of quite basic software tools
- Some objects depend on more complex but generic or widely available software; eg HTML, standard image formats such as TIFF and other formats designed to work on interchangeable platforms
- Some objects depend on specific application software and are not designed to work outside their original operating environment – although manufacturers often provide tools that allow them to be read or used more widely; eg word processing formats, spreadsheets, some databases, drawing and GIS mapping formats
- Some objects essentially *are* software; eg executable files, software programmes
- Many complex materials contain combinations of objects with different levels of software dependency.

The degree of software dependency may limit the choice of strategies that are available. For example, 'data-' or 'document-type' objects may be effectively presented by a range of software, while 'software' objects may have far fewer options for retaining access once their original operating environment has been lost.

17.10 Choosing appropriate strategies

There is, as yet, no universally applicable and practical solution to the problem of technological obsolescence for digital materials. Several approaches have been proposed but it is unlikely that there will be a single solution that offers a cost-effective means of access for all materials, for all purposes, for all time. At this stage, it is reasonable for preservation programmes to look for multiple strategies, especially if they are responsible for a range of materials over extended periods.

It is important to take active steps now, even small ones, which will preserve access for the 'manageable future', while also planning for whatever long-term approaches appear to be the most practical.

The current front-runners as long-term strategies appear to be: the use of *standards* for data encoding, structuring and description that can be expected to remain recognisable

for long periods; *emulation* of obsolete software or hardware in a new environment; and *migration* of data from one operating technology to another. These are all strategies that have been demonstrated to work in certain circumstances over limited periods of time. Necessarily, they have not proven themselves against unknown threats over centuries of change. But they do have current applications in the management of data, and it seems likely that combinations of them will continue to be researched and proposed for large-scale, long-term preservation.

17.11 The principles behind current approaches

In searching for ways to overcome the impact of technological change, most approaches that have been proposed are based on one or more of the following principles:

- Converting data to a human readable form on a carrier that is easy to maintain (such as paper, film or stable metal carriers)
- Creating data in, or converting data to, a highly standardised form of encoding and/or document structure (or file format) that will continue to be widely recognised by computer systems for a long time
- Making the data 'self-describing' and 'self-sustaining' by packaging it with metadata and with links to software that will continue to provide access for some time, (and perhaps even packaging the software with the data)
- Converting the data to a format where the means of access will be easier to find
- Maintaining the data in its original form (or a simplified version), and providing tools that will re-present it as originally, either using the original software and hardware (which have been maintained as well), or using new software that emulates the behaviour of the original software and/or hardware
- Providing specifications for emulating the original means of access on a theoretical intermediate computer platform, as a bridge to later emulation in a future operating environment
- Converting (migrating) the data to new formats that are accessible with each new operating technology
- Supporting later migration on demand by maintaining the data and recording enough information about it to allow a future user or manager to convert it to a then-readable form
- Maintaining the data and providing new presentation software (*viewers*) that will render an acceptable presentation of it for each new operating environment.

17.12 Critical support for preservation strategies

Whatever strategies are chosen, they must be supported by:

- Appropriate organisational commitments of responsibility, policy, procedures and resources

- Appropriate legal clearances
- Protection of the data
- Access to specifications of standards and file formats for reference
- Metadata that establishes the identity, integrity and technical requirements of the material throughout its life
- Attention to quality control issues at all stages
- Monitoring of threats such as impending changes in technology that would indicate re-activation of the strategy is needed.

17.13 Contingency planning

With all strategies, it is good practice to retain and protect the original object data stream, as well as the modified data streams that the strategy may produce. Retaining the original data stream should be seen as contingency planning, providing an opportunity to pursue other strategies should the chosen strategy fail. Such an approach does imply extra costs to manage the additional data, and to manage the relationship between parallel data streams. Despite the costs, the uncertain status of most preservation strategies makes this approach very attractive.

TECHNICAL AND PRACTICAL ISSUES

17.14 Introduction

This section is devoted to a discussion of some of the most commonly proposed strategies. These have been arranged as follows:

- 'Investment' strategies (primarily involving investment of effort at the start):
  - Use of standards
  - Data extraction and structuring
  - Encapsulation
  - Restricting the range of formats to be managed
  - 'UVC' (Universal Virtual Computer) approach
- Short-term strategies (likely to work best over the short-term only):
  - Technology preservation
  - Backwards compatibility and version migration
  - Migration (which may also work over longer periods)
- Medium- to long-term strategies (likely to work over longer periods):
  - (Migration)
  - Viewers
  - Emulation
  - (UVC approach)

- Alternative strategies:
  - Non-digital approaches
  - Data recovery
- Combinations.

17.15 'Investment' strategies

1. Use of standards

Description:

This strategy involves the use of preferably open, widely available, supported or agreed standards and file formats, for which there is an increased likelihood of stability and longer term support. Such standards or formats may either be formally agreed or may be *de facto* standard formats that have been widely adopted by industry. Compliance to standards may also either simplify the application or maximise the effectiveness of later preservation strategies. This strategy can be related to No.4 – Restricting the range of formats to be managed.

A particular refinement of the standards approach is proposed in conjunction with the UVC approach (see below at No. 5), as *durable encoding* (Gladney and Lorie, 2002), which recommends encoding data to confirm with well-known data processing standards down to the level of encoding bits as ASCII or Unicode UTF-8, and objects as XML. For objects that cannot be encoded in this way, programmes that can interpret them can be so encoded and packaged with them.

Examples:

- A majority of digitisation programmes choose TIFF (Tagged Image File Format) as an open, stable and widely supported standard for creation of preservation master images, with expectations of the format's longevity
- The Victorian Electronic Records Strategy (VERS) primarily stores digital documents in Adobe Portable Document format (PDF) and encapsulates them in an XML metadata wrapper. PDF was chosen, in part, due to the public availability of the proprietary standard, from which independent viewing tools have been constructed.

More information: (all links viewed March 2003)

Gibbs R, Heazlewood J (2000). 'Electronic Records – Problem Solved?: the Victorian Electronic Records Strategy and the future of electronic record keeping in Victoria'. In: *Books and Bytes : Technologies for the Hybrid Library : Proceedings, 10th Biennial Conference and Exhibition, 16-18 February, 2000, Melbourne Convention Centre*. Victorian Association for Library Automation, Inc., Melbourne, 2000.

Gladney H, Lorie R (2002). *Trustworthy 100-Year Digital Objects: Durable Encoding for When It's Too Late to Ask*. Saratoga CA, HMG Consulting, 2002. Available, with later relevant papers, from HMG Consulting and via *Digital Document Quarterly*,http://home.pacbell.net/hgladney/ddq_1_4.htm

Some available data interchange standards for various areas of activity are listed in:

The Diffuse Project (2002). *Diffuse Standards and Specification List*. The Diffuse Project Consortium, 2002. http://www.diffuse.org/standards.html

Potential advantages of using standards

- Should simplify the preservation process by slowing the rate of change in the technology required for access; encoding data in very basic standards like ASCII should make it 'readable' by computer systems for a long time
- Widely supported formats may have a range of tools available for interpretation
- Use of available, published standards is more likely to allow re-interpretation of the data or re-construction of tools in the future, if necessary.

Difficulties, disadvantages and risks

- May involve some investment to convert material to standard; may involve losing some elements in conversion; may not be any standardised format available for some types of objects.

Specific requirements

- Knowledge of appropriate standards and ongoing monitoring of standards developments
- Standard file formats need to be well chosen, both with regard to the effect of any transformation on the essential characteristics of objects and the expected longevity of tools to work with these formats.

Indications for use

- Use of standards should be generally encouraged, but particularly when a caretaker organisation has some influence over the creation of materials or the format in which materials may be deposited
- Suitable where open, standard formats are available that can encode the required complexity of the original objects, without unacceptable loss of essential characteristics.

2. Data extraction and structuring

Description:

Data abstraction, sometimes also called *normalisation*, involves analysing and tagging data so that the functions, relationships and structure of specific elements can be described. The re-presentation of content can be liberated from specific software applications and be achieved using different applications as technology changes.

Examples:

- The San Diego Supercomputer Center have used custom algorithms to apply XML tags to a collection of one million emails (Moore et al, 2000 [2]). Application of this approach to word processing documents and geospatial datasets has also been investigated (Moore, 2001). The National Archives of Australia is also investigating this approach to emails, with extension to other formats to follow (Heslop and Davis, 2001)
- The VERS programme of the Public Record Office Victoria (Australia) is investigating XML representation of database tables
- The Universal Virtual Computer approach (Lorie, 2000) proposes the inclusion of tags in original data streams to mark sections of data for interpretation using a documented set of rules for each data type.

More information:

Heslop H, Davis S (2002) (unpublished). *An Approach to the Preservation of Digital Records*. National Archives of Australia, Canberra

Lorie RA (2000). *Long-Term Archiving of Digital Information, IBM Research Report RJ10185*. IBM Research Division, San Jose, California. http://domino.watson.ibm.com/library/CyberDig.nsf/7d11afdf5c7cda94852566de006b4127/be2a2b188544df2c8525690d00517082

Moore R, Baru C, Rajasekar A, Ludaescher B, Marciano R, Wan M *et al*. (2000). Collection-Based Persistent Digital Archives – Part 1. *D-Lib Magazine* 6(3). http://www.dlib.org/dlib/march00/moore/03moore-pt1.html

Moore R, Baru C, Rajasekar A, Ludaescher B, Marciano R, Wan M *et al*. (2000). Collection-Based Persistent Digital Archives – Part 2. *D-Lib Magazine* 6(4). http://www.dlib.org/dlib/april00/moore/04moore-pt2.html

Moore R (2001). Final Report for the Research Project on Application of Distributed Object Computation Testbed Technologies to Archival Preservation and Access Requirements, SDSC TR-2001-8. San Diego Supercomputer Center. http://www.sdsc.edu/TR/TR-2001-08.doc.pdf

Potential advantages of data extraction

- Infrastructure independence simplifies the transport of data between platforms and over generations of technology.

Difficulties, disadvantages and risks

- Not all object types can be abstracted in this way
- Requires extensive development of tools and methods for analysis and processing in order to correctly represent and tag each type of data
- The technology eventually used for presentation may still limit what functions can be represented.

Specific requirements

- Appropriate tools to tag and transform data
- A high degree of quality control during the development of methods to ensure that all semantic relationships and anomalies are represented.

Indications for use

- Structured or semi-structured data or documents for which retention of content, semantics and relationships is more important than any particular display characteristics.

3. Encapsulation

Description:

Encapsulation is a widely adopted means of binding together data and the means of providing access to it, preferably in a 'wrapper' that describes what it is in a way that can be understood by a wide range of technologies (such as an XML document). Because it is often impractical and unnecessary to encapsulate the actual means of access such as software and hardware, encapsulation usually bundles metadata describing or linking to the correct tools. An alternative approach is to include or link to a specification for the software or hardware so that it could be rebuilt in the future if necessary.

Examples:

- The Reference Model for an Open Archival Information System (OAIS) describes incorporating data objects and their associated metadata into Archival Information Packages (AIPs). Metadata may either be bundled directly with the archived object or logically associated within the system
- The VERS strategy involves creation of "onion records", in which data objects are wrapped directly in XML-encoded metadata, making them independent of a management system
- The Universal Preservation Format (UPF) strategy seeks to make objects independent of applications and operating systems by wrapping the content in "self-describing" metadata that includes the technical specifications to access the encapsulated materials.

More information:

Consultative Committee for Space Data Systems (2002). *Reference Model for an Open Archival Information System (OAIS)*. CCSDS 650.0-B-1. Blue Book. Issue 1. January 2002. Washington D.C., CCSDS Secretariat, 2002. http://wwwclassic.ccsds.org/documents/pdf/CCSDS-650.0-B-1.pdf

Gibbs R, Heazlewood J (2000). Electronic Records – Problem Solved?: the Victorian Electronic Records Strategy and the future of electronic record keeping in Victoria. In: *Books and Bytes : Technologies for the Hybrid Library : Proceedings, 10th Biennial Conference and Exhibition, 16-18 February, 2000, Melbourne Convention Centre*. Victorian Association for Library Automation, Inc., Melbourne, 2000.

Shepard T, MacCarn D (1999). *The Universal Preservation Format: A Recommended Practice for Archiving Media and Electronic Records.* WGBH Educational Foundation, Boston. http://info.wgbh.org/upf/pdfs/991231_UPF_RP.pdf

Potential advantages of encapsulation

- Provides information that will make it easier either to find a current means of access or to develop one.

Difficulties, disadvantages and risks

- Providing a link to a current means of access does not really address the basic problem of technological change

- May be too difficult to find or build a replacement means of access even with the encapsulated information.

