

Preserving the digital heritage



© Some rights reserved.

Usage and distribution of this work is defined in the Creative Commons Attribution-Non-Commercial-Share Alike 2.5 Netherlands License. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-sa/2.5/nl/>.

Published by

Netherlands National Commission for UNESCO, www.unesco.nl

European Commission on Preservation and Access, Amsterdam, www.knaw.nl/ecpa

June 2007

ISBN 978-90-6984-523-4

Available as pdf at www.unesco.nl and www.knaw.nl/ecpa

Typesetting and design: Ellen Bouma, Edita-KNAW

The paper in this publication meets the requirements of the  ISO-standard 9706 (1994) for permanence.

Netherlands National Commission for UNESCO

Founded in 1946, the Netherlands National Commission for UNESCO has three main tasks. It advises the Dutch government on UNESCO matters in the field of education, science, culture and communication. It provides information to the public about the aims, the programme and the work of UNESCO. It is the link between the Netherlands and UNESCO, a UN organization in which almost all countries in the world are represented.

Netherlands National Commission for UNESCO

P.O. Box 29777, 2502 LT The Hague,

The Netherlands

scunesco@unesco.nl

www.unesco.nl



Netherlands
National Commission
for UNESCO

European Commission on Preservation and Access

The European Commission on Preservation and Access (ECPA) promotes activities aimed at keeping collections in European archives and libraries accessible over time. The ECPA acts as a European platform for discussion and cooperation of heritage organizations in areas of preservation and access. To promote the exchange of experience and expertise, the ECPA organizes conferences, meetings and workshops.

European Commission on Preservation and Access (ECPA)

Royal Netherlands Academy of Arts and Sciences

P.O. Box 19121, 1000 GC Amsterdam, The Netherlands

ecpa@bureau.knaw.nl

www.knaw.nl/ecpa

Preserving the digital heritage

Principles and policies

Edited by
Yola de Lusenet
Vincent Wintermans

Selected papers of the international conference organized by
Netherlands National Commission for UNESCO
Koninklijke Bibliotheek (National Library of the Netherlands)

The Hague, 4-5 November 2005

Netherlands National Commission for UNESCO
European Commission on Preservation and Access

2007

Preface

In November 2005, the Netherlands National Commission for UNESCO, in collaboration with the Koninklijke Bibliotheek (National Library of the Netherlands) and UNESCO's Information Society Division, organized a conference entitled *Preserving the Digital Heritage* (The Hague, The Netherlands, 4-5 November 2005). It focused on two important issues: the selection of material to be preserved, and the division of tasks and responsibilities between institutions. This publication contains the four speeches given by the keynote speakers, preceded by a synthesis report of the conference.

The Netherlands has been involved in UNESCO's work on digital heritage from the very beginning. The decision to develop a recommendation on digital heritage was taken by the 31st General Conference in 2001 following a draft resolution submitted by the Netherlands. Dutch specialists were involved in the formulation of the first draft of what would eventually become the Charter on the Preservation of the Digital Heritage. The Conference in The Hague was part of the global action plan to promote and disseminate the Charter after its adoption by the 32nd General Conference in 2003.

The Netherlands National Commission for UNESCO hopes that this publication will contribute to the ongoing discussion of the many thorny problems associated with the long-term preservation of the world's digital heritage.

Lieteke van Vucht Tijssen
President, Netherlands National Commission for UNESCO

The preservation of the digital cultural heritage is a logical extension of the traditional tasks of national libraries, archives and museums. Executing this traditional task in the digital world, however, requires new knowledge and expertise, and also new policy decisions. In 2000 the Conference of Directors of National Libraries (CDNL) decided to draw up a manifesto with the aim of making national governments aware of the threats and challenges. The Koninklijke Bibliotheek was responsible for preparing this document together with the Library of Congress (United States of America) and the National Library of Australia. It was the first step towards the 2001 UNESCO Resolution and the 2003 Charter on the Preservation of the Digital Heritage.

Since 2001 much progress has been made. However, many questions are still awaiting an answer as the reader will learn from this publication. In order to find solutions for technical problems a permanent research effort on an international scale is needed. Research should involve heritage institutions as well as large research organisations which have large interests at stake. This publication presents different views on the most effective division of responsibilities. UNESCO is a valuable platform for bringing these views together in order to find new models which would meet the needs of all parties involved.

Wim van Drimmelen

Director General, Koninklijke Bibliotheek, National Library of the Netherlands

Contents

| | |
|---------------------------------------------------------------------------------------------------------------------|----|
| Introduction <i>Annemieke de Jong and Vincent Wintermans</i> | 1 |
| Safeguarding our digital heritage: a new preservation paradigm <i>Abdelaziz Abid</i> | 7 |
| Moving beyond the artifact: lessons from participatory culture <i>William Uricchio</i> | 15 |
| Addressing selection and digital preservation as systemic problems <i>David Bearman</i> | 26 |
| Preserving the digital heritage: roles and responsibilities for heritage repositories <i>John Mackenzie Owen</i> | 45 |
| Appendix A: Charter on the preservation of the digital heritage | 50 |
| About the authors | 54 |
| List of participants | 55 |

Introduction¹

Annemieke de Jong and Vincent Wintermans

During its 32nd General Conference in 2003, UNESCO adopted the Charter on the Preservation of the Digital Heritage. Digital heritage consists of digitally created or digitized resources of human knowledge or expression with lasting value and significance. The Charter draws attention to the risk of loss which digital heritage is exposed to due to 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. It urges Member States to take legal, economic and technical measures to safeguard this heritage.

After the adoption of the Charter, awareness-raising activities were launched in various parts of the world. The Netherlands National Commission for UNESCO, the Koninklijke Bibliotheek (National Library of the Netherlands) and the Information Society Division of the UNESCO Secretariat organized the international conference on Preserving Digital Heritage as a contribution to this campaign in the European region. The project group digital heritage that functions under the Netherlands National Commission set the theme and identified the speakers. This project group is made up of specialists from the European Commission on Preservation and Access (ECPA), the Koninklijke Bibliotheek, the Netherlands Institute for Sound and Vision, and the Nationaal Archief. Thanks to a contribution from UNESCO's Participation Programme, a number of specialists from outside the European region were also able to participate.

Representatives from various heritage institutions – libraries, archives and museums – were invited to discuss two important issues: the selection of materials to be safeguarded and the division of tasks and responsibilities between institutions for preservation purposes. The Organizing Committee invited three keynote speakers with a predominately academic background, who sketched the problems and opportunities for long-term preservation of heritage in the digital age. They critically reviewed the current preservation efforts by heritage institutions and proposed various, and sometimes revolutionary, solutions. These papers are printed in their entirety in this publication. The other presentations will be accessible on the Internet².

As a counterpoint to this, three speakers from heritage institutions commented on each keynote speech, providing the participants with a range of state of the art examples from a variety of heritage institutions and cooperation schemes. They

¹ Based on the article *Digitalisering: Digitaal erfgoed: paradigmaverschuiving of 'business as usual'?* by Annemieke de Jong, *Archievenblad*, maart 2006

² http://www.UNESCO.nl/main_6-3.php

provided clear proof – if any such proof were needed – that the urgency of the matter is widely understood, and that a lot of European institutions have taken the advice of the Memory of the World Guidelines for the Preservation of the Digital heritage (National Library of Australia, 2003) to heart: ‘where necessary, it is usually better for non-comprehensive and non-reliable action to be taken than for no action at all’.

The discussions at the Conference were to a large extent dominated by this tension between the visionary and the practical. This confrontation was deliberately sought by the Organizing Committee by inviting speakers with strongly diverging responsibilities, views and backgrounds. It was impossible to reach joint recommendations. Some of the recurring themes of the keynote speeches, and the reactions from speakers and other participants are presented below.

The end of the paradigm of the written archive

Abdelaziz Aziz (UNESCO, information society division) stressed the need to reconsider our certainties about ‘preservation’. Before the internet, it was still possible to bide our time. We could still preserve new carriers, like CDs or CD-ROMS, as physical objects and hope that these would remain stable and readable. Now we know that this option will not work: ‘now, with the Internet, the time is close when we will no longer go out from the virtual spaces in order to be able to use the content that we access digitally’. Digital information is not subject to gradual decay: it either exists, or it does not. Information on paper can be preserved by ‘benign neglect’. Not so with digital information: frequent use and widespread circulation enhance the chances of survival. The preservation of a document is no longer dependent on the lifespan of the carrier, but on the ability to transport the information to new media. If we fail to construct a new regime of preservation, our societies will suffer irreversible damage in their collective social memory.

Participatory culture

William Uricchio (Massachusetts Institute for Technology and Utrecht University) discussed the growing importance of on-line communities in the development of hitherto unknown types of culture. He admits that the digital revolution has serious implications for the media industries, but these can nonetheless continue to work with the assumptions that have been elaborated in the previous period. The real challenges are found in what is known as the ‘social media’, like blogs and massively multiplayer on-line role-playing games. These new, participatory types of culture cut across all established classifications. These social media will be of great interest to future historians, provided that the interaction that constitutes their specific nature can be adequately salvaged. Traditional archiving environments might not be in a position to do this. There is hope, however, that the distributed logic that characterizes these media can also be used to solve the storage problem.

Gary Carp (Director of Internet Strategy & Technology of the Swedish Museum of Natural History) argues that the differences between the analogue and digital ages is not so deep. There is no fundamental difference between a collection of wax rolls of a nineteenth century phonograph and a set of MP3 files. Both have to be presented in a setting that can make them significant to the visitors. The museological canon is extendable to the digital world.

Catherine Lupovici (International Internet Preservation Consortium (IIPC) Programme Officer, Bibliothèque nationale de France) discussed the IIPC's experiences with and philosophy of web archiving. The need to archive the web presented itself when publications that used to be collected under legal deposit legislation started to appear in electronic formats. Collecting these works became more difficult when their characteristics changed in this new environment (loss of stability, hypertext links). The existence of the deep web which cannot be reached by crawlers is another complicating factor.

Libraries have used different strategies for harvesting, either focusing on manageable parts of it (electronic journals), or taking regular snapshots of the national domain, or harvesting on a thematic basis (like the American Presidential Elections 2000 harvest of the Library of Congress).

Eleven National Libraries, including the Bibliothèque nationale de France, and the Internet Archive started the International Internet Preservation Consortium (IIPC) in 2003 to share experiences and to develop standards. These partners have become aware that the only way to record the real web for future users is to handle it on a large scale, at least at the domain level. It will be important to combine extensive automatic harvesting, focused intensive selection for harvesting and deep web deposit. Archiving on a large scale will allow users in the future to apply smart mining tools on historical collections. Archiving on a smaller scale as a result of a careful selection only will rob future researchers of this possibility.

Marie-Anne Chabin (Archive 17) made a distinction between archives 'by birth' (archival records), systematically retained records of the activities of organizations that are destined to become part of the archive from the very beginning of the lifecycle, and archives 'by baptism' (miscellaneous archives), documents from various provenances, selected in less systematic way. This second type has been growing in importance over the past decades, and will continue to do so. Selection of miscellaneous archives is difficult, and therefore the need for regular assessment of selection policies and processes is increasingly important.

There are important and complicating new developments in the digital age: the physical unit of the document disappears, the quantity of records is growing at an unprecedented scale, and the life expectancy of digital materials is short.

Internet communities develop freely across borders, but memory institutions are in the main organized along national lines. The risk that valuable materials will be forgotten is therefore very real.

The relationship between archivists and producers is changing due to the fact that materials can be captured without the creator's involvement. Archivists should involve the communities that produce records more actively in the archiv-

ing process. In turn, this means that archivists must become more involved in the process of information creation.

Systemic solutions for preservation

David Bearman (Archives & Museum Informatics), the second keynote speaker of the conference, argued that the link between selection and safeguarding, which used to be indissoluble, should be thought over. The current situation, in which memory institutions individually attempt to maintain, preserve and make available the heritage is uneconomical, inefficient and ineffective. Archives, museums and libraries should concentrate on making materials accessible to the public. The network should take care of the long-term availability of all digital materials, so that the necessity of selection is no longer there. How this can be done is explained in Bearman's paper. His solution involves extending mandatory legal deposit to digital publications, the divulgence of the source codes of proprietary software to a new type of specialized institution, the Trusted Standards Bodies, which should be responsible in the network for migration to new formats.

Rens Frommé works for V2_, the Institute for the Unstable Media, that started in 1981 as a multimedia artists' initiative in 's Hertogenbosch. In the mid-eighties it became a centre for art and media technology. The opening of the V2_Lab, an international media laboratory, in May 1998 meant an extension of V2_'s practices towards art production, research and knowledge exchange. In V2_'s interdisciplinary workspace, national and international artists, scientists and technicians collaborate on electronic art projects and technical research projects, for example, hardware or open source software development. The often long-term research projects focus on the use of new technical opportunities for artistic means, research on the cultural and social implications of these techniques, and the development of technically innovative (web) applications.

V2_ focuses on the documentation and presentation of representations, not of artefacts. It preserves only its own projects, as research on documentation aspects of the presentation of electronic art activities.

It has no ambition to capture the general heritage of multimedia art.

Eric Oltmans of the Acquisitions & Processing Division of the Koninklijke Bibliotheek (KB), explained the workings of the e-depot of his institution. For the systematic preservation of electronic journals, the traditional model – based as it is on geographical boundaries – is no longer sufficient. International collaboration and coordination are essential. Safe-place networks, like the KB's e-depot, can function as centres of expertise.

The e-depot started in 1993. After a pioneering phase, the first depository contract at the KB was signed with Elsevier in 2002. Other companies followed suit. The current volume of the depot is 5 million items. This will reach 11 million when current agreements come to an end. The depot uses both migration and emulation as elements of its preservation strategy. It does not provide open

access, because the collection in the e-depot must not interfere with publishers' business interests. However, there is on-site access at the library, and retrieval and download for private use is allowed. The agreements with the publishers stipulate that open access to the journals will be provided if the publisher is no longer capable of making the journals available.

The digital fabric of society

The third and last keynote speaker, John Mackenzie Owen (University of Amsterdam) concurred with Uricchio in stressing the importance of participatory culture and with Bearman in pointing to the changing roles of traditional memory institutions. These institutes are successfully making the shift from analogue to digital materials, but this is not enough. Websites, blogs, podcasts, discussion lists and all the other new types of objects that owe their existence to the worldwide web have peculiarities that set them apart from traditional heritage materials: they know no boundaries, they are dynamic, interactive and fragmented. As a consequence, it is not the materials as such that one would like to preserve, but their uses; processes, not artifacts. To capture the digital fabric of society, a new type of heritage institution is needed. These institutions should not replace museums, libraries and archives, but complement them.

Maggie Jones (Digital Preservation Coalition) related some elements of the UNESCO Charter on the Preservation of the Digital Heritage and the accompanying Guidelines with the work of the Digital Preservation Coalition (DPC) in the U.K. Founded in 2001, the organization is comprised of institutions from many different fields, and it is this cross-sectoral nature that is one of its main strengths.