Specific requirements

- Detailed knowledge of the technical requirements for access
- Secure bundling of the package so that data and metadata are not separated
- Metadata describing the means of providing access must be kept up to date
- A self-describing layer such as an XML wrapper very desirable.

Indications for use

- Probably should be seen as a basic good practice for all objects that may facilitate other strategies.

4. Restricting the range of formats to be managed

Description:

Preservation programmes may decide to only store data in a limited range of formats. This can be achieved either by only accepting material already in those formats, or by converting material from other formats before storage.

Examples:

- The UK Archaeology Data Service (ADS) specifies a preferred (but not exclusive) range of formats for deposit and provides guidelines for depositors on creating or preparing materials for submission
- In prescribing the types of records that must be maintained by contributing institutions, government archival bodies may also be able to prescribe the formats that they will accept for deposit.

More information:

Archaeology Data Service (2001). *Guidelines for Depositors, Version 1.1*. Archaeology Data Service, York. http://ads.ahds.ac.uk/project/userinfo/deposit.html

Potential advantages of restricting formats

- Reduces the range of problems needing to be managed
- May be used as a refinement of the standards approach, in which case it offers the benefits of that approach as well.

Difficulties, disadvantages and risks

- Does not necessarily solve the access problem unless the formats used are effective through some other strategy
- May restrict the range of materials the programme will accept
- Conversion may cause some loss of essential elements.

Specific requirements

- A basis for deciding what formats will be accepted and how to deal with submissions that do not comply

- Either clear submission rules or conversion software to migrate data
- Rigorous quality control checking.

Indications for use

- Reasonably straightforward, easily standardised materials
- Collections with large numbers of uniform items.

5. 'Universal Virtual Computer' approach

Description:

The Universal Virtual Computer (UVC) approach seeks to specify an intermediate platform, a virtual machine, which is general, but may be completely and accurately defined. UVC operations are simple enough to be re-implemented from the specification at any time in the future on an available platform.

For object preservation, at the time of archiving a logical schema representing a data type is developed, along with a decoding programme that is capable of interpreting the object according to this schema. The decoding programme is written for execution by an implementation of the UVC.

At the time of object restoration, an emulator for the defined UVC is implemented on an available platform. The UVC executes the archived decoder programme, which interprets the archived object, and passes the results to a restore programme, which restores a representation of the object according to the archived logical schema.

Examples:

- The proof-of-concept prototype for the UVC approach (Lorie, 2002) has been used to produce a logical schema, decoder programme and representation mechanism for PDF documents, such that the document content can be represented using a UVC interpreter and restore programme.

More information:

Gladney H, Lorie R (2002). *Trustworthy 100-Year Digital Objects: Durable Encoding for When It's Too Late to Ask*. Saratoga CA, HMG Consulting, 2002. Available, with later relevant papers, from HMG Consulting and via *Digital Document Quarterly*, http://home.pacbell.net/hgladney/ddq_1_4.htm

Lorie R (2002). *The UVC: a Method for Preserving Digital Documents – Proof of Concept*. Amsterdam, IBM Netherlands, 2002. http://www.kb.nl/kb/ict/dea/ltp/reports/4-uvc.pdf

Lorie RA (2000). *Long-Term Archiving of Digital Information, IBM Research Report RJ10185*. IBM Research Division, San Jose, California.
http://domino.watson.ibm.com/library/CyberDig.nsf/7d11afdf5c7cda94852566de006b4127/be2a2b188544df2c8525690d00517082

Potential advantages of UVC approach

- May provide options for preserving the behaviour of both document-type materials and software programmes
- A single, defined intermediate platform may reduce the development work required to accommodate different software and platform combinations
- The specification of the UVC is intended to be simple, allowing use by

programmers of average competence and possibly simplifying construction of UVC interpreters or emulators in the future
- May be designed to interpret the original object data stream, or a transformed or abstracted representation
- Data encodings and decoder programmes could be tested at the time of creation on a contemporary UVC implementation. Future implementations of the UVC specification could then be expected to reproduce the current behaviour.

Difficulties, disadvantages and risks
- The approach is currently in development and has been prototyped for a transformed representation of an original document. Further work is required to apply the approach to software programmes. As with emulation (see No. 11), the complexity of programme behaviours may be problematic
- Investment required at time of archiving in development of encoding methods or UVC-native interpretive programmes for each data type
- May require substantial support from information producers to provide UVC-compatible versions of their products (Gladney and Lorie, 2002), for which they may have little incentive or business case
- Investment required at restoration time in developing a UVC emulator and restore programmes
- If original data objects are abstracted or transformed for encoding purposes, such transformation may discard essential characteristics.

Specific requirements
- Development of a logical schema or representation for each data type or programme at the encoding stage.
- Development at the encoding stage of a decoder programme to interpret each data type or programme, written for execution by a UVC interpreter or emulator.
- Development of a UVC interpreter or emulator at the time of object restoration, to suit a prevailing platform.
- Development of restore programmes to return a representation of the original object, based on the logical schema and data retrieved by the UVC in executing the archived decoder programme.
- Archiving of the data object or programme (or its transformed representation), any associated logical schemata, the UVC-executable decoder programme, and the UVC specification and restoration instructions.
- Sufficient expertise for development of logical schemata, encoding, decoder programmes, UVC emulator implementation from the specification and restore programmes.

Indications for use
- May be suitable where objects may be sufficiently represented, encoded, interpreted and restored using tools developed from the UVC specification. At the time of writing, the UVC specification is under development.

17.16 Short-term strategies

6. Technology preservation

Description:

This strategy involves keeping and maintaining the original software and hardware with which digital objects were presented. It is the most basic, and in some ways the most important first step in preserving access if no other strategy is in place. If the hardware and software required for access are discarded before other strategies are available, it may be effectively impossible to provide later access without expensive and uncertain data recovery work.

Examples:

- Maintaining old disk drives that will accommodate diskettes of a size that are no longer accommodated by current computer equipment
- Maintaining superseded software for use with legacy materials
- Maintaining old operating systems to support software that does not work on current platforms.

More information:

Jones M, Beagrie N (2001). *Preservation Management of Digital Materials: A Handbook*. The British Library, London, 2001.

Potential advantages of technology preservation

- Presenting digital objects through their intended hardware and software ensures the full range of intended elements and functions are presented
- Provides a period in which alternative strategies may be developed or implemented
- As a side benefit, documenting the hardware and software that needs to be kept may lead to a better understanding of the collection and its dependencies, which is likely to be useful information for planning and implementing other strategies.

Difficulties, disadvantages and risks

- Long-term maintenance of equipment, with increasingly scarce parts and expertise, is very unlikely
- Even with active management, the window of access using this approach may be as narrow as five to ten years from the time the original format is superseded. (However, that may be much better than losing access immediately)
- Requires the management and maintenance of a wide range of equipment and software, along with ancillary materials such as manuals and licences, which may be difficult and expensive to achieve
- The necessary expertise and technical support may simply not be available.

Specific requirements

- Requires active identification of hardware and software needed for access
- Requires active and ongoing maintenance arrangements for equipment and preservation and licence arrangements for software
- Requires steps to ensure expertise is shared and is not dependent on one person
- It may be possible for a number of organisations to pool superseded equipment or parts, and to use shared or third party software archives
- As a matter of principle, if the required access software is available, it should be sought and retained at least until another strategy has been put in place.
- Retained software should be treated like any other digital objects, requiring control, documentation, media refreshing and maintenance, subject to copyright requirements.

Indications for use

- Recommended as an initial strategy for all preservation programmes, in the absence of longer term strategies or while they are being developed
- May be the only available option for a longer period for complex digital objects such as software and multimedia objects
- Recommended for software required to support of a range of other strategies.

7. Backwards compatibility and version migration

Description:

This strategy relies on the ability of some software to interpret and present objects created with previous versions of the same software. In the case of backwards compatibility, the presentation may be limited to temporary viewing, whereas version migration permanently converts documents into a format that can be presented by the current version of the software.

Examples:

- Web browsers are usually capable of interpreting and displaying material written using earlier versions of the HTML standard
- Office document applications, such as word processing, spreadsheet and database applications, usually allow previous versions of their file formats to be transformed and resaved in a new version, as part of application upgrade paths
- The Digital Preservation Testbed (Digitale Testbed Bewaring) project in the Netherlands has investigated migration of documents through and over generations of application versions.

More information:

Potter M (2002). Researching Long Term Digital Preservation Approaches in the Dutch Digital Preservation Testbed (Testbed Digitale Bewaring). *RLG DigiNews* 6(3). http://www.rlg.org/preserv/diginews/v6-n3-a2.html

Potential advantages of backwards compatibility

- Availability: software developers often build in a suitable backwards compatibility or migration path for documents
- May extend the period before more extensive transformation or treatment is needed
- In some cases offers functionality similar to the original presentation.

Difficulties, disadvantages and risks

- It is unlikely that compatibility will be retained over many generations of the software
- Likely to introduce unwanted changes incrementally if used over many generations
- Such paths may not be available for all types of objects
- May be abandoned by software developers for any new generation of their software, so reliability may be unpredictable
- Even between nearest versions of the same applications, there may be unwanted changes introduced to the materials.

Specific requirements

- As with any migration step, quality control checking of migrated documents is required to detect any unacceptable changes

Indications for use

- May provide a simple, short-term migration path for document-type objects in formats that offer a succession of versions, so long as conversions do not introduce unwanted changes
- May be an alternative to technology preservation for objects such as complex spreadsheets and databases, for which no alternative strategies are yet available.

8. Migration

Description:

Migration involves transferring digital materials from one hardware or software generation to another. As distinct from refreshing, which maintains the data stream by transferring it from one carrier to another, migration entails transforming the logical form of a digital object, so that the conceptual object can be rendered or presented by new hardware or software.

There are a number of strategies that can be considered as forms of migration, differing in the time when transformation happens and in the types of objects transformed. The most commonly proposed migration method involves permanently transforming one logical format into another in line with technological change, so that all migrated objects can be presented with prevailing technology.

It is also possible to propose a 'migration on demand' or 'migration at the point of access' model. This approach is discussed under No. 10 ('Viewers') below.

NB. Because of the likely cumulative effects of repeated migrations, this approach has been included amongst short-term strategies. However, for some data and format types it is likely that migration may prove to be a useful long-term strategy.

Examples:

- Collections of heterogeneous materials in well-characterised formats, such as image collections, are likely to be suited to format transformation.

More information:

Lawrence GW, Kehoe WR, Rieger OY, Walters WH, Kenney AR (2000). *Risk Management of Digital Information: A File Format Investigation*. Council on Library and Information Resources, Washington, D.C. http://www.clir.org/pubs/reports/pub93/contents.html

National Archives of Australia. *Managing Electronic Records – Appendix 3: Preserving Electronic Records through Migration*. National Archives of Australia, Canberra. http://www.naa.gov.au/recordkeeping/er/manage_er/append_3.html

Task Force on Archiving of Digital Information (1996). *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information.* Commission on Preservation and Access and Research Libraries Group. ftp://ftp.rlg.org/pub/archtf/final-report.pdf

Potential advantages of migration

- Simple migration procedures are well established for some formats
- Migrations carried out in response to changes in technology allow the migrated objects to be checked against unmigrated copies to see whether essential elements have been retained
- If the migration has worked, users can confidently expect the material to be presented with prevailing technology, without the need for special hardware or software.

Difficulties, disadvantages and risks

- It may not be possible to provide access to some materials such as complex objects using format migration, because there may be no way of representing complex functions in the new format
- While apparently working, transformation of the logical encoding may compromise the integrity or essential elements of the material
- Objects will need to be transformed regularly to keep pace with technology, creating an ongoing cost burden. Large-scale migrations involve detailed analysis of data structures, development of rules to control the transformation, writing programmes to change the data encoding, and extensive quality control and 'cleaning up'. This may be easily justified for large, business critical databases but such rigour may not be feasible for less critical materials in a diverse range of file formats
- Small changes between generations may accumulate into major alterations or losses as a result of repeated migrations.

Specific requirements

- Requires programmes and tools to carry out the conversion
- Rigorous quality control checking, both while methods are being developed and after migration
- Documentation of the migration method should be stored in metadata, as part of object history and authenticity
- If possible, migration processes should be made completely reversible by documenting the nature and location of all transformations
- Alternatively, a copy of the source digital objects should be retained if a transformation is not reversible or if some essential elements may be lost. (Retaining a copy of the original format is good practice in any case)
- The migration process should be tested before full implementation, and its success established before destroying any intermediate generations.