The UNESCO Charter particularly stresses the need for collaboration, and this is what DPC does, through its forums, workshops, websites and awareness bulletin. The Guidelines point to the fact that collaboration can save money, but still requires investment in terms of time and effort.

The Digital Preservation Training Workshop that DPC organized with the University of Cornell was in line with the pragmatic line of the Charter and Guidelines: it showed participants that one can and should work on preservation even if a comprehensive strategy is lacking. Digital preservation should start with creating responsible stewardship through awareness-raising, and continue with targeted projects that can gradually encompass a wider range of stakeholders. Awareness-raising among creators of digital content is at the moment a crucial task for DPC.

Ute Schwens presented the Network of Expertise in Long-Term Storage of Digital Resources (Nestor). The German Federal Ministry of Education and Research launched this project in 2003 to enable cross-sectoral cooperation and to share the experiences of the various institutions in Germany, both with each other and with the rest of the world.

Nestor offers an efficient infrastructure to address and involve institutions, organizations, companies and individuals and to make their competence available to others, thus thereby facilitating communication between interested parties. Nestor has special working groups for Trusted Repository Certification and for the preservation of multimedia applications and objects. Awareness Building is an important part of its activities.

Nestor argues for better cooperation between the institutions that preserve analogue materials and those that concentrate on digital materials. Not only because digitization blurs the boundary between them, but also because of the shared goal: the preservation of cultural heritage.

The discussion between the participants and speakers focused on the role of private parties in long-term preservation. Search engines like Google and Yahoo, or the photo sharing website Flickr are good at providing services, but can they safeguard in the long term materials that have lost their commercial value? Public-private partnership is the keyword here. In these partnerships, the issue of commercial rights will be crucial.

Institutions that want to harvest the 'digital fabric of society' in the sense of Mackenzie Owen must come to terms with another fundamental issue: the right to privacy.

It is often stated that the internet is a global phenomenon and that borders no longer count. Yet this is not the whole truth. International cooperation is generally seen as an important means to overcome the many problems of safeguarding digital materials in the long term. Yet left to itself, the movement to forge partnerships will not tend to be truly all-inclusive. Heritage institutions in the developed world tend to seek partnerships first and foremost between themselves. UNESCO possesses a real global outlook, and this is one of the most valuable elements the Organization brings to the debate.

Safeguarding our digital heritage: a new preservation paradigm

Abdelaziz Abid

A question that has become extremely pressing is which of the enormous trove of digital materials should be kept for future generations, and how to go about selecting and preserving them. This conference is based on the assumption that digital objects demand new preservation policies and agreements. Traditional preservation policies as we know them in the analogue domain, based as they are on the environmental control of the physical conditions of conservation, are no longer relevant. Or, at least, they are no longer sufficient to guarantee the preservation of digital objects.

A new preservation paradigm is evolving in the digital domain. Of the many problems that librarians, archivists and other information professionals are facing, this conference focuses on two: the selection of material to be preserved and the sharing of tasks and responsibilities between preservation institutions. These issues are mentioned in the 'Charter on the Preservation of the Digital Heritage', adopted at the 32nd session of the General Conference of UNESCO in October 2003

This initiative started in 2001 when, 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), an Amsterdam-based non-profit foundation, which outlined the issues involved in digital preservation.³

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.

UNESCO's campaign

UNESCO has developed a strategy for the promotion of digital preservation. This strategy is centred on:

- a wide consultation process with governments, policymakers, producers of information, heritage institutions and experts, the software industry as well as standard-setting organizations;
- dissemination of technical guidelines;
- implementation of pilot projects and human capacity-building based on the Guidelines;

³ <http://unesdoc.unesco.org/images/0012/001255/125523e.pdf>

– and adoption of an international charter on the preservation of digital heritage by the General Conference at its 32nd session, held in October 2003.

The Guidelines were prepared by the National Library of Australia and are based on extensive reviews of literature, the Library's own experience, and UNESCO's organized consultations in various regional centres.

Activities carried out since the adoption of the Charter include mainly human capacity-building through training workshops held in Brazil, China, Ecuador, Ethiopia, Malaysia and Tunisia.

Of particular interest is cooperation with the Committee on Data for Science and Technology (CODATA) and InterPARES, the International Research on Permanent Authentic Records in Electronic Systems. This programme is aimed at developing the theoretical and methodological knowledge essential for the long-term preservation of authentic digital records. UNESCO has contracted InterPARES to support the dissemination and adaptation of its findings to the Caribbean and Latin American countries. The countries directly involved in this initiative are Argentina, Brazil, Cuba, Mexico and Peru. Beginning on 7 November 2005, the InterPARES Project will be host to five scholars from the Caribbean and Latin America for a three-week workshop. While in Vancouver, they will be learning about the InterPARES research methodology, products and findings in order to take back to their own countries the new knowledge generated in the context of the project. The scholars will return on 6 February 2006 for further elaboration of the knowledge acquired and adaptation to the Caribbean and Latin American environments.

Cooperation with CODATA in this field is concerned with training scientists and archivists in preserving digital scientific data. CODATA works to improve the quality, reliability, management and accessibility of data in all fields of science and technology. An international workshop on 'Creating the Information Commons for e-Science' was organized by CODATA last August at UNESCO Headquarters in Paris. The two-day event was a contribution towards the goals of the World Summit on the Information Society (WSIS). The workshop focused on practical steps to ensure that policies related to the production, dissemination, management, preservation and application of scientific data and information support the WSIS Principles and Plan of Action.

The domain of digital heritage

In its traditional sense, heritage, whether tangible or intangible, can be defined as monuments, cultural and natural sites, museum collections, archives, manuscripts, etc., or practices that a society inherits from its past, and which it intends to preserve and transmit to future generations. This transmission is aimed at constituting a common foundation of values and references on which a feeling of membership and a sharing of common social values can develop.

The principles guiding the selection of these assets and the actions of cultural and heritage organizations rest on the fundamental characteristic of lasting value

and significance. Assessment of significance and value rests in turn on criteria such as uniqueness of the object, its irreplaceable character, and a number of other criteria such as time and place, form and style.

A significant part of digital heritage consists of the product of the digital reproduction of pre-existing works, which may consist of texts, images, and sounds, or which may be of an audiovisual, graphic, photographic or cinematographic nature. This digital 'double' does not claim to be an identical copy of the initial work, but only a representation of it.

The second component of digital heritage comes from data which exist only in digital form, whether they are Internet sites, electronic publications, multimedia productions, or cultural or scientific databases containing and organizing textual or graphic documents, sounds, still images or audiovisual or multimedia productions.

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 in the same way 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 involved also poses a problem. The Internet features billions of pages whose average lifespan is extremely short.

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 saved in archives remains accessible in its original form. 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 databases – becomes inaccessible.

Another complex issue concerns copyright, including copyright in the software required to access digital files. An agreement on the principle of 'the right to copy for preservation' still has to be developed worldwide.

A paradigm shift

Before the Internet developed, it was still possible for us to bide our time. Preserving these new disconcerting carriers in a more traditional form remained an option. Although digital technology was spreading very quickly to all spheres of creation and science, it was possible to circumvent it: the virtual was still often just another stage in a circular process from reality to reality. Even the earliest virtual worlds, 3D productions, were finished on film or videotape to be made accessible to the public.

Before the Internet, the computer had only been a means of obtaining real-world results or creating real-world objects, which could then be filed in their final state, independent of their digital existence. The computer constituted a sort of transitory stage in a loop which went from reality to reality.

The Internet sharpens the issues of the digital domain and heritage:

- as a tool for finding information, it is the most gigantic data reservoir;
- as a new vehicle for electronic publishing, supplementing and sometimes substituting traditional modes of content distribution;
- as a tool for commercial distribution and intermediation, and for the provision of services;
- finally, as a tool for bringing about convergence, for merging texts, still and moving images, sounds and audio-visual creations to offer new modes of expression and creativity.

With the Internet, the question is clear: the time is close when we will no longer go out from these virtual spaces in order to be able to use them. We still often print out documents on paper, because reading from paper still feels a little more comfortable, but for how much longer? The Internet sharpens the issues of the digital world and heritage. It obliges us to reconsider all our certainties about the very meaning of the word ‘preserve’, a meaning which comes to us from the remotest of past ages when humans for the first time inscribed what they knew on objects that were longer-lasting than they were, so that their memory could traverse the generations and reach us.

All matter tends to disappear gradually, to dissolve, to disintegrate, to yellow, to age – but not digital information. Information either is, or is not. Storing digital information will be like preserving the flame of a fire: you have to tend to it constantly, maintain it, nourish it. Otherwise it will die out and be lost. On the other hand, it will remain eternally young.

This will not happen without significant change on the part of those institutions responsible for preserving documentary heritage. Letting documents lie on shelves in appropriate physical conditions was the best guarantee of preservation, and even allowing people to look at documents was long considered to be the worst enemy of conservation. On the contrary, the ability to allow digital information to circulate rapidly on new carriers, to migrate from one carrier to another, will be the ultimate guarantee of its permanent existence.

If we are not careful, our societies – under pressure from technology cycles – risk witnessing the loss of whole areas of memory as we move further into the new millennium.

As long as information has used physical media in order to move around, it has left traces. Even if we do nothing about it, something always remains of those traces, something which can be made into an archive. But the digital domain has one congenital defect: if you do not save something, you effectively erase it. In other words, the preservation of heritage must henceforth be a deliberate, voluntary act, organized in the present.

Furthermore, the digital domain reverses the very propositions which seemed the most certain: the survival of a document is not dependent on how long the medium carrying it will last, but on the capacity of that document to be trans-

ferred from one medium to another as often as possible. A new preservation paradigm is born.

Since time immemorial, the methods and practices of documentary heritage conservation have given the highest priority to preservation of carriers: paper and ink, the various generations of computer disks, magnetic tapes or emulsions for film, photography or microfilm. In the digital domain, it is the wine that is to be saved not the bottle.

Fragile heritage

The digitization of all data produced by human intelligence, whatever their original form, simultaneously affects the process of creating content, the way in which content is disseminated, and the ways it can be preserved over time. This digitization is happening to a greater or lesser degree in all spheres of activity, in the production and marketing of goods and services, in artistic, intellectual and scientific creation, and in public administration.

The effects of this on our modes of production of and access to culture and knowledge cannot yet be fully measured.

Currently, scientific institutions face the challenge of managing incredible quantities of diverse data – in some cases, several hundreds of gigabytes per day – resulting from laboratory experiments, life-size experiments or observations from various instruments (satellites, radar, telescopes, probes, sensors, microscopic cameras, etc.), some of which constitute actual historical events which can never be repeated. This is the case, for example, with meteorological phenomena which absolutely must be preserved to allow the development of weather forecasting techniques by analysing data accumulated over several decades.

It could be argued that the safe-keeping of this knowledge capital is just as important for the world of science as creating and interpreting that knowledge. In the absence of a specialized structure capable of coping with these growing masses of information, the perpetuation of ‘scientific heritage’, itself a spring-board for new discoveries, still remains a real challenge for our modern societies: its loss would be an irremediable decline.

The end of the 1970s and the beginning of the 1980s saw the birth of office automation. The world of publishing and the press were the next to be massively affected by this revolution in page design and printing.

Over the last few years, libraries have digitized their collections and put them online; ever more powerful search engines excavate millions of pages to satisfy our curiosity. A gigantic universal library is setting itself up before our very eyes.

Sound recording, a pioneer in the digital field, entered very early into general use. Today, music has become the largest cultural industry by number of files exchanged on the Internet (no less than three million MP3 files were exchanged every day in 2002).

Sound was followed by images. Digitization was initially limited to the processing of fixed images, while the development of scanners and image processing software made this increasingly accessible to the general public. Sim-

ilarly, the relatively lighter weight of fixed image files very soon opened up the possibility for photographic agencies to circulate images across their networks.

From then on, and with the development of standards for digital compression and image processing, the digitization of the audiovisual system was set to affect all its components, from production to editing, from editing to transmission control, from transmission control to broadcasting networks, and finally to the private individual's TV sets.

E-commerce and e-governance

Today, online public services and administration are forming part of our landscape. However, they are only the visible face of the process of computerization of organizations. In-house e-mail and intranets are contributing to profound changes in methods of production and information flow. Traditional power centres, founded on the possession of rare and invaluable information, are being called into question, as are pyramidal work organizations which are losing ground to more collaborative ways of working.

This is not without its effects on archiving practices. These are based, for example, on the origin and hierarchical position of the information producer: the higher the level in the hierarchy of the decision-makers, the greater the testimonial value of documents they generate, and consequently the more they deserve to be preserved. Will the weakening of this selection criterion lead archivists to want to preserve everything, for fear of losing the essential?

E-culture and e-learning

The Internet naturally finds its place in the field of culture and education while supporting the emergence of new cultural practices.

Virtual museums represent one of the most innovative manifestations of this, and their dynamism led to the creation in 2001 of the domain name '.museum' for museum (see the site of ICOM, the International Council of Museums) reserved for the museum community in order to allow them to improve their visibility and presence on the Internet. By facilitating access to their works and exhibitions, and by using teaching devices based on interactivity and hypermedia, they are contributing to a renewal of modes of appropriation and understanding of cultural assets. By reducing the barriers of geographical distance and by allowing new cultural practices to emerge, they provide powerful support to policies for the democratization of culture.

What is true for museums is also true for libraries, which for several years now have been exploiting the potential of the Internet to broaden access to their collections, by putting their bibliographical databases online, and gradually also their collections of digitized works. Moreover, they make it possible to access rare and invaluable works, often kept in storerooms so that they are not exposed to the risk of physical damage.

This broader access to cultural sources is bringing considerable benefit to the scientific and educational world. In particular, it opens the way to new modes of

cooperation, especially in the area of training, where e-campuses, virtual classes and e-learning programmes have been developed in recent years. As a communication tool, the Internet also helps cement communities of interest around a set of themes or areas of knowledge through dedicated websites and forums.

All this contributes to building an ever-increasing digital heritage.

Infinite heritage?

But important as it is, the problem of volume, from a purely technological point of view, is not without a solution, since progress in electronics allows a constant increase in the capacity of storage media for a progressively lower cost per stored megabyte.

For all that, is everything produced suitable material to become heritage? And even if it is, which avenues should it follow, which treatments should it undergo to enter the domain of heritage? Should it be randomly left to technological progress or should it be the result of a voluntary, controlled heritage preservation process?

If we consider the production resulting from the digitization programmes of cultural institutions, we are clearly on familiar ground: the works concerned are defined, identified, and listed, even if the specialist techniques employed are not yet completely familiar.

Ultimately, and in every case, the objective is to end up with a duly described digital archive of each object, which will be recorded either on a movable physical medium or on a data server.