Indications for use

- Migration is likely to be suitable for a range of digital objects, particularly document and dataset types of object
- Where the essential elements to be preserved are reasonably straightforward and do not depend on the look and feel of the material, and do not involve executable files
- May be most cost-effective for homogeneous collections such as digital image and audio collections that are in very widely used, well-standardised, non-proprietary formats
- Some widely used proprietary formats may also be suitable if patent

owners or licensers either supply or allow others to develop format specifications or conversion tools.

9. Re-engineering

Description:

Being highly dependent on a specific system or platform in order to function, software objects are perhaps the most affected by changes in technology and are also usually unsuited for many preservation strategies, including regular migration. Software re-engineering may offer several strategies for transforming software as technologies change, similar to transformation of data formats. Some possibilities include:

- Adjustment and re-compiling of source code for a new platform
- Reverse-engineering of compiled code into higher level code and porting that to the new platform
- Re-coding of the software from scratch, or re-coding in another programming language (Wheatley, 2001)
- Translation of compiled binary instructions for one platform directly into binary instructions for another platform. (Researchers at the University of Queensland (Cifuentes et al, 1999) are investigating this concept.)

More information:

Cifuentes C, Van Emmerik M, Ramsey N (1999). The Design of a Resourceable and Retargetable Binary Translator. In: *Proceedings: Sixth Working Conference on Reverse Engineering, October 6-8, 1999, Atlanta, Georgia, USA*. IEEE Computer Society, New Jersey, 1999, pp 280-291

Wheatley P (2001). Migration – a CAMiLEON discussion paper. *Ariadne* 29. http://www.ariadne.ac.uk/issue29/camileon/

Potential advantages of re-engineering

- Has the potential to port software objects from one platform to another.

Difficulties, disadvantages and risks

- Except for open source programmes and software developed in-house, source code is often not available or within rights to use
- Even when source code is available, porting to other platforms is not trivial, and in general, compilers or interpreters are required for the new platform for the code language
- Requires considerable time and effort per object
- Any form of reverse engineering is usually strictly prohibited by end user license agreements and seriously violates intellectual property rights. Other forms of transformation may also infringe such rights.

Specific requirements

- A high level of expertise
- Tools to transform human-readable code into machine-readable code
- Explicit permission to reverse-engineer.

Indications for use

- Should only be contemplated where appropriate rights are expressly granted, and when expertise, tools and, preferably, source code are available.

17.17 Medium- to long-term strategies

10. Viewers and migration at the point of access

Description:

A number of alternatives to recurring, incremental migration have been proposed, involving the use of viewers, software tools or transformation methods that provide accessibility at the time of access, using the original data stream.

Examples:

- The 'migration on request' approach developed in conjunction with the CEDARS and CAMiLEON projects includes a software tool with the digital object and uses the object's metadata to record a method for accessing the object using the tool. As technology changes, the metadata is updated to reflect changes in the method of access (Cedars, 2002; Mellor, Sergeant and Wheatley, 2003).
- The TOMS (Typed Object Model Server) approach provides transformation methods for common document and data types, allowing a server to choose a suitable transformation path for a range of object types. (Thibodeau, 2002)
- The VERS strategy converts documents to a PDF format on the basis that third-party viewers for PDF may be constructed from the format specification.
- The Rosetta Stones approach includes methods for data format interpretation and sample files in both the original format and a reference format showing what the files should look like if being interpreted correctly. Software tools may then be constructed to follow the interpretation method proposed for the files, and to check for correct interpretation by comparing sample files against the reference display. (Thibodeau, 2002)

More information:

Cedars Project (2001). *The Cedars Project Report, April 1998 – March 2001*. Cedars, University of Leeds.
http://www.leeds.ac.uk/cedars/pubconf/papers/projectReports/CedarsProjectReportToMar01.pdf

Cedars Project, (2002). *Cedars Guide to: Digital Preservation Strategies*. Cedars, University of Leeds.
http://www.leeds.ac.uk/cedars/guideto/dpstrategies/dpstrategies.html

Mellor P, Sergeant D, Wheatley P (2002). *Migration on Request: A Practical Technique for Preservation*. CAMiLEON Project, University of Michigan.
http://www.si.umich.edu/CAMILEON/reports/migreq.pdf

Thibodeau K (2002). Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years. In: *The State of Digital Preservation: An International Perspective – Conference Proceedings, Documentation Abstracts, Inc., Institutes for Information Science, Washington, D.C., April 24025, 2002*. Council on Library and Information Resources, Washington, D.C.
http://www.clir.org/pubs/reports/pub107/thibodeau.html

Potential advantages of using viewers, etc

- The original data stream is interpreted and presented by the viewer, tools or transformation method, rather than an incrementally migrated data stream, so the risk of cumulative distortions of content or function over

generations of migration may be avoided
- Objects are only interpreted or transformed when they are accessed, so the cost of regularly migrating objects regardless of access demand is avoided.

Difficulties, disadvantages and risks
- There may not be viewers or tools available for complex materials including executable files
- Viewers may be able to represent some, but not all, elements of some materials (although this may be an advantage where 'view-only' functionality is required)
- The gap between the original format and the prevailing technologies at the time of access may be too great for the tools or methods to cope with
- Viewers, tools or methods, and corresponding metadata must also be maintained or adjusted as technologies change
- If not demonstrated beforehand, there is a risk that viewers, tools or methods may not present the conceptual objects satisfactorily.

Specific requirements
- Thorough documentation of file formats and transformation methods must be kept up to date
- Extensive upkeep of technical metadata in response to technology changes
- Technical metadata and methods for access should be linked but stored separately from the digital objects so that the metadata or methods can be updated centrally.

Indications for use
- May be preferred to recurring migration where the cost of repeated migrations is an issue or where there are likely to be long gaps between access requests
- May be suitable where it can be demonstrated in advance that it is feasible to construct tools or viewers that will interpret file formats from included instructions, specifications or methods.

11. Emulation

Description:

Emulation involves using software that makes one technology behave as another. In the long-term digital preservation context, this would entail making future technologies behave like the original environment of a preserved digital object, so that the original object could be presented in its original form from the original data stream.

Hardware emulation is often proposed as a widely applicable strategy, as hardware specifications are likely to be more complete or easily defined than software specifications. Emulation of a hardware platform also offers good leverage, in that it would allow a range of systems and digital objects to operate, thus solving the problem for a very wide range of digital objects. Alternatively, emulation of specific software applications or behaviours may be considered. One argument against this is that individual emulation efforts would be required for each application; on the other hand, if the need for emulation is small, it may be overkill to expend effort in emulating an entire platform or system for a small number or range of objects.

Examples:

- Researchers from the CAMiLEON project have investigated emulation as a digital preservation approach, including experimental use of available emulators (Hedstrom and Lampe, 2001) and construction of an emulator for a 1970s system, George3 (Holdsworth and Wheatley, 2001)
- A Universal Virtual Computer (UVC) has been proposed as an intermediate virtual platform that could be used across future systems, so that emulation of programme behaviour can be targeted to a single persistent platform, minimising the need for additional layers of emulation (Lorie, 2000)
- The possibility of postponing emulator construction until required, preserving instead detailed specifications for such emulators that would be generated when they were needed, has also been proposed (Rothenburg, 2000).

More information:

Hedstrom M, Lampe C (2001). Emulation vs. Migration: Do Users Care? *RLG DigiNews* 5(6). http://www.rlg.org/preserv/diginews/diginews5-6.html#feature1

Holdsworth D, Wheatley P (2001). Emulation, Preservation and Abstraction. *RLG DigiNews* 5(4). http://www.rlg.org/preserv/diginews/diginews5-4.html#feature2

Lorie RA (2000). *Long-Term Archiving of Digital Information, IBM Research Report RJ10185*. IBM Research Division, San Jose, California. http://domino.watson.ibm.com/library/CyberDig.nsf/7d11afdf5c7cda94852566de006b4127/be2a2b188544df2c8525690d00517082

Rothenberg J (2000).Using Emulation to Preserve Digital Documents. Koninklijke Bibliotheek, The Hague. http://www.kb.nl/kb/pr/fonds/emulation/usingemulation.pdf

Potential advantages of emulation

- Emulation is an established principle in computer science, and is often used for developing and testing new software before production
- Emulators do currently exist for various platforms and systems, ranging from emulators for obsolete systems constructed by enthusiasts, to commercial systems for cross-platform use or testing of software
- In its widest possible application, emulation would allow a range of digital objects to be recreated with full functionality, including software objects, using the original, untransformed data stream in combination with original preserved software.

Difficulties, disadvantages and risks

- Emulation is technically complex, requiring a high degree of effort and specific expertise so it is likely to be very costly
- As a widely applicable digital preservation technique, emulation is still in the research stage
- Effective emulation could be frustrated by inadequate documentation of software, or by non-standard use of file formats such as 'workarounds'
- As systems become more complex, so will the requirements for emulation, which may need to include multiple components. Emulation of all aspects of a system or application may not be possible
- It may be difficult for future users to know how to interact with a wide range of archaic applications operating under emulation, so contemporary presentation tools will probably still be needed, adding a further layer of changing software tools required to access the emulated object
- As technology and platforms change over time, emulators themselves will either have to migrate to, or have their host systems emulated on, the new platform, potentially leading to layers upon layers of emulators.

Specific requirements

- A sufficient level of expertise to develop emulators, or access to emulators developed by someone else
- Thorough, accurate documentation of the systems to be emulated
- Clarity about the level of emulation required, e.g. full hardware emulation vs. specific software functions
- Emulator code should be produced using standard software engineering techniques, including good code structure and thorough commenting and documentation
- Code for the emulation programme should be written in open source, in a standard programming language with good prospects for longevity and future compatibility
- Any non-standard code required (e.g. for specific peripheral functions) should be written as a separate but connected module, and well documented.

Indications for use

- Where suitable emulators are already available for the required platforms
- Where sufficient expertise is available for emulator construction

- For very complex objects or those such as executable software, which may only work with specific systems or hardware
- For objects for which alternative paths such as migration do not work
- For objects whose value relies on being viewed in their original environment.

17.18 Alternative strategies

12. Non-digital approaches

Description:

An alternative to digital preservation methods is to 'print out' the objects onto relatively stable analogue media, such as paper, microfilm or even nickel plates (as with HD-Rosetta technology, which micro-engraves document images on to nickel with an ion beam and allows viewing with optical magnifiers), shifting the preservation burden to an analogue copy in place of the digital object.

Examples:

- An institution has custody of an early database in an obsolete proprietary format which will be unreadable in the next system upgrade and for which there is presently no way to extract or migrate the contained data. The institution chooses to print the entire contents of the database to paper as individual records to preserve at least some access to the data, though inefficient. The institution also retains the database in digital form, in the event that an access mechanism becomes available.
- A digitisation programme creates Computer Output Microfilm from their digital image masters as a physical back-up to the collection and an alternative source for preservation, distribution and access.

More information:

Hedstrom M, Lampe C (2001). Emulation vs. Migration: Do Users Care? *RLG DigiNews* 5(6). http://www.rlg.org/preserv/diginews/diginews5-6.html#feature1

Norsam Technologies, (2001), *HD-Rosetta Archival Preservation Services*. http://www.norsam.com/hdrosetta.htm

Potential advantages of non-digital methods

- Objects are captured in human-readable form and are removed from the threat of technological obsolescence and the pressure of ongoing digital preservation cycles
- Provides a simpler preservation alternative, as analogue materials may be preserved for the long-term using traditional preservation methods
- Likely to involve a once-only conversion cost.

Difficulties, disadvantages and risks

- Likely to lose advantages afforded by digital technology such as convenience of use or storage efficiency
- Loses typical functionalities supported by digital technology, such as

spreadsheet calculations, embedded sound or moving images (although some of these could be saved to separate analogue form as well), or search and navigate functions. (This may not be a disadvantage if these elements did not need to be preserved anyway)

- May not completely remove the threat of technological obsolescence as reader technology for some formats may change over time. Even though it may always be theoretically possible to use optical magnifiers to read the information, this may be impractical, making some material effectively unusable for most users. For sound recordings, analogue access may be more unreliable than digital
- The long-term stability of analogue carriers may depend on expensive storage environments that prove to be less reliable than well-managed computer systems based on high levels of redundancy.

Specific requirements

- Carrier materials used for conversion to analogue should be of archival quality and be stored under archival conditions
- Retention of digital objects is still recommended, where possible, in the event that a suitable access pathway is developed in the future.

Indications for use

- Only suitable for objects that do not require the functions of digital technology to achieve their purpose, e.g. textual, image or data type documents that require no functionality above 'flat' display
- May be a pragmatic step while other strategies are being developed
- May be required as a last resort where no other strategy is available and such limited accessibility is better than no accessibility at all.