However, the approach is quite different when it comes to the Internet. Here, the unity of the document is lost in hyperlinks; flow replaces the finished object. In this universe, traditional methods of collection or acquisition no longer apply, and there are scarcely any other solutions available to heritage organizations than to set up automatic collection devices. These are based on software 'harvesters' which traverse the Web, carrying out regular recordings. Their work is guided by a search plan which makes it possible to select the pages to be recorded in order to ensure their conservation.

Various procedures can be employed. For example, random samples may be taken, the search software providing a snapshot of the temporary state of the Web at any given moment. This was how the American pioneers went about building up the first archives of the Web, Brewster Kahle's Internet Archive.

Other heritage organizations have implemented selection strategies based on well-defined criteria, either by subject, form, language, or nationality. These make it possible to create partitions in the whole of the Web, so as to control its mass in the long term. They also make it possible to control the harvester robot inside a site, as it surfs from link to link.

Other collection strategies can be based on criteria of content or theme, which makes it possible to constitute specific archives. Lastly, this selection can be carried out according to formal criteria, by considering the form of expression as such, which returns us to the nature of the media present on Web sites. In France,

for example, INA (the National Audiovisual Institute) concentrates on the conservation of Web radio and TV, whereas the BNF (National Library of France) is more interested in the products of electronic publishing.

In this digital universe, it can clearly be seen that all the efforts of heritage institutions will be concentrated on taming this flow, channelling it into thematic, geographical, linguistic or formal categories, and organizing this prolific and polymorphous data mine

In the analogue universe, the transfer operations carried out to make backup copies always caused losses in quality of the content. But although digital technology makes it possible to avoid quality losses in the signal, it still cannot guarantee survival. Even if the supporting medium remains intact, it may become impossible to decode the data, given the considerable instability in encoding standards and formats when the document was born.

Concluding remarks

Our societies have witnessed the end of the paradigm of the written archive, a paradigm that had developed over hundreds of years. Throughout the twentieth century, new media have wisely and modestly joined this prestigious tradition. This paradigm has already been transformed, and the devices in place are unable to deal with the brutal advance of information technologies, and the quantitative inflation which they cause. This goes beyond those institutions specializing in the management of memory: a whole new regime of information will have to be constructed, and quickly, completely transforming old memory and archiving systems. If this shift does not take place, our societies will suffer irreversible damage in their collective social memory.

Moving beyond the artifact: lessons from participatory culture

William Uricchio

I can't seem to shake a *déjà vu* feeling when considering the current moment of media change, as if, knowing what I do now, I was present during cinema's first decade. How, I often wonder, did that new medium's contemporaries fail to preserve films, key texts, and audience experiences? But when I consider the lax and, with a few notable exceptions, unsystematic efforts that are being made to preserve our latest 'new' medium, it should come as no surprise. In the pages that follow, I will speak as a cultural historian who makes extensive use of archives and has an abiding interest in heritage and access. But I also speak as someone whose personal and working life is increasingly bound up in new media and particularly social media. These two positions are hardly incompatible. On the contrary, they seem to me to offer not only real advantages, thanks to the computerized archive, but real opportunities for historians to reflect upon the birth and development of the latest 'new' medium. And they bring with them the responsibility of learning from the past and preparing for the future, and urging the thoughtful and systematic archiving of this latest chapter in our cultural history. What follows will therefore necessarily have a polemical edge. The stakes for our culture and our ability to understand are both immediate and profound.

Posing the problem

The 'digital turn' has enabled the transformation of such familiar media as photography, film, recorded sound and the printed word, introducing new platforms, standards and applications. As Hollywood's growing dependence on digital production techniques and DVD sales shows, and the press's routine embrace of computer-based information flow attests, the media industry has quickly accommodated these transformations. There are of course some concerns about putting low-cost, professional digital production equipment in the hands of ordinary citizens, but the media industry is quickly learning to turn a profit on the results. The revolution provoked by digital distribution seems more of a problem, and although 'piracy' is the term most often asserted by the industry, its own failure to develop adequate digital distribution models is a far more serious liability. But these problems largely reside in the domain of business strategy, and while many in the business and political community fret about the implications of digitization for the bottom line, few have bothered to ask whether digitization has fundamentally altered the nature of the media with which they are concerned.

The archival world has, by contrast, shown deep concern with what might best be termed ‘media ontology’ and the implications of the differences between, for example, the grain and pixels that constitute a film image. Although digital technologies have demonstrated their advantages for helping with the task of image and sound restoration and preservation, a significant cohort of archivists has argued that a shift in medium, from analogue to digital, fundamentally undercuts their mission. They argue that preserving film is not the same as storing compressed digital data. Yet while profound in its implication, even this shift leaves the basic parameters of the traditional archive largely intact. Whether digital or analogue, pixel or grain, the archive remains concerned with the restoration and preservation of cultural artifacts that are still called films, books, audio recordings and photos. True, a fundamental debate regarding media ontology remains unresolved and the indiscriminate use of nomenclature such as ‘film’ and ‘photography’ is problematic, but collections still adhere to the basic selection and cataloguing parameters of the analogue past.

Of far greater concern to our culture’s memory institutes should be new categories of digital culture that have been enabled particularly by networked computers – categories that are growing exponentially. Consider the case of email – computerized versions of the physical letter (of course with a fabric of stylistic and formal differences) that generally speaking enjoy no systematic archiving or even technical standard. Future historians of the ‘information age’ will ironically face a gap in key evidentiary domains when attempting to make sense of their topic since most business archives are designed to account for paper records, but not electronic. And emails are a best-case scenario.... Social media such as blogs (web logs), wikis, massively multi-player role playing games, and various on-line social spaces that lack any homologies to traditional archival objects face a far more difficult situation. In the case of social media, there are neither pre-existing archival categories nor memory institutes charged with collection, selection, restoration, preservation and access. Indeed, there seems to be a very real debate about how to even think about these emerging media forms. They do not adhere to the familiar ‘mass’ distribution of the traditional media (although their reach can be greater than many traditional media), nor are they as atomized as telephone calls and individual correspondence (although they can be uniquely person to person), nor is their textual identity necessarily stable and fixed in the way that we think of photographs or films or books (although they can inhabit a range of positions from dynamic, like games, to stable, like e-books). They fall outside of the familiar limits of our cultural habits and expectations, and since memory institutions are largely involved in the business of creating and maintaining tradition, it’s easy to see why these new forms are so awkward.

The challenges to those concerned with preserving cultural heritage are enormous. Not only must we move beyond familiar objects and homologies in our

choice of what to save, but we must also attend to the larger cultural shift towards participation and collaboration. Blogs and wikis are not only highly dynamic as texts; they are examples of networked and collaborative cultural production. They depart from the strictures (and legal definitions) of authorship familiar from the book or film, and are instead accretive, multiply voiced, collective, and ongoing. Although we have some historical precedents for this type of authorship (examples ranging from the Old and New Testaments to quilts come to mind), these have tended either to acquire an institutionally imposed stability (the Bible) or been repositioned from media text to artistic artifact (quilts).⁴ Yet this collaborative dynamism is a core defining aspect of today's new social media. Without familiar criteria such as authorship, or the text as an authoritative and fixed entity, or the many institutions that help us to construct hierarchies of importance, how will we know what to save? What evaluative criteria should we deploy? And even if we could answer these questions, what would be an appropriate point to 'freeze' and hold these dynamic artifacts in the archive? Where do we draw the line between the cultural and the social, between the artifact and the means of its production? And even if we accept the social and acknowledge that patterns of interaction are as important as the text, how do we go about documenting them in meaningful ways? The answers are not clear, but there might at least be productive new ways of posing the questions.

The stakes of facing such questions go beyond the need to document a significant change in how our cultures represent themselves and their experiences and how they circulate those representations. These are, of course, vital issues particularly at moments of transition, where new constituencies are empowered to represent and circulate, or where people develop new technological affordances and new ways of using them. Such broad concepts as knowledge and the public are redefined and empowered at such moments. And while of great interest to cultural historians such as myself, these concepts also help to define the fabric of our daily lives. Consider our changing relationship to the news. Once vetted and produced by trained journalists and centralized institutions with carefully cultivated ideological profiles and professional reputations, news has slipped onto the Internet where it is often neither vetted nor contextualized and where it circulates at the speed of light. When there were well-established filters and a clear line of responsibility, one could agree or disagree, but the status of the information was known. Having 'seen it on the Internet' does little to pin down critical parameters or accurate sources of information. Yet, perhaps even more than rumour, such information can take on a life of its own, confusing public discourse. As we now know about so-called misinformation campaigns, digitally altered photographs,

⁴ In fact, most of our cultural texts are profoundly intertextual, relying upon layer upon layer of precedent, commentary and reference. Particularly since the Romantic Era, we have tended to suppress this unruly aspect of authorship, a practice today enforced by the strictures of intellectual property. The digital turn has forced the question upon us, and as we think through the implications for new media forms, we might reconsider the questions put forward by the likes of Michel Foucault and Roland Barthes and reflect upon what we have long forgotten about the reign of authorship.

and re-worked press releases, and even the highest levels of some national governments have made use of the new media to manipulate the record and with it public perception and knowledge. The policy implications both for day-to-day politics and for the historical record couldn't be more profound. Here, too, is an area where systematic archiving of social media – dynamic texts as well as circulation patterns – could greatly assist the functioning of our political lives by allowing us to go back to the record and check instances of manipulation.⁵

Such an intervention, it needs to be said, can only serve the present to the extent that the public is educated in the critical assessment and use of these new media. Our cultural knowledge and assumptions are grounded in the traditional centralized media. The press and book publishers, film and music companies, broadcasting authorities, and the rest have long exercised a near paternalistic control over media production and circulation. More importantly, at least in the West, interpretive systems grounded in 2,000+ years of biblical exegesis, legal precedent working out the meanings of terms, and dictionaries and lexicons pinning down the precise functioning of language, have all worked hand in glove with centralized media systems to circumscribe meanings and enforce responsibility. Audio and visual culture challenged the precision of the word, but even here rather strict protocols developed in order to minimize charges of libel and maximize impressions of consistency and professionalism. But the digital turn and the steady emergence of bottom-up and collaborative media require new ways of encountering texts and new skill sets on the part of media users. This new literacy can no longer take the news at face value, but rather must consider it as partial, as a construction, and as data that must be critically assessed and combined with other sources in order to yield some sense of the truth. But I digress....

Web 2.0: social media

In a world where people increasingly manage, direct and re-define when and where they experience media, how they share it, and what it means to produce and consume, we need to think about cultural processes in dramatically new ways. The shifts from analogue to digital, from centralized to dispersed, from mass media to social media, from information transmission to collective intelligence, from old statistical extrapolations to new data feeds, all point to media

⁵ But one example involves President Bush's 2003 'mission accomplished' declaration in regard to Iraq on the USS Abraham Lincoln. If we set aside for a moment the staging of the press conference (the background contained a banner reading 'mission accomplished' and photographers were forbidden to take any shots that would show that the ship was in fact not at sea, but instead docked in the San Diego harbour), the White House took an additional step to control information that amounted to rewriting history. The office of the Press Secretary initially issued a statement on its website on 1 May 2003 entitled 'President Bush Announces Combat Operations in Iraq Have Ended.' Later, after it was clear that combat had not ended, the headline was modified to read '... Major Operations...'. The press release, a matter of record, was re-worked to accord with White House spin, but back dated to 1 May, and attempts were made to prohibit access to the earlier and now telltale version. Thanks to regular archiving of the website, the White House was found out and future historians will understand that the criticism generated by the first version was based on real events, not the fantasies and misinterpretations alleged by the White House after its revision of the historical record.

use that is social by design, not social by default. The descriptors used by Tim O'Reilly and others for Web 2.0 are revealing: 'an architecture of participation,' 'harnessing collective intelligence,' 'exchanging,' 'pooling,' 'collaboration,' Whether we look to social media communities such as *technorati*, or social networking sites such as *friendster* and *myspace*, or media fashion communities such as *digg*, we see new logics that have to be understood if we are to make use of them. I'd like to borrow a term from the worlds of biology, neurology, and computer science that speaks to these new conditions: high connectivity. High Connectivity in these various domains refers to dynamic networks, high magnitudes of contact among many nodes, and both interactive and iterative behaviour.

Before looking a bit more closely at what these developments portend for culture, and by consequence, our memory institutions, first a few words about the technologies that have been drawn upon. The World Wide Web was largely made possible thanks to the introduction of Mosaic and the Pentium chip a little over ten years ago (ca 1995). Making far more effective use of networked computers, these developments were also aided by significant increases in cable carrying capacity (broadband), compression algorithms, and cheap memory. These developments have continued, and two parameters in particular have a bearing on the archive: memory and transmission speed. Memory capacity is growing ever vaster and cheaper as a look at the retail electronics sector will demonstrate. As production capacities grow, so too does our ability to hold the massive amounts of data that are being produced. Transmission speeds have also improved significantly: just a week before this UNESCO gathering, Japan's Kansai Electric demonstrated a transmission rate of one terabit per second, or the equivalent of a two hour film in half a second.⁶ Together, massive increases in memory and the new transmission speeds combine with a near geometrical progression of wired households in some national contexts to suggest that the developments we've seen since the introduction of Mosaic nearly a decade ago are but the tip of the iceberg. We have a sense of what is looming in the distance, but its magnitude is not yet visible or even imaginable. Yet we also know that these changes are touching most parts of our lives, from banking and buying train tickets, to entertainment and leisure, to driving and working. But the most interesting developments – and the developments that we need to think about archiving – remain submerged and out of sight.

As suggested at the outset, the implications of the digital turn for media production have enjoyed most of the press coverage. Armed with a modestly priced computer and video camera, a skilled teenager can produce films or music that achieve industry technical standards. And while a big step up from the 8mm or 16mm home movies of two or more decades ago, the real advantage of digital culture has been in transforming 'home' movies into 'world' movies. New logics of distribution have enabled grassroots producers to sidestep the control long

⁶ Between delivering this paper and correcting the proofs, transmission speeds have continued to accelerate. In March 2007, Alcatel-Lucent announced speeds of 25,6 terabits per second, or something like the transmission of 60 two-hour films per second.

exercised by media corporations and governments, and reach a global audience. The rapid circulation of digital texts has also stimulated the growth of cultural hunters and gatherers who cut and mix, collect and reassemble, borrow and repurpose, and who do so as collectives. These practices are not so distant from those evident in pre-industrial and agrarian cultures (or said another way, in the era of pre-commoditized culture) – again, consider the work of quilters, folk singers and storytellers that might be characterized in precisely the same terms. And just as notions of ownership were blurred in folk cultures, so are they unclear with many of today’s new media practitioners. The contemporary blurring of intellectual property can be attributed in part to the new logics of digital culture, and in part to the increasingly draconian control over our cultural heritage asserted by corporate copyright and trademark holders. As intellectual property protection steadily expands at the behest of corporate interests, the public is increasingly deprived of anything other than paid access to their own popular culture. And as corporate profits invariably decline in traditional sectors such as music and film, the pressure to expand control over intellectual property only grows, even though the decline correlates to the increasing market share of cell phones, computer games, and exponentially increasing involvement with social media.⁷