13. Data recovery

Description:

Data recovery (sometimes referred to as *data archaeology*) usually involves recovering data as bits from physical media followed by steps to restore the intelligibility of the recovered data. It is most often employed in recovery of data from failed, damaged or degraded media, but methods to restore intelligibility have been used to rescue documents in obsolete formats. However, to assume that one will be able to carry out such rescue in the future is a very unreliable and high-risk substitute for an active preservation programme now.

Examples:

- The UK Archaeology Data Service carried out data recovery of discs from the Newham Museum Archaeological Service (Dunning, 2001). A number of files were found to be corrupted and not recoverable. For those that were recovered, many were in obsolete data formats that required specialist software for interpretation, or were inadequately documented, such that the context of the data could not be satisfactorily established.

More information:

Ross S, Gow A (1999). *Digital Archaeology: Rescuing Neglected and Damaged Data Resources*. Library Information Technology Centre, South Bank University, London. http://www.ukoln.ac.uk/services/elib/papers/supporting/pdf/p2con.pdf

Woodyard D (2001). Data Recovery and Providing Access to Digital Manuscripts. Paper presented at Digital Dancing: New Steps, New Partners - Information Online 2001, Tenth Exhibition and Conference, 16-18th January, 2001, Sydney Convention and Exhibition Centre, Darling Harbour, Sydney. http://www.nla.gov.au/nla.arc-14099-20020211-www.csu.edu.au/special/online2001/papers/digital_issues_iia.htm

Dunning A (2001) *Excavating Data – The Retrieval of the Newham Archive*. Arts and Humanities Data Service. http://ahds.ac.uk/newham.pdf

Potential advantages of data recovery

- May allow recovery of data that would otherwise be permanently lost.

Difficulties, disadvantages and risks

- There is no guarantee of recovery from media, nor recovery of data intelligibility
- Without sufficient documentation, data interpretation is often a 'best guess' and identity, integrity and context are difficult to establish
- Often expensive, with considerable effort required per item
- Without sufficient documentation, it is impossible to judge beforehand whether the effort and expense will be justified.

Specific requirements

- Greatly assisted by good documentation of the file types and content
- May require specialist forensic data recovery services or recognition software

Indications for use

- Recommended for use only as a data recovery and restoration strategy in the event of media damage, or where obsolete media or file formats are found and where the value or importance of the data is likely to warrant the potential costs.

17.19 Combination strategies

As previously noted, for a diverse collection a number of strategies may be necessary to cover the range of objects and characteristics to be preserved; different approaches may also articulate well with each other over time. Preservation programmes should also consider the potential benefits of redundancy in pursuing more than one strategy: even with good planning, a single strategy may fail leaving the programme with no means of access. Several examples noted above use more than one approach; for example:

- Standards such as TIFF for image collections are often chosen in preparation for eventual migration to other standard formats over the long term
- The VERS strategy couples the use of standards (PDF, XML) to the future use of viewers and the likely migration of XML encoded metadata in the future
- Persistent archives (Moore, 2001) use data abstraction with the view to eventual migration – migration of the data, the mark up system and the supporting software, and upgrading of hardware
- The Universal Virtual Computer (UVC) approach combines data abstraction with rules for migration of data objects at the point of access, and an emulation approach for software objects. The 'durable encoding' approach adds the use of fundamental standards for encoding data, including encoding that could be understood by the UVC.

FOR PRESERVATION PROGRAMMES WITH FEW RESOURCES

17.20 Choosing low cost options

Preservation programmes with few resources may have to limit the amount of material they have to manage. With regard to the strategies used for providing access:

- It may be possible to adopt a minimal access approach, storing data safely with good documentation about the original means of access, trusting that at some future stage it may be possible to use that information to devise a means of access
- It may be possible to identify priority material that could be migrated to a format providing at least some level of access, while storing the original for

later preservation work that may offer fuller accessibility

- Insisting on accepting material in only a few very well standardised and widely used formats may greatly reduce workloads and special tools needed to provide access. It may even make it possible to migrate some material forward satisfactorily with consumer grade tools.

CASE STUDIES

17.21 Some possible strategies for different data types (for discussion)

- Datasets – standardised encoding; metadata describing structure; may migrate but expect data to remain readable for a long period without further action
- Databases – capturing data and documenting structure by data extraction; capturing software used to interrogate data; capturing interface and snapshots of query results; migration of data to new database structure and user interface
- Image and sound files – use of standards, including attention to things likely to cause complications such as compression; migration to new standard format
- Text files – encoding (ie migrating to standard encoding and standard XML structure); possible printing out; migrating to new format
- HTML files – sorting into kinds of formats and migrating as browser standard changes
- Software and software-based materials – technical preservation; emulation; re-engineering
- Emails – data extraction and standardised structuring; migration when necessary
- Office records – viewers; data extraction and format normalisation

---

REFERENCES – where to look for more information

Cross references

Essential elements also see Deciding what should be kept: chapter 12

The relationship between data and software also see Understanding digital preservation: chapter 8

Offsite references

The *Preserving Access to Digital Information (PADI)* website provides a comprehensive and updated set of references for studying strategies for preserving accessibility. It is available at http://www.nla.gov.au/padi/

Chapter 18. Some starting points

INTRODUCTION

18.1 Aims

It is difficult to set up fully-fledged preservation programmes from scratch, even when well resourced. It is a daunting prospect for those with very limited resources. The purpose of this chapter is to suggest some starting points, both in overview and in response to a number of possible hypothetical scenarios. This information is offered in a spirit of stimulating thought and discussion, as each programme's situation is different, requiring individual and often imaginative responses.

KEY MANAGEMENT ISSUES

18.2 Some beginning steps

The following steps may help in setting up a preservation programme:

1. Determine what kind of materials you are responsible for, or what kind of materials you are interested in preserving.

2. Liaise with others who have similar interests or responsibilities to see if a cooperative approach is possible.

3. Liaise with others who have experience in preserving or at least managing the kind of materials you are interested in, and seek their guidance and mentoring.

4. Try to work out who creates the material you are interested in; who publishes, distributes or holds the material, and what interest or capability they might have in preserving it for at least a defined period.

5. Try to work out who the potential and current users of the material are, and how they wish to use the material.

6. It may be too difficult at this stage to identify the essential characteristics that must be maintained, but it is important to try to determine the level of functionality you want to keep, eg whether users need to be able to interact with material and modify it, or simply to view it in a read-only form.

7. There appear to be two widely used models for taking the first practical steps:

- Start small with a modest amount of material, possibly limited to relatively straightforward and 'plain' material, with the aims of providing the best level of preservation you can within those constraints and learning as you go, with a commitment to build up policy, objectives, expertise and infrastructure from there.

- Seek to conceptualise the whole programme and how all challenges will be solved, before starting.

Both approaches have problems and benefits, but these guidelines recommend the former approach for people with limited resources who must deal with pressing needs. Such an approach can hardly claim to offer comprehensive and reliable preservation, but it may offer preservation that can develop comprehensiveness and reliability over time.

8. Develop at least basic policies that that will guide the early commitments that you make.

9. Identify the most pressing threats that require immediate attention to prevent valuable material from being lost. (It may be necessary to accept that some materials will be lost, and to focus on saving at least some of the most important material.)

10. Identify any immediate steps that you might need to take that will enable you to deal with the threats, such as getting a better understanding of the material you will have to deal with, or establishing contact with creators.

11. Identify resources – people, expertise, funds, equipment, time – that could be committed to the task of dealing with the threats.

12. Identify actions you could take, especially simple steps that can be implemented quickly, that would either buy some time, or if you have enough resources and support, would allow you to embark on a more ambitious preservation programme. (Some examples of pressing threats and possible actions can be found later in this chapter.)

13. Work out the rights or permissions you would need in order to take this action,

14. Sort out permissions, either by clarifying existing rights, or by asking rights owners for permission to do what you are about to do.

15. Plan and take the action you have decided on, and evaluate it at every step.

16. Discuss with creators how they create materials, offering them advice and information on practices they could follow that would make preservation easier.

17. Review what you have done and decide whether it is sustainable, extendable, or not feasible; in need of further development or severe modification.

SAMPLE SCENARIOS

18.3 Scenario 1

For Web publications, there may be a strong likelihood that they will be changed or removed from the Web without being saved.

Possible actions in response:

- Develop criteria for selecting material that is worth saving
- Contact publishers and discuss their plans, and yours
- If the material appears to be at risk of loss, and you are able to get approval,

copy the relevant files to a local computer, check the quality of the transfer, document what you have copied and how you copied it, and start looking after the data stream, making backup copies and storing them in a safe place offsite

- On the basis of the undertakings given to the publisher, decide whether it is appropriate to make the material available for public access. If it is, you will need an interface of some kind that allows users to find it and to understand what they are seeing
- This is a very short-term arrangement that will eventually need the support of systems to search for material, manage negotiations with publishers, capture and download a range of materials, record and manage appropriate metadata, manage access conditions, look after the ongoing maintenance of the data streams, identify the essential characteristics of the material that must be maintained when the current access technologies do not work, and find ways of representing the material which is likely to include complex multimedia and executable objects. Systems and arrangements to do all of this require very significant investments of time and money, and it may take some years of development and procurement work to put such arrangements in place
- There are a number of preservation programmes around the world that are developing these capabilities: their technical specifications, preservation policies and procedural manuals should be consulted.

18.4 Scenario 2

For records in a record keeping system (RKS), there could be an impending change in administrative arrangements, such as a change of government, which may lead to large-scale transfer or discarding of records; or an imminent replacement of systems.

Possible actions in response:

- If there is time, embark on an appraisal project to determine how the RKS works, what business activities are recorded, what kinds of records are important, and whether the RKS allows a presumption of authenticity to be made. Records should be sentenced in accord with a disposal schedule prepared from the appraisal process, and records selected for preservation should be transferred, with their metadata, to archival custody, where they may be checked, converted to a standard format chosen to accommodate their essential characteristics, and placed in storage with appropriate back ups
- If there is not time for the necessary appraisal before transfer, it may be necessary to do a quick appraisal to determine that there is sufficient documentation to allow later processing to happen when the materials have been removed from their working context. The records and any tools for accessing them – possibly the entire RKS – could be transferred to archival custody
- Once they have been transferred, set up arrangements for secure storage of the records, their documentation, and any access tools until they can be appraised. Because many RKS are tied to specific operating platforms, it may be

necessary to maintain the equipment and software that originally supported it

- Such material would presumably be a high priority for appraisal, and for establishing suitable strategies to maintain authenticity and accessibility, such as data extraction and migration to a standard format, before the original RKS becomes unsustainable.

18.5 Scenario 3

For commercially produced audio-visual materials, there could be a risk that small producers may go out of business.

Possible action in response:

- Develop criteria for selecting what is worth keeping
- Identify and contact producers or distributors and seek to obtain copies of highest quality source files free of anti-copying devices
- If material is in a non-standard format, ask for copies in a standard format
- Transfer material that has been selected for preservation to the archive, and copy it either to a more stable carrier or to a well-managed computer storage system with proper backups
- Check the quality of the data transfer
- Record documentation about the material including rights information likely to be needed in managing intellectual property rights, which are often complicated for this kind of material
- Because digital audio-visual material involves very large amounts of data, it is unlikely that a non-specialist archive will have the facilities needed to store many items before experiencing capacity problems. It may be better to store items on a reasonably stable stand-alone carrier such as CD
- More ambitious programmes need to develop deposit and rights arrangements with producers or distributors, and to set up sophisticated systems to handle and store large amounts of data, metadata, and possibly arcane playback equipment and software if it has not been possible to convert material to the archive's standard playback system without unacceptable loss.

18.6 Scenario 4

For material issued on short-term magnetic carriers such as floppy disks or tapes, there is the strong likelihood of media deterioration.

Possible action in response

- Seek to determine what material is worth preserving
- Copy material from floppy disks or tapes either to more stable carriers like CD, or to other unstable carriers like tapes that are actively managed by data maintenance systems

- Check the quality of the copying
- Record metadata about the material and the transfer
- If possible, use the transfer as an opportunity to document the software dependencies of the material
- Plan how to deal with the software dependency problem.

18.7 Scenario 5

For legacy material, there may be an impending loss of equipment and software that is required for access.

Possible actions in response:

- Seek to determine what material is worth preserving
- If possible, arrange for the current custodian to transfer the material to a carrier and a file format that can be handled by the equipment and software you have, if it can be done without significant loss
- If necessary, arrange for the material in its original state to be transferred to your custody along with the equipment and software currently used for access
- Either look for some way to maintain the equipment for as long as possible, or immediately copy the material to a different carrier and/or format
- Check for unwanted changes in the material
- Document the material, its provenance and any changes
- Store the material securely with proper back ups
- Plan how you will deal with the software dependencies of the material, especially if it has not been possible to convert it to a file format that you will continue to use
- For material that has already lost the hardware and software needed for access, it will probably be necessary to find someone with the same equipment in working order who is willing to let you use it. This may require the use of forensic data recovery services or purchase of specialist software for data recovery.