The challenges posed to the old order by the collaborative logics of Web 2.0 can be seen in many sectors. Collaborative news networks such as *Slashdot* (www.slashdot.org) and *Kuro5hin* (www.kuro5hin.org) have blurred the distinctions among editors, readers and writers, with their participants fulfilling all roles. They draw on correspondents and commentaries from around the world, complicating, contradicting, and compositing various bits of information so that the reader must actively consider multiple points of view and sources before making a determination about the news. In a similar way, open source software initiatives such as Linux draw upon a community of collaborators, and by keeping the source code open, they direct their energies towards improving functionality rather than building encryption systems. Moreover, with ample networked input and development, they have the advantage that their software mutates and improves more quickly than proprietary models. Add low costs into the mix, and it is little wonder that Linux is steadily winning terrain from centralized companies such as Microsoft.

Decentralized, networked, collaborative, accretive, ephemeral and dynamic... these developments and others like them bear a closer resemblance to oral cultures than to the more stable regimes of print (writing and the printing press) and the trace (photography, film, recorded sound). Time-bound and contingent, they are at odds with the durability of the printed word and photographed im-

⁷ The market success of Japanese *manga* and *anime* in the US (where *manga* outsells domestic comic books by a factor of four to one) is due in large part to the work of fans who imported, translated and circulated these cultural products without copyright clearance. They effectively created a commercial market, tested it, and their translations in many cases remain superior to commercial alternatives. The industry understands the value of these ‘pirates’ and both sides have worked out a harmonious co-existence.

age et al. Spread from person to person, with the always-present possibility of manipulation and mutability, they differ in the main from the relative stability and uniformity of the traditional fixed media.⁸ And like oral cultures, they seem to evade the preservation frameworks that we have put in place in our institutions of memory, built as they are around tangible media. And yet, despite these conditions, these collaborative efforts also enjoy embodiment as digital text, image and sound, and as such differ from oral culture. They can be apprehended, but the question is, at what point? What constitutes a sufficient 'capture' in a dynamic and fast evolving distributed network where any of the nodes is capable of change?

If we take an instance of social media such as the *Wikipedia* (see <http://en.wikipedia.org/wiki/Wikipedia> for an excellent summary), we can see the benefits of collaboration and mutability. A reader-written and reader-edited encyclopaedia, its entries change over time to reflect new developments, complications in and regional inflections of meaning, and the lively debates among its contributors. Its entries evolve, and as such are responsive to the latest undulations of scholarship. Yet, for all of the *Wikipedia*'s value as a documentary history of changing cultural conceptions and definitions, it confronts the archivist with a problem: what point is the right point to fix and hold for posterity? Social networks such as *Friendster* (www.friendster.com) and *Flickr* (www.flickr.com) will offer future researchers invaluable information about the construction and functioning of our society. Not only are they as textually dynamic as the *Wikipedia*, but the fabric of thematic tags (the elements that permit associations on the basis of location, interest, or background) and the patterns of links (as users form elaborate networks) are themselves valuable data. Indeed, the textual data only acquires meaning thanks to these other parameters. Again, what is the archivist to do? It's as if we would expect librarians not only to keep books, but to track their circulation as well. Yet with social media, circulation is a crucial and defining measure of meaning and import. A third complicating parameter is introduced by online games, for example MMORPGs (massively multiplayer on-line role playing games) in which the texts are extraordinarily dynamic, the social networks crucial to the substance of the game, and the interactions of players with one another actually constitutive of the gaming experience. Pity the archivist! Yet these games occupy a significant amount of cultural space, have social implications, and provide a voice to their users on a variety of topics. *America's Army* (www.americasarmy.com) not only offers an interesting text against which to read US involvement in Iraq, but it includes remarkable debates among its users over issues that range from the political to the tactical, and offer informed readings from voices that researchers rarely have access to. In some cultural settings, South Korea for example, the penetration of virtual worlds is enormous and growing

⁸ Although in truth, written texts such as Shakespeare's plays have undergone hundreds of small (and sometimes major) transformations, and films routinely change as they pass through the hands of censors and countless projectionists. Nevertheless, they are relatively stable compared to the dynamic state of wikis, blogs, and other social media.

(South Korea Telecom's Cyworld includes nearly half of the population). Is this something that we can afford to ignore, fixating instead on the extension of traditional 19th and 20th century cultural forms in our digital and networked present?

Signs of success

Social media differ fundamentally from the media around which our archival policies have been constructed. Their rootedness in community and collaborative interactions, and their responsiveness to an ever-shifting present, gives them a unique quality as *Zeitdokumenten*, finely-grained embodiments of culture. Of all the differences from traditional media discussed above, perhaps the two most difficult ones for memory institutes are 1) the notion that these media are dynamic and always in progress rather than having a final, completed state; and 2) that these media are networked, and that the web of connections bears heavily on the meaning of any one site. These two parameters are not only unfamiliar, but somewhat daunting in terms of their potential storage requirements. Fortunately, as mentioned at the outset, memory is getting cheaper by the day and transmission speeds are improving significantly. But just as importantly, the very distributed logic that enables social media and peer-to-peer (P2P) applications holds a potential solution to the storage problem. Consider music file exchanges such as *Napster* or *Grokster* and the like. They managed something well beyond the physical (let alone financial) capacities of the traditional music industry by digitizing a vast amount of music, storing it, and distributing it on demand to millions of users. They did this by networking millions of personal computers in homes across the world, distributing the tasks of digitization, storage and access in such a way that the cost was negligible and the labour involved was freely given. Millions of modestly-sized memory chips and processors, when properly linked, emerge as a formidable and robust computing force far beyond the potentials of any one centralized system. This network, the basis after all of social media, could also potentially be put to use as a decentralized memory institute, complete with the redundancies, checks, and quality controls currently evident in many existing applications. And the best part is that there are working examples of this principle!

One of America's founding fathers, Thomas Jefferson, summarized the principle upon which a new distributed archive would operate:

...let us save what remains: not by vaults and locks which fence them from the public eye and use in consigning them to the waste of time, but by such a multiplication of copies, as shall place them beyond the reach of accident.⁹

Jefferson was speaking of ways to assure the preservation of American's founding documents, but a clearer and more succinct summary of the differences between the traditional and the new archive is difficult to imagine. In this spirit, LOCKSS ('Lots of Copies Keep Stuff Safe' www.lockss.org) offers a splendid

⁹ Thomas Jefferson. [1791] 1984. Thomas Jefferson to Ebenezer Hazard, Philadelphia, 18 February 1791. In Thomas Jefferson: *Writings: Autobiography, Notes on the State of Virginia, Public and Private Papers, Addresses, Letters*, edited by Merrill D. Peterson. New York: Library of America. Cited on the homepage of LOCKSS.

working example. In their own words, 'LOCKSS is open source software that provides librarians with an easy and inexpensive way to collect, store, preserve, and provide access to their own, local copy of authorized content they purchase. Running on standard desktop hardware and requiring almost no technical administration, LOCKSS converts a personal computer into a digital preservation appliance, creating low-cost, persistent, accessible copies of e-journal content as it is published. Since pages in these appliances are never flushed, the local community's access to that content is safeguarded. The accuracy and completeness of LOCKSS appliances is assured through a robust and secure, peer-to-peer polling and reputation system.'