18.8 Scenario 6

For complex datasets, there may be an impending loss of staff who know how the data are coded and how the dataset works.

Possible actions in response:

- Determine whether the dataset is worth keeping
- With the help of existing staff who are familiar with the dataset, document it fully so that other staff or contractors can take over its management

- Ensure the dataset is copied and adequately backed up in a secure storage system
- Plan how you will continue to provide access once the current operating environment has been superseded.

MINIMAL PRESERVATION PROGRAMMES

This section outlines possible steps in setting up a minimal preservation programme, which may be applicable in cases where some action is needed but an organisation is not able to commit to anything more ambitious. Some scenarios are also included.

Understand your preservation responsibilities and needs and resources

- Are there digital materials that you should preserve? Is there anyone else who is likely to preserve them? What permissions are needed? What risks or threats need to be addressed? Determine what resources you could apply to the task.

Influence the preservation task

- At least decide on the formats that will be accepted. If possible negotiate with producers to use widely accepted standards and to provide adequate documentation.

Protect the data

- Store media in appropriate conditions
- Copy data to more stable media and make backup copies, using good quality media
- Store data securely, including offsite storage for backups if possible
- Check data for errors regularly
- Establish a data refresh regime suited to the life of the media.

Do something about the means of providing access

- Record information that will be needed to provide short-term access – the identity of the material, access requirements, passwords, etc
- Retain necessary access equipment and software, maintaining hardware and protecting software within licence arrangements
- Other action will be needed to maintain accessibility as it becomes impractical to maintain hardware in working condition
- If further action is not feasible for minimal programme resources, plan to pass the material to another suitable caretaker who will take responsibility
- Alternatively, find ways to adequately reflect the material in a stable, non-digital form (such as printing out). This is likely to be unsatisfactory for

anything other than text or still images.

Minimal programmes can play a positive, but obviously limited role in preserving digital heritage materials.

A minimal programme - Scenario 1

A small production house maintains an archive of its files for each publishing job. The archive is used to provide content and reference source files for re-use of content for its publications. The company needs access to the archived material for its own purposes for at least five years. A small number of its publications meet legal deposit requirements and must be deposited in digital form with a collecting institution for longer-term preservation. The company takes no further preservation responsibility for its material.

The company's archive is stored on CD-R in a secure area in-house, with an additional copy of each CD-R being stored off-site. Early archive files were stored on magnetic and magneto-optical media. The material on these disks is being transferred to CD-R, as the magnetic media may not be sufficiently stable over the required period and the specialised drives required to read these disks are being phased out.

The archive comprises both pre-production and production files in specialised proprietary publishing software formats. The company relies on backwards compatibility or software upgrade versions to retain access to its earlier files. For critical image or design reference files, reference versions are also created in more widely renderable formats such as TIFF or PDF for extended access. These versions are intended as reference for recreating working files *de novo* in the future, should later software versions prove no longer able to render earlier versions.

A minimal programme - Scenario 2

A research institution sponsors an indigenous community heritage programme. Members of the community create a collection of folklore, artwork, genealogical records and interviews, which it makes available to the community via the Internet. The material is hosted by the research institution and is of lasting value as a record of the indigenous community's traditions and culture. Rights have been negotiated with the community to allow the material to be preserved indefinitely.

The site uses current standard web mark-up and media files, at the request of the host institution, which takes initial preservation responsibility for the site and maintains a small collection of software tools that can be used to correctly render its hosted content.

The institution maintains regular backups of all material hosted on its servers, including off-site storage of at least one backup. Individual sites may also be regularly archived to CD-R, particularly when updates to sites are made. Standard IT security measures are in place to protect online content.

The institution has funding to continue the programme for another two years and is in

the process of negotiating the transfer of the material to another institution for ongoing preservation.

A post-minimal programme - Scenario 3

A higher education and research institution maintains a programme for archiving dissertations, research data sets, analyses and models. It intends to preserve and provide access to these materials indefinitely.

The institution limits the range of formats in which it will accept deposited materials, in order to reduce the range of preservation pathways that must be devised over the long-term. The accepted formats conform to open, widely used standards, which are expected to be accessible over a longer period. The institution maintains software capable of rendering these formats at the present time. Depositors are requested to provide detailed metadata.

The institution provides secure storage on both disk array and tape, and has both a backup regime and disaster recovery measures in place, including multiple copies in separate locations and on alternative media.

System and equipment upgrade plans are being devised, and research into methods for translating or interpreting the deposited formats in the future is being conducted. The technological environment is also monitored for signs of impending obsolescence of employed technologies.

SECTION 4

FURTHER INFORMATION

19. Glossary

19.1 Aims

This selective glossary explains terms as used in these Guidelines.

19.2 Terms

Accessibility The ability to access the essential, authentic meaning or purpose of a digital object.

ASCII American Standard Code for Information Interchange. Internationally used standard for encoding to represent all upper and lower case Latin letters, numbers, punctuation, etc.

Authenticity Quality of genuineness and trustworthiness of some digital materials, as being what they purport to be, either as an original object or as a reliable copy derived by fully documented processes from an original.

Bit (Binary digIT) Smallest unit of computerised data, being a single digit (1 or 0).

Blog (Weblog) A log or diary of postings to a web site, often by the site owner, but also often by other invited correspondents.

Browser Software that provides access to World Wide Web pages.

Byte Set of (usually 8) bits representing a single character in computer code.

Cams, cam sites, live cams, webcams Web sites that broadcast images from a video camera attached to a computer, either as a succession of still images or as streaming video.

Certification Process of assessing the degree to which a preservation program complies with an agreed set of minimum standards or practices.

Compression Reduction of the amount of data required to store, transmit and re-present a digital object.

Conceptual objects Digital objects as humans interact with them in a human-understandable form.

Data protection Processes of protecting bit-level data of digital objects from unauthorised changes or loss.

Digital heritage Those digital materials which are valued sufficiently to be retained for future access and use.

Digital preservation The processes of maintaining accessibility of digital objects over time.

Distributed arrangements Arrangements for digital preservation that draw on the responsibility of a number of partners.

Download Process of copying data from a remote computer to local computer storage.

DTD Document Type Definition. A formal definition of the elements, structures and rules for constructing all SGML documents if a given type.

Encryption Process of encoding data into secret code so that only authorised users are able to convert the data back to its original encoding for presentation.

E-prints Digital texts of peer-reviewed research papers, made accessible through the Internet before, during or after refereeing.

Essential elements The elements, characteristics and attributes of a given digital object that must be preserved in order to re-present its essential meaning or purpose. Also called *significant properties* by

some researchers.

HTML HyperText Markup Language. The encoding used to create World Wide Web pages, including markers for text formatting, insertion of objects, and hyperlinks.

Identity of data objects. The state of being distinguishable from other digital objects, including other versions or copies of the same content.

Ingest Process of bringing digital objects and their associated documentation into safe storage.

Integrity of data objects. The state of being whole, uncorrupted and free of unauthorised and undocumented changes.

(**The) Internet** The largest collection of interconnected networks (or internets) in the world, all using the TCP/IP (Transmission Control Protocol/Internet Protocol) protocols.

Logical objects Digital objects as computer encoding, underlying conceptual objects.

Means of access Tools (usually particular combinations of software and hardware) required to provide access to digital objects and present them in a human-readable form.

Metadata Data about data, usually in a highly structured form and often encoded for computer processing and interrogation.

Online publications Digital documents made available to users through a computer network such as the Internet.

Open standards Specifications that are defined through a publicly available process and publicly available for reference and use.

Operating system Software that controls the way a computer operates.

Physical objects Digital objects as physical phenomena that record the logical encoding, such as polarity states in magnetic media, or reflectivity states in optical media.

Plug-ins Pieces of software (sometimes hardware), that add features to a larger software program such as programs to display specific file types.

Porting Process of translating a piece of software from one computer system to another.

Producers Agents responsible for designing, creating, and distributing digital materials.

Preservation metadata Metadata intended to support preservation management of digital materials, by documenting their identity, technical characteristics, means of access, responsibility, history, context, history and preservation objectives.

Preservation program The set of arrangements, and those responsible for them, that are put in place to manage digital materials for ongoing accessibility.

Public domain software Software programs that are free of copyright restrictions.

Refreshing Process of copying data from one carrier to another, without changing how the data is encoded, in order to avoid data loss due to media deterioration or replacement.

Resource discovery metadata Metadata intended to make the existence and description of digital materials visible to those who may wish to access them.

Rights Legally enforceable entitlements associated with digital materials, such as copyright, privacy, confidentiality, and national or organisational security restrictions.

Risk management Process of identifying and assessing risks presented by threats, and if appropriate, taking steps to bring the level of risk down to an acceptable level.

Service providers Organisations or individuals contracted to carry out some or all functions of a preservation program, under the program's overarching responsibility.

SGML Standard Generalized Markup Language. A standard for specifying a tag set or markup language for documents. SGML describes how to specify (in a DTD) the underlying structure of a given type of documents, without defining how they will be displayed. HTML and XML are based on SGML

Standards Agreed specifications or practices for achieving given objectives. Some standards are formally prepared, agreed, endorsed and published by standards-setting bodies, while others become *de facto* standards by common adoption and use. Some standards, such as many file formats, are developed and patented by intellectual property owners who may or may not make their specifications public.

Verification process of checking that a digital object in a given file format is complete and complies with the format specification.

World Wide Web The total collection of resources and servers accessible by the Internet, using the HTTP protocol, which is only one of a number of ways that information can be accessed through the Internet. (Email is another.)

XML Extensible Markup Language. A pared down version of SGML that is expected to become a widely used standard for describing standardised document structures so they can be understood by most computer systems.

20. Reading list

20.1 Aims

This reading list is intended to suggest further sources of information that will take the reader beyond the level of detail possible in these Guidelines. Readers should be aware that understandings and methods in digital preservation are neither universally agreed nor fixed, so these readings may present differing views on some issues.

20.2 Content

This is a highly selective reading list. The available information on digital preservation is extensive; information in associated areas of interest for preservation programs, such as resource discovery metadata or rights management is vast. However, there are few 'standard texts' that could form a core reading list. Most of the references below are extracted from the PADI subject gateway (itself referenced under *Current awareness sources*).

To keep the size of the list manageable:

- Most references already included in individual chapters have not been repeated here. They should also be seen as an important (perhaps the most important) part of the list
- Most conference papers have been omitted, as they are so numerous and usually have a more theoretical orientation. A search on the PADI subject gateway using the term "conference paper" provides access to many useful resources
- References to organisational and project sites have generally been preferred to the multiple (and often repetitive) papers describing them. In most cases, it is possible to find relevant papers through the sites themselves, as well as other technical information.

20.3 Current awareness sources

The following information sites are good sources of up-to-date information and discussion of digital preservation issues:

Web sites

PADI (Preserving Access to Digital Information). National Library of Australia. (Regularly Updated)

PADI is an international subject gateway devoted to the subject of digital preservation. In partnership with the Digital Preservation Coalition, it also produces a quarterly digest of significant new developments.
http://www.nla.gov.au/padi/

ERPANET: Electronic Resource Preservation and Access NETwork (Regularly Updated)

ERPANET, funded by the European Commission, aims to establish an international consortium to provide a virtual clearinghouse and knowledge-base on state-of-the-art developments in digital preservation.

http://www.erpanet.org/

Journals (and regularly released monograph series):

CLIR issues. Council on Library and Information Resources (Regularly Updated)
http://www.clir.org/pubs/issues/issues.html

CLIR reports. Council on Library and Information Resources (Issued regularly)
http://www.clir.org/pubs/reports/reports.html

D-Lib magazine. (Regularly Updated)
An online journal of digital library research
http://www.dlib.org/

Digital Document Quarterly. Henry Gladney (Regularly Updated)
A privately published journal with a focus on digital preservation issues.
http://home.pacbell.net/hgladney/ddq.htm

Publications of the European DigCULT Forum. (Regularly released)
http://www.digicult.info/pages/publications.php

RLG DigiNews. Research Libraries Group
A bi-monthly newsletter providing information on digital initiatives with a preservation component or rationale, on image conversion and digital archiving projects, and current announcements. Archive available.
http://www.rlg.org/preserv/diginews/

Discussion lists: core

DIGITAL-PRESERVATION. JISC
Carries announcements and information on activities relevant to the preservation and management of digital materials in the UK.
http://www.jiscmail.ac.uk/lists/digital-preservation.html

OAIS Implementers Discussion List (oais-implementers@lists2.rlg.org)
A discussion list intended for individuals and institutions actively working with the Open Archival Information Systems (OAIS) Reference Model in an effort to model, build and manage digital archives or repositories.
http://www.rlg.org/longterm/oais.html

padiforum-l
padiforum-l is a moderated discussion list for the exchange of news and ideas about digital preservation issues.
http://www.nla.gov.au/padi/forum/

WEB-ARCHIVE
Comité Réseau des Universités
Focused on on-line content archiving, from the technical, legal and organisational point of view
http://listes.cru.fr/wws/info/web-archive

Discussion lists: non-core but also useful

ERECS-L
A moderated list for archivists and other information professionals which provides a forum for discussion of ideas, techniques, and issues associated with the management and preservation of electronic records.
http://listserv.albany.edu:8080/archives/erecs-l.html

Preservation Administration Discussion Group (PADG-L)
This list covers preservation of both digital and traditional materials. A searchable archive is available.

http://palimpsest.stanford.edu/byform/mailing-lists/padg/

ShelfLife

A weekly executive news summary for information professionals worldwide, published by the Research Libraries Group (RLG) in collaboration with NewsScan, Inc.
http://www.rlg.org/shelflife/index.html

IASA list

Discussion list of the International Association of Sound and AudioVisual Archives.
http://www.rlg.org/shelflife/index.html

20.4 General interest

CAMiLEON : Creative Archiving at Michigan and Leeds : Emulating the Old on the New (Regularly Updated)

Is examining issues relating to the implementation of technology emulation as a digital preservation strategy, and hopes to develop tools, guidelines and costings for emulation compared with other digital preservation strategies.
http://www.si.umich.edu/CAMILEON/

Cedars : Curl Exemplars in Digital Archives Project. (Updated to 2002)

Under the overall direction of the Consortium of University Research Libraries, the project (April 1998-March 2002) aimed to address strategic, methodological and practical issues of digital preservation. Website links to Cedars Guidance documents on intellectual property rights, preservation metadata, collection management, technical strategies and the Cedars Distributed Digital Archiving Prototype System, and to project working papers and articles.
http://www.leeds.ac.uk/cedars/

Changing Trains at Wigan: Digital Preservation and the Future of Scholarship. Seamus Ross (Date Created: Nov 2000)

Looks at the emergence of digital documentary materials for scholarly and evidentiary purposes, and examines the challenges and issues in their effective preservation from a case study perspective.
http://www.bl.uk/services/preservation/occpaper.pdf

A Continuing Access and Digital Preservation Strategy for the Joint Information Systems Committee (JISC) 2002-2005. Neil Beagrie (Date Created: 01 Dec 2002)

Proposes the role that JISC should undertake on behalf of funding councils and institutions as part of a national digital preservation programme.
http://www.jisc.ac.uk/index.cfm?name=pres_continuing

Cyberculture, Cultural Asset Management, and Ethnohistory : Preserving the Process and Understanding the Past. Seamus Ross (Date Created: Jun 2001)

Emphasises the importance of preserving the cultural context in which the Internet operates, highlighting eight challenges for ensuring long-term access to materials in cyberspace, and compares the advantages of centralised, decentralised and distributed archiving models.
http://www.deflink.dk/upload/doc_filer/doc_alle/740_sross_cyberculture_rev2.doc

Digital Division is Cultural Exclusion. But Is Digital Inclusion Cultural Inclusion? Karen Worcman in: *D-Lib magazine* (Date Created: Mar 2002)

Examines "the extent to which digital technologies and the Internet can be instruments of social and cultural inclusion" and "how the use of these technologies can be linked to the preservation of the history of a particular

cultural group." It also notes the impacts of digital technology on history and the collective memory of communities and the challenges in overcoming digital exclusion of economically disadvantaged groups, in the creation and preservation of digital history and of sustainable projects and resources.
http://www.dlib.org/dlib/march02/worcman/03worcman.html

Digital Electronic Archiving : the State of the Art and the State of the Practice. B. C. Carroll; G. Hodge; Information International Associates Inc. (Date Created: 26 Apr 1999)
Study undertaken to provide information on the state of the art and practice in digital electronic archiving policies, models and best practices. International in scope and includes a variety of data types applicable to scientific and technical information including data, text, images, audio, video and multimedia, and a variety of object types such as electronic journals and monographs, satellite imagery, biological sequence data and patents. Several "cutting edge" projects are identified for more detailed analysis.
http://www.icsti.org/99ga/digarch99_TOCP.pdf

Digital Preservation and Deep Infrastructure. Stewart Granger in: *D-Lib magazine* (Date Created: Feb 2002)
http://www.dlib.org/dlib/february02/granger/02granger.html#

European Commission on Preservation and Access (ECPA). (Regularly Updated)
ECPA "acts as a European platform for discussion and cooperation of heritage organisations in areas of preservation and access". Website contains information about projects, activities, publications and other resources related to the preservation of documentary heritage (including digital material) in Europe.
http://www.knaw.nl/ecpa/about.html

JISC Digital Preservation Focus. Joint Informations Systems Committee (Regularly Updated)
http://www.jisc.ac.uk/dner/preservation/

Levels of Service for Digital Repositories. William LeFurgy in: *D-Lib magazine* (Date Created: May 2002)
William LeFurgy of the US National Archives and Records Administration (NARA) outlines conditions governing persistence of digital objects, such as system architecture and material specification, and suggests a model for future levels of service for digital repositories.
http://www.dlib.org/dlib/may02/lefurgy/05lefurgy.html

A Metadata Approach to Preservation of Digital Resources: The University of North Texas Libraries' Experience. Cathy Nelson Hartman; Daniel Gelaw Alemneh; Samantha Kelly Hastings (Date Created: Aug 2002)
This paper discusses the issues related to digital preservation and demonstrates the role of preservation metadata in facilitating preservation activities in general. In particular, it describes the efforts being made by the UNT libraries to ensure the long-term access and preservation of various digital information resources.
http://www.firstmonday.org/issues/issue7_8/alemneh/index.html

Preserving Digital Information : Final Report and Recommendations. John Garrett (co-chair); Task Force on Archiving of Digital Information; Donald Waters (chair) (Date Created: 20 May 1996)
Arose from a decision by the Commission on Preservation and Access (CPA) and the Research Libraries Group (RLG) to commission a Task Force to investigate and recommend means of ensuring "continued access indefinitely into the future of records stored in digital electronic form." This watershed exercise generated discussion worldwide.
http://www.rlg.org/ArchTF/

The state of the art and practice in digital preservation. Kyong-Ho Lee; Oliver Slattery; Richang Lu; Victor McCrary; Victor Tang (Date Created: Jan 2002)

This paper published in the *Journal of Research of the National Institute of Standards and Technology* (vol. 107, no. 1, pp. 93-106) surveys ideas and practice as of late 2001. A final section recommends the development of preservation standards based on XML, and outlines some critical issues that still need to be resolved.
http://nvl.nist.gov/pub/nistpubs/jres/107/1/j71lee.pdf

20.5 Preservation advocacy

Digital Preservation Coalition. JISC Digital Preservation Focus (Last Updated: 14 Feb 2001)
Established in 2001, the (UK) Digital Preservation Coalition aims to develop and pursue a UK digital preservation agenda within an international context.
http://www.dpconline.org/

20.6 Preservation of published materials (library focus)

Access to web archives: the Nordic Web Archive Access Project. Svein Arne Brygfjeld (Date Created: 22 Aug 2002)
Presented at the 68th IFLA General Conference, Glasgow, 2002. Describes a prototype system for access to large-scale web archives, as developed by the Nordic Web Archive Access Project, an initiative of the National Libraries of Denmark, Finland, Iceland, Norway and Sweden.
http://www.ifla.org/IV/ifla68/papers/090-163e.pdf
Also available in French at http://www.ifla.org/IV/ifla68/papers/090-163f.pdf

Berkeley Digital Library SunSITE. University of California Berkeley Library and Sun Microsystems Inc. (Regularly Updated)
This site builds digital collections and services, as well as providing information and support to digital library developers worldwide. Includes links to information on copyright, metadata, preservation and standards; digital library projects; tools for building digital libraries; and training for digital librarians.
http://sunsite.berkeley.edu/

Collecting and Preserving the Web : Developing and Testing the NEDLIB Harvester. Juha Hakala, in: *RLG DigiNews* (Date Created: Apr 2001)
Outlines the outcomes of the NEDLIB Harvester Project for the archiving of Web resources. Some key issues in using this form of technology for capturing materials on the Web are reviewed.
http://www.rlg.org/preserv/diginews/diginews5-2.html#feature2

Columbia University Libraries Policy for Preservation of Digital Resources. (Date Created: Jul 2000)
Statement of policy, including commitment to digital lifecycle management.
http://www.columbia.edu/cu/lweb/services/preservation/dlpolicy.html

DACHS: Digital Archive for Chinese Studies. Institute of Chinese Studies, University of Heidelberg (Regularly Updated)
Operating since August 2001, DACHS "aims at identifying, archiving and making accessible Internet resources relevant for Chinese Studies, with special emphasis on social and political discourse as reflected by articulations on the Chinese Internet." Collected resources include websites, discussion boards, journals, newsletters and single documents. On overview of the archive's collection policy, workflow and technical infrastructure is available.
http://www.sino.uni-heidelberg.de/dachs/

DELOS Network of Excellence (NoE) on Digitial Libraries. (Regularly Updated)
Established in 2000 to facilitate development of an open agenda for digital libraries research. The group is a reference point for all 5th Framework

Programme projects funded by the Information Societies Technologies (IST) Programme.
http://www.ercim.org/delos/

Digital Imaging and Preservation Policy Research (DIPPR). Department of Preservation and Conservation, Cornell University Library (Last Updated: 22 May 2002)

DIPPR draws its members from the Department of Preservation and Conservation at the Cornell University Library and is involved in research, implementation, publication and training, with an emphasis on digital preservation and on mainstreaming the results of ongoing research projects. Activities include research on technical aspects of digital imaging, digital preservation research through Project Prism, and publication of *RLG DigiNews*.
http://www.library.cornell.edu/iris/research/dippr.html

Dspace: Durable, Digital, Depository. (Last Updated: 2002)

Website of the MIT Dspace initiative established with Hewlett-Packard to create a web-based electronic archive of the intellectual output of MIT and other federated partners. Details of the staffing, governance, planning, technical architecture and functionality are available on the website.
http://dspace.org/index.html

A First Experience in Archiving the French Web. Serge Abiteboul; Gregory Cobéna; Julien Masanès; Gerald Sedrati (Date Created: Sep 2002)

Describes preliminary work by the Bibliothèque nationale de France and INRIA in archiving the French web under legal deposit legislation. Defining the perimeter of the French web and versioning issues are also discussed.
ftp://ftp.inria.fr/INRIA/Projects/verso/gemo/GemoReport-229.pdf

IFLANET: International Federation of Library Associations and Institutions (Regulary updated)

Site includes a wide range of information including resources about Electronic Collections and Services.
http://www.ifla.org/

Internet Archive (Regularly Updated)

A non-profit commercial venture that collects and stores public materials from the Internet such as the World Wide Web, Netnews, and downloadable software donated by Alexa Internet. Archived web pages may be accessed using the Wayback Machine interface. Also provides access to films documenting 20th century North American life and culture, digitised from the Prelinger Archives of ephemeral films in San Francisco.
http://www.archive.org/
also see: The Internet Archive, an Interview with Brewster Kahle, in: *RLG DigiNews* (Date Created: 15 Jun 2002)
http://www.rlg.org/preserv/diginews/diginews6-3.html#interview

JSTOR (Journal Storage): The Scholarly Journal Archive. (Regularly Updated)

Aims is to build a reliable and comprehensive archive of important scholarly journal literature. Back issues of paper journals have been converted into electronic formats allowing savings in space while improving access to journal content.
http://www.jstor.org/

Kulturarw^3. National Library of Sweden (Regularly Updated)

Aim of this project is to test methods of collecting, preserving and providing access to Swedish online documents.
http://kulturarw3.kb.se/html/kulturarw3.eng.html

The Last Page of the Internet : the Importance of Preserving the Dynamic Aspects of the Internet. Niels Brugger (Date Created: Jun 2001)

Discusses the complications involved in preserving the dynamic features of the Internet, as identified by media scholar Niels Brugger.
http://www.deflink.dk/upload/doc_filer/doc_alle/1023_NBR.doc

LOCKSS. Stanford University Libraries (Regularly Updated)
Project is building "persistent access" software for libraries. LOCKSS (Lots of Copies Keeps Stuff Safe) provides tools which use local, library-controlled computers to safeguard readers' long-term access to web based journals.
http://lockss.stanford.edu/

Long Term Preservation Study. Koninklijke Bibliotheek (Regularly Updated)
Documents progress in the Long Term Preservation Study undertaken as part of the Project Depot van Nederlandse Elektronische Publicaties (DNEP) in association with IBM. A research plan and a presentation on strategies being investigated, such as the Preservation Layer Model, are available from the site.
http://www.kb.nl/kb/ict/dea/ltp/ltp-en.html

National Digital Information Infrastructure and Preservation Program (NDIIPP). Library of Congress (Regularly Updated)
Contains information on the national planning effort for long-term preservation of digital content in collaboration with representatives of other federal, research, library, and business organisations. Links to many program publications.
http://www.digitalpreservation.gov/ndiipp/

National Library of Australia.
Website includes links to range of prgrams and papers in digital preservation.
http://www.nla.gov.au/

National Library of Canada Electronic Collection. (Regularly Updated)
Provides access to archived versions of Canadian online material. Includes a link to information about the electronic collection and its history and tips on archiving an online publication.
http://collection.nlc-bnc.ca/e-coll-e/index-e.htm

NEDLIB : Networked European Deposit Library. National Library of the Netherlands (Updated to February 2002)
Homepage of NEDLIB, a collaborative project consortium headed by the National Library of the Netherlands (Koninklijke Bibliotheek) and including eight other European national libraries, a national archive and three main publishers. The main goal was to find ways to preserve access to both online and offline (physical format) digital publications as a basic infrastructure upon which a networked European deposit library can be built. Project Technical Working Papers are available by following the link 'Working Papers' on the home page.
http://www.kb.nl/coop/nedlib/

netarchive.dk. Denmark's Electronic Research Library (Regularly Updated)
A joint-initiative of the Royal Library, the State & University Library and the Centre for Internet Research, University of Aarhus in Denmark. This year long study, in 2001 - 2002, examined the capture and archiving of Danish internet activity relating 2001 municipal elections.
http://www.netarchive.dk

Online Computer Library Center Inc. (OCLC) Digital and Preservation Resources. (Regularly Updated)
OCLC is a nonprofit computer service and research organisation whose network and services link more than 30,000 libraries in 65 countries and territories.
http://www.oclc.org/digitalpreservation/

Preservation of Scientific Serials : Three Current Examples. William Y Arms, in: *The Journal of Electronic Publishing* Volume 5, Issue 2 (Date Created: Dec 1999)
Examines three journals in digital form: the *ACM Digital Library*, the *Internet*

RFC series, and *D-Lib Magazine* and discusses measures which can be taken today to preserve access to the information contained within these journals. Solutions proposed are "partly technical and partly organizational". Proposes three levels of preservation: preserving the "look-and-feel"; preservation of access, maintaining both the underlying material and an effective system of access; and preservation of content.
http://www.press.umich.edu/jep/05-02/arms.html

Research Libraries Group (RLG). (Regularly Updated)
Preservation is one of the focuses of the RLG's activities with the long-term retention of digital research materials comprising a key area.
http://www.rlg.org/longterm/

UNESCO Libraries Portal.
Website includes links to great many library sites and information on UNESCO projects.
http://portal.unesco.org/ci/ev.php?URL_ID=6513&URL_DO=DO_TOPIC&URL_SECTION=201&reload=1041937729

20.7 Preservation of records materials (archives focus)

Archival Preservation of Smithsonian Web Resources: Strategies, Principles, and Best Practices. Dollar Consulting. (Last Updated: 18 Oct 2001)
Commissioned by the Smithsonian Institution Archives, this report provides guidelines on the capture, management and long-term preservation of Smithsonian Institution web sites. Incorporating an integrated records life cycle process model, it recommends best practices, plus appendices on documentation of web sites and a preservation metadata model.
http://www.si.edu/archives/archives/dollar%20report.html

Conversion and Migration Criteria in Records Keeping Systems. Association for Information Management Professionals (Regularly Updated)
The proposed standard will address fundamental policy, procedural, and technical issues associated with conversion and migration from one record keeping system to another regardless of record format, so that these systems will insure the context, content, and structure of authentic records.
http://www.arma.org///publications/standards/workinprogress.cfm

DAVID: Digitale Archivering in Vlaamse Instellingen en Diensten (Digital Archiving in Flemish Institutions and Administrations). City Archives of Antwerp (Last Updated: 10 Jan 2003)
The DAVID Project is a collaboration of the City of Antwerp Archives and ICRI to research digital durability in a governmental environment. It seeks to develop best practices for archiving electoral and population data, emails and websites.
http://www.antwerpen.be/david/

Diffuse: Guide to Archiving. (Last Updated: May 2002)
A data archiving guide developed within the Diffuse project of the EU IST program. Discusses major requirements, links to key standards, specfications, best practice examples and white papers on public record archiving.
http://www.diffuse.org/archive_guide.html

Effective Records Management Project. University of Glasgow
Project aimed to produce tools and protocols, and a pilot system for creation and distribution of committee papers within the University. A final report was published in early 2002 and is available in PDF on the project's Web pages.
http://www.gla.ac.uk/InfoStrat/ERM/

Enduring Paradigm, New Opportunities : The Value of the Archival Perspective in the Digital Environment. Anne J. Gilliland-Swetland, in: *CLIR Reports* (Date Created: Feb 2000)

Examines usefulness of the archival perspective in addressing problems in preserving digital information.
http://www.clir.org/pubs/reports/pub89/contents.html

Guidelines for Electronic Records Management on State and Federal Agency Websites. Charles R. McClure; J. Timothy Sprehe
Guidelines developed as part of a research project on records management and preservation strategies for electronic information contained in (US) state and federal agency websites.
http://istweb.syr.edu/~mcclure/guidelines.html

International Council on Archives. (Regularly Updated)
http://www.ica.org/

International Records Management Trust (IRMT). (Regularly Updated)
A London-based organisation, IRMT was established in 1989 to assist developing countries in managing their official records. Searchable on this website are links to IRMTsponsored projects and papers, including resources examining electronic records management issues.
http://www.irmt.org/index.html

National Archives and Records Administration (NARA) (USA) (Regularly Updated)
http://www.archives.gov/

National Archives of Australia (Regularly Updated)
http://www.naa.gov.au/recordkeeping/preservation/summary.html

National Archives of Singapore. (Regularly Updated)
http://www.museum.org.sg/NAS/nas.shtml

National Historical Publications and Records Commission (NHPRC). (Regularly Updated)
The NHPRC has funded much research on long-term preservation of and access to electronic records. Website provides links to project reports. Of particular interest is the section on Electronic Records Projects.
http://www.archives.gov/grants/index.html

Public Record Office (PRO)
The UK Public Record Office aims to assist and promote the study of the past by selecting, preserving and providing access to public records. Two of PRO's digital preservation projects are: EROS (Electronic Records from Office Systems) and NDAD (UK National Digital Archive of Datasets).
http://www.pro.gov.uk/

UNESCO Archives Portal.
Website includes links to great many archives sites and information on UNESCO projects.
http://portal.unesco.org/ci/ev.php?URL_ID=5761&URL_DO=DO_TOPIC&URL_SECTION=201&reload=1036751929

Victorian Electronic Records Strategy Project. Public Records Office, Victoria (Last Updated: 31 Mar 1999)
The project aimed to demonstrate the feasibility of capturing and preserving electronic records; and to provide a set of functional descriptions for electronic archiving. The projects findings, functional descriptions and a general description of the demonstrator system have been published in the Victorian Electronic Records Strategy Final Report.
http://www.prov.vic.gov.au/vers/published/final.htm

20.8 Preservation of audio and audio visual materials

Building a National Strategy for Digital Preservation: Issues in Digital Media Archiving.

Council on Library and Information Resources (CLIR). (Date Created: Apr 2002)

A collection of papers commissioned by the Library of Congress and CLIR as background for the National Digital Information Infrastructure and Preservation Program. Topics of the papers covered six principal areas presenting collection-management issues: large Web sites, electronic books, electronic journals, digitally recorded sound, digital film, and digital television.
http://www.clir.org/pubs/reports/pub106/contents.html

The Care and Handling of Recorded Sound Materials. National Library of Canada; Gilles St-Laurent. (Last Updated: 15 Feb 2002)

Provides a good basic explanation of what sound is, how it is recorded and identifies many different recording media. Discusses handling and preservation.
http://palimpsest.stanford.edu/byauth/st-laurent/care.html

Digital Preservation of Moving Image Material? Howard Besser. (Date Created: 2001)

This article describes the digital technology induced changes occurring in the production and distribution processes of moving image material. Indicates two paradigm shifts likely for moving image preservation: complete works vs asset management and the physical artifact vs content. General approaches to digital preservation and problems for moving image archivists are also discussed.
http://www.gseis.ucla.edu/~howard/Papers/amia-longevity.html

European Convention for the Protection of the Audiovisual Heritage and Protocol on the Protection of Television Productions .... Council of Europe (Date Created: 06 Feb 2001)

This convention, issued by the European parliament, provides for the safeguarding and preservation of European moving image heritage. Parties to the agreement are obliged to introduce legal or voluntary mechanisms for the deposit of audiovisual media in designated archival repositories in their territories. The text is broadly worded so that the legislation will apply to electronic and other new forms of audiovisual expression as they are created.
http://www.coe.int/t/e/cultural_co-operation/culture/Resources/Reference_texts/Conventions/econpataud.asp

PRESTO - IST-1999-20013: Key Links System Specification Document. Presto Consortium (Last Updated: 26 Jun 2001)

A 135 page report developed from the findings of a preservation survey of audio-visual materials in European Broadcast Archives. Includes information on system requirements, technological upgrades, processing methodologies and metadata specifications for digitally archiving film, audio and video.
http://presto.joanneum.ac.at/Public/D32.pdf

The Safeguarding of the Audio Heritage: Ethics, Principles and Preservation Strategy. IASA Technical Committee. (Date Created: Sep 2001)

Identifies problem areas and recommends practices for audio and AV archives, balancing between the ideal situation and the real world.
http://www.iasa-web.org/iasa0013.htm

UPF (Universal Preservation Format) Home Page

Sponsored by the WGBH Educational Foundation and funded in part by a grant from the National Historical Publications and Records Commission, the initiative advocates a platform-independent format that will help make accessible a wide range of data types. The UPF is characterized as "self-described" because it includes, within its metadata, all the technical specifications required to build and rebuild appropriate media browsers to access contained materials throughout time.
http://info.wgbh.org/upf/

20.9 Preservation of data collections

Archiving Scientific Data. Peter Buneman; Sanjeev Khanna; Keishi Tajima; Wang-Chiew

Tan. (Date Created: Jun 2002)

Describes development of an archiving tool for XML data that allows retention of all previous states of the data as it changes over time. Meaningful change descriptions, retrieval of specific versions and history over time for any element in the archive are supported.
http://db.cis.upenn.edu/Research/ki.html

Arts and Humanities Data Service (AHDS) (Regularly Updated)

AHDS is a national (UK) service to collect, describe and preserve the electronic resources resulting from research and teaching in the humanities. One of its aims is to develop strategies for preserving digital cultural heritage.
http://ahds.ac.uk/

Geophysical Data in Archaeology: a guide to good practice. Armin Schmidt (Date Created: 2002)

Offers an introduction to archaeological geophysics and the variety of data that is produced including the raw measurement data, processed data and interpretative drawings. It also provides an invaluable introduction to storage and archiving of geophysical datasets
http://ads.ahds.ac.uk/project/goodguides/geophys/

Inter-University Consortium for Political and Social Research. (Regularly Updated)

Within the University of Michigan, ICPSR acquires and preserves social science data on behalf of 400 member colleges and universities in the US and abroad. Uses migration methods to ensure continuing access to the archived data.
http://www.icpsr.umich.edu/index.html

Long Term Archiving of Digital Documents in Physics - Conference report. Dr Arthur P. Smith

Report of the IUPAP (International Union of Pure and Applied Physics) conference summarizes the discussion about what a digital archive consists of, and lists the conference recommendations.
http://publish.aps.org/IUPAP/ltaddp_report.html

NDAD: UK National Digital Archive of Datasets. University of London Computing Centre

NDAD contains archived digital data from UK government departments and agencies.
http://ndad.ulcc.ac.uk/

Preservation of the Electronic Assets of a University. Oxford University Computing Services; T. Alex Reid. (Date Created: Oct 1997)

Describes how Oxford University has approached the management, storage and preservation of its electronic assets.
http://users.ox.ac.uk/~alex/hfs-AXIS-paper.html

20.10 Preservation of digital art

Archiving the Avant Garde: Documenting and Preserving Variable Media Art. Berkeley Art Museum and Pacific Film Archive (Date Created: 2001)

Collaborative project to develop, document, and disseminate strategies for describing and preserving non-traditional, intermedia, and variable media art forms, such as performance, installation, conceptual, and digital art.
http://www.bampfa.berkeley.edu/ciao/avant_garde.html

Longevity of Electronic Art. Howard Besser. (Date Created: Feb 2001)

Highlights the problems in preserving electronic art, noting the special characteristics of electronic artworks that pose challenges for preservation and proposes practical strategies for preserving electronic art.
http://www.gseis.ucla.edu/~howard/Papers/elect-art-longevity.html

Rhizome.org: The New Media Art Resource

Rhizome.org is a non-profit organisation which aims to preserve electronic art.
http://rhizome.org/info/index.php

Variable Media Initiative. Guggenheim Museum

Seeks to identify artist-approved strategies for preserving variable media artwork (installation, performance, interactive, digital). Artists are encouraged to define their work independent of medium and provide guidelines on how their work may be recast in new formats.
http://www.guggenheim.org/variablemedia/

20.11 Preservation of email

Archiving E-mails. Filip Boudrez; Sofie Van den Eynde. (Date Created: Aug 2002)

A report from the Flemish DAVID Project examining in detail the legal and technical issues related to the preservation of email records.
http://www.antwerpen.be/david/teksten/Report4.pdf

E-Mail and Potential Loss to Future Archives and Scholarship or The Dog that Didn't Bark. Susan Lukesh, in: *First Monday* (Date Created: Sep 1999)

This paper discusses the importance of informal communication and how it is increasingly created in electronic formats which need to be actively preserved. Lukesh recommends action by archivists, software vendors, public institutions and creators, particularly scholars, to aid preservation of e-mail.
http://www.firstmonday.dk/issues/issue4_9/lukesh/

E-Mail-XML Demonstrator: Technical Description. Testbed Digitale Bewaring (Date Created: Oct 2002)

This report describes the prototype software developed by the Dutch Testbed Digitale Bewaring project in its investigations of long-term preservation of email messages. The solution is based on customisation of Microsoft Outlook to allow communication with a central server responsible for metadata collection, conversion and archiving of both messages and metadata in XML.
http://www.digitaleduurzaamheid.nl/bibliotheek/docs/email-demo-en.pdf

Strategies for Capturing and Managing Emails as Records and as Organisational Assets Adrian Cunningham. (Date Created: 18 Apr 2002)

http://www.naa.gov.au/recordkeeping/noticeboard/emails_as_records_files/frame.htm

20.12 Preservation of e-print collections

E-print Services and Long-term Access to the Record of Scholarly and Scientific Research. Michael Day, in: *Ariadne* (Date Created: 22 Jun 2001)

Considers some of the long-term preservation issues for e-print services. Some of the major implications such as responsiblility for preservation, and authenticity are discussed.
http://www.ariadne.ac.uk/issue28/metadata/

Setting Up An Institutional E-print Archive. Michael Gardner; John MacColl; Stephen Pinfield, in: *Ariadne* (Date Created: 16 Apr 2002)

Based on experiences at the universities of Edinburgh and Nottingham in setting up pilot e-print servers; provides an account of several practical issues, including document types and formats, submission procedures, metadata standards and digital preservation issues.
http://www.ariadne.ac.uk/issue31/eprint-archives/

SHERPA: Securing a hybrid environment for research, preservation and access (Last Updated: 2002)

SHERPA is a structured three year project funded by JISC to create "e-print archives" for leading UK research institutions. The archives will comply with the Open Archives Initiative metadata harvesting protocol and consider digital preservation by investigating the application of the OAIS reference model.
http://www.sherpa.ac.uk

20.13 Preservation of physical format digital materials

Bits is Bits: Pitfalls in Digital Reformatting. Walt Crawford (Date Created: May 1999)

This article describes some of the impediments to reformatting digital materials - such as copy protection technology, software and hardware dependencies and encryption.

American Libraries Vol. 30 No. 5 (05/99)

CD-R and CD-RW Questions and Answers. Optical Storage Technology Association (OSTA) (Date Created: 2001)

This series of pages, provided by the Optical Storage Technology Association, covers a number of topics about CD-R and CD-RW media, including some term definitions, media longevity, handling, labelling, speed and quality.
http://www.osta.org/technology/cdqa.htm

Farewell my Floppy: a strategy for migration of digital information. Deborah Woodyard (Last Updated: 29 Apr 1998)

This paper describes a survey of National Library of Australia collection material stored on disk and reports on the practical aspects of migrating floppy disks to CD-R.
http://www.nla.gov.au/nla/staffpaper/valadw.html

Mapping Functionality of Off-line Archiving and Provision Systems to OAIS. Jorg Berkemeyer; Die Deutsche Bibliothek (Date Created: Jan 1999)

Discusses the preservation of physical format digital material by national libraries and in the context of the OAIS reference model.
http://www.kb.nl/coop/nedlib/meetings/frankfurt/GEN-232.doc

20.14 Digitisation

Colorado Digitization Project Digital Toolbox. (Regularly Updated)

The Digital Toolbox is designed to guide administrators through the questions to ask in the initial planning stages of a digital project. Provides information on the technical aspects of digitisation.
http://www.cdpheritage.org/resource/index.html

Guides to Quality in Visual Resource Imaging. Donald P .D'Amato; Franziska Frey; Linda Serenson Colet; Don Williams (Date Created: Jul 2000)

Five guides written in conjunction with the Digital Library Federation, CLIR, and RLG to identify imaging technologies and practices for visual resources. Practical information on project planning, selecting a scanner, factors affecting image quality, measuring image quality, and file formats for master files.
http://www.rlg.org/visguides/

Handbook for Digital Projects : A Management Tool for Preservation and Access. Maxine K Sitts, (Ed) (Date Created: Dec 2000)

A web resource providing information on issues surrounding digital conversion of collection materials. Contributions from many School for Scanning presenters provide information on project selection and management, technical and copyright considerations, digital longevity.
http://www.nedcc.org/digital/dighome.htm

nof-digitise Technical Standards and Guidelines. People's Network Development Team (Regularly Updated)

A technical guide for digitisation projects developed by UKOLN and Resource: The Council for Museums, Archives & Libraries for the New Opportunities Fund. Adopts a life-cycle approach and outlines successive stages in the creation, development, management, access and re-use of digital information.
http://www.peoplesnetwork.gov.uk/content/technical.asp

Selection Criteria for Digital Imaging. Columbia University Library (Last Updated: 14 Jan 2001)

http://www.cc.columbia.edu/cu/libraries/digital/criteria.html

The Society for Imaging Science and Technology. (Regularly Updated)

An international non-profit society whose goal is to keep members aware of the latest scientific and technological developments in the field of imaging.
http://www.imaging.org/

Technical Advisory Service for Images (TASI) (Regularly Updated)

TASI is a service set up to advise and support the UK academic community on the digital creation, storage and delivery of image-related information. Provides information on preserving access to digital images.
http://www.tasi.ac.uk/

20.15 Legal and voluntary deposit

Depot legal et numerotations. Bibliotheque nationale de France (Regularly Updated) (France)

Updated version of the Bibliotheque nationale de France's legal deposit web pages. As well as providing background information to the mandatory deposit scheme, links are provided to current legislation, and to the recommendations of the Conseil scientifique du dépôt légal (the French legal deposit advisory body) that deposit be extended to include online publications.
http://www.bnf.fr/pages/zNavigat/frame/infopro.htm

Legal Deposit from the Internet in Denmark : Experiences with the Law from 1997 and the Need for Adjustments (Date Created: Jun 2001)

In *Papers from the Preserving the Present for the Future : Strategies for the Internet* conference, held at the Royal Library, Copenhagen.
http://www.deflink.dk/eng/arkiv/dokumenter2.asp?id=695

Legal Deposit. National Library of Scotland (Regularly Updated)

Links to documentation about deposit of UK non-print publications, including a 1999 Revised Version of the Code of Practice for the Voluntary Deposit of Non-Print Publications and related explanatory notes. Refers to the deposit of both offline and online electronic publications, the latter being subject to experimental deposit testing.
http://www.nls.uk/professional/legaldeposit/index.html

Management of Networked Electronic Publications: A Table of Status in Various Countries. Elizabeth Martin. (Last Updated: Nov 2001)

A comparison of 16 national libraries on deposit legislation and arrangements, approach and policy, plans, negotiations with publishers, access arrangements and implementation for networked electronic publications.
http://www.nlc-bnc.ca/obj/r7/f2/r7-100-e.pdf
Available in French at http://nlc-bnc.ca/obj/r7/f2/r7-100-f.pdf

A standard for the legal deposit of online publications. Giovanni Bergamin. (Date Created: 4 Jun 1999)

Abstract in English; abstract in Italian available at:

http://www.aib.it/aib/commiss/cnur/dliberga.htm
http://www.aib.it/aib/commiss/cnur/dleberga.htm

Statement on the Development and Establishment of Codes of Practice for the Voluntary Deposit of Electronic Publications. Conference of European National Librarians (Date Created: 2000)

Official joint statement by the Conference of European National Librarians and the Federation of European Publishers. A draft Code of Practice, to facilitate the drafting of locally-endorsed voluntary deposit arrangements, is included.
http://minos.bl.uk/gabriel/fep/

20.16 Metadata

Digital Libraries: Metadata Resources. International Federation of Library Associations and Institutions (Last Updated: 22 Sep 1999)

Links to many articles and sites relating to data documentation and standards.
http://www.ifla.org/II/metadata.htm

Meta Matters. National Library of Australia

This website is intended to help Web content providers improve the effectiveness of searching for information resources on the World Wide Web through the use of metadata standards.
http://www.nla.gov.au/meta/

Metadata Encoding & Transmission Standard (METS). Library of Congress (Date Created: 14 Jun 2001)

Official web site of the METS XML schema for encoding descriptive, administrative, and structural metadata.
http://www.loc.gov/standards/mets/

Preservation Metadata and Digital Continuity. Steve Knight, in: *DigiCULT.Info Newsletter* (Date Created: Feb 2003)

Describes the National Library of New Zealand's digital preservation programme generally and the development of a preservation metadata schema.
http://data.digicult.info/download/digicult_info3_low.pdf

UKOLN Metadata. Michael Day (Regularly Updated)

A general metadata site providing links to projects, initiatives, registries and resources including some software tools for handling metadata and a glossary.
http://www.ukoln.ac.uk/metadata/

20.17 Standards

Digital Library Standards. University of California Libraries (Regularly Updated)

Provides links to resources about a range of digital library standards.
http://sunsite.berkeley.edu/Info/standards.html

National Information Standards Organisation – NISO. (Regularly Updated)

Develops and promotes international technical standards used in information services.
http://www.niso.org/

PDF-Archive Project (PDF/A). Association for Information and Image Management, International (AIIM, International)

A joint activity of the Association for Suppliers of Printing, Publishing and Converting Technologies (NPES) and AIIM, International, to develop an international standard defining the use of Adobe's Portable Document Format (PDF) for archiving and preservation of electronic documents. The project will address support of multipage documents featuring combinations of text and

graphics and the requirements for reading devices to render archived documents.
http://www.aiim.org/standards.asp?ID=25013

Standards for Libraries. National Library of Australia
This site provides links to information about library and related standards, lists of standards, and key standards bodies.
http://www.nla.gov.au/services/stndard3.html

W3C World Wide Web Consortium (Regularly Updated)
The W3C, an international industry consortium, aims to lead the World Wide Web to its full potential by developing common protocols.
http://www.w3.org/

XML for Digital Preservation: XML Implementation Options for E-Mails. Maureen Potter (Date Created: 11 Oct 2002)
Reports on progress at the Digital Preservation Testbed (Testbed Digital Bewaring) of the Netherlands in using XML as a preservation approach.
http://www.digitaleduurzaamheid.nl/bibliotheek/docs/email-xml-imp.pdf

20.18 Some interesting tools

The Computer History Simulation Project (Regularly Updated)
A loose Internet-based collective of people interested in restoring historically significant computer hardware and software systems by simulation.
http://simh.trailing-edge.com/

My File Formats (Regularly Updated)
A web site with information about over 1,000 file formats.
http://myfileformats.com/?old=manufacturers&truespace=.com.html

Software Archaeology. Andrew Hunt; David Thomas in *IEEE Software*, Volume 19, Number 2 (March/April 2002)
A short article describing the problems of understanding software code with little or no documentation. It ends with some suggestions as to how current developers could make code easier to work with in the future.
ISSN: 0740-7459

Windows Desktop Product Life Cycle Support and Availability Policies for Businesses. (Regularly Updated)
An article outlining Microsoft's policy for ongoing support for its desktop business products with timelines and details on specific products.
http://www.microsoft.com/windows/lifecycle/default.msp

21. Index