Just as LOCKSS is a good example of a solution that makes use of distributed computing, other examples have tackled the problem of capturing the dynamic and extensive character of the web. Nearly synonymously with the appearance of the World Wide Web, Brewster Kahle initiated his remarkable Internet archive known as the *Wayback Machine* (www.archive.org). With over 80 billion web pages archived since its inception in 1996, the archive is composed of 'snapshots' of websites taken at regular intervals, allowing the interested researcher to go back in time and track changes. The previously mentioned 'discrepancy' in the White House Press Secretary's re-writing of a press release, as well as attempts to re-write the CVs of Enron's officials after their company's collapse, came to light thanks to this archive. But for the cultural historian, having access to the first decade of the World Wide Web's existence, and with it, the ability to track the emergence of social protocols, home page formats, early games, etc., is an invaluable resource. And the best part is that Kahle managed to capture the dynamic character of the web with a modest infrastructure and limited budget. With a mirror site in Alexandria, Egypt, the archive's discursive claim is loud, clear and, I think, not at all overstated. With a more fulsome budget, the archive could expand its operations and the frequency of its 'snapshots', but for the moment, its beauty is that it works efficiently and outside the entanglements of national governments and funding agencies. For once in our media history, someone has managed to both think ahead and act accordingly.

LOCKSS and the *Wayback Machine* offer two splendid examples of what is possible, one growing out of the efforts of progressive librarians and the other from an insightful and resourceful individual. Together, they have solved several of the problems unique to the digital domain. The issue of mapping links and social networking patterns has been taken on in the aggregate (social network mapping and analysis is a developed field, and can take on the face-to-face as easily as the digitally networked). The difficulty here turns on issues of privacy and the potentials for misuse, just as the problems potentially facing internet archives turn on issues of intellectual property. Social problems rather than conceptual or technological difficulties seem to be the order of the day, which is where an organization like UNESCO can be of great help.

Conclusions

Digital media have blurred relations between the once clearly demarcated realms of producers and consumers. As these digital technologies have become networked and have entered a state of high connectivity, a process that is roughly a decade old, we have seen the fast emergence of new social media forms. Social media have enabled new forms of collaboration, and they provide what Pierre Levy describes as a collective intelligence. In so doing, they have rapidly intensified the erosion of traditional and highly centralized organizations through a process of redefinition. Collaborative news networks compete with established journalism, in the process redefining the news from an institutional assertion of facticity to an act of critical engagement and a struggle for meaning. Music file exchanges such as KaZaa compete with the established recording industry, in the process transforming music from a commodity to be bought and sold, into communities of taste built upon distributed sharing. And open source software developers compete with the Microsofts of the world, in the process redefining software from a commodity to a collaborative language and community tool.

So too the world of the archive. The traditional archive has served as a social agent active in the reproduction of culture. By serving as a repository of what key institutions deem valuable, archives preserve cultural values and sustain hierarchies of social and cultural power. Given the very real limits of space and resources, difficult decisions have been made regarding whose letters to keep, which newspapers to store, and which sort of artifactual ephemera to hold. Not surprisingly, these decisions reflect an institutional sense of importance and relevance, which in turn largely maps onto the vision of the dominant classes. Little wonder that researchers seeking traces of immigrant life in the US at the turn of the century or marginal political and social movements of the present have such difficulty in finding relevant archival sources. But new archival practices that attempt to account for social media forms such as blogs, wikis, chat spaces, games, etc., enabled through distributed logics, enhanced storage capacity and accelerated transmission speeds, can redefine the archive from social agent to social practice. By embracing bottom-up dynamics, they will better reflect a wide range of social values, not just the ruling elite. By addressing cultural production that takes place outside of confines of corporate media, they will assume a much wider range of social granularity. And by taking advantage of the new affordances of digitally networked culture, they will encourage widespread participation.

Archives reflect the environments in which they are situated, and new environments require new policies. As I've attempted to argue, networked digital technologies and 'participatory culture' offer new challenges to both cultural producers and archives, generating new producers and users, new content and collaborators. But they also offer solutions that might be used to rethink the capacities and logics of the archive (storage, distribution, dynamic capture). The examples of the *Wayback Machine* and LOCKSS show what is possible, and provoke more traditional archives into considering these new possibilities.

Challenges, of course, remain. We need to think carefully about issues of privacy (for the archiving of network maps) and intellectual property (long overdue for an overhaul in the age of cut and paste). We must make careful distinctions in the weights of various cultural artifacts between, say, diplomatic documents and coke bottle collections, and develop different strategies for safeguarding and making both available. But above all, we must build upon the pioneering efforts of the few visionaries who have ventured into this new world, archiving the fast development of this dynamic cultural sector, and guarding against their selective and strategic use (or misuse) by those in positions of power, and make information accessible to all in accordance with the cultural logic of the age.

Addressing selection and digital preservation as systemic problems

David Bearman

The organizers of this conference asked this session: ‘how has the selection of various kinds of material been organized?’¹⁰ Their intention, I believe, was to suggest that different solutions be applied to different information formats, which is one of the assumptions reflected in the paper by Abdelaziz Abid in this volume. But the answer I will give goes beyond strategies appropriate to specific digital genres and formats to address the limitations of existing institutional structures and to suggest ways to move towards solutions that work for all current and future formats and genres.

I see four fundamental problems in our current situation: selection results in collection; digital preservation results in representations that favour aspects of the original perceived as critical within the different traditions of libraries, archives and museums; authenticity is being sacrificed by the timing of selection decisions; and the entire enterprise is currently organized in an uneconomical, inefficient and ineffective way. After exploring these dysfunctions, I propose a series of systemic solutions that I believe we must pursue to solve them.

Where we stand

Selection

Selection is presently conducted within individual repositories by staff trained in a particular approach to heritage, as librarians, archivists, museum curators, archaeologists and the like. They approach the question of selection based on their distinctive training and their understanding of the missions of their repositories, but they share the fundamental assumption that in selecting they are collecting, and in collecting they are adopting responsibility for the preservation of what they select. I will argue that the effect of these assumptions, which were oxymoronic in the world of analogue heritage, is perverse in the age of digital heritage.

Libraries select those books, videos, sound recordings, maps and other published items that they believe will be of most interest to their clientele, whether defined by locality or by speciality. The result is that most published items are

¹⁰ ‘Background & Programme’, Netherlands National Commission for UNESCO and the Koninklijke Bibliotheek, Conference: ‘Preserving the Digital Heritage: Principles and Policies’, The Hague, the Netherlands, 4-5 November 2005, http://www.unesco.nl/images/xprogramme_copy.doc

selected by many libraries and wait for the users of those institutions to request them. Most languish on shelves, while in some repositories the same item may be on a waiting list, being wanted by several patrons at once. By entering its collection, each individual instance becomes an issue for preservation to some individual institution, yet most libraries will be technically unable to preserve their copy. In a digital world, only one preserving institution need do so, and could do it universally for all that is published, within one country or internationally.¹¹

Archives do not select.¹² They transfer records created by the transaction of business within the organization for which they serve as an archive. Virtually all these records are digital now or will be soon, but unlike library materials, archival records are not fungible. Archives can't decide to keep different records; they must keep evidence of business activity, and preservability cannot be a criterion for whether something enters the collection. The problems archives face are that adequate digital records are not being created, they are not being transferred to archives in a timely fashion, and the means for transforming them from the formats in which they were created or stored to preservable formats are not widely available. Archives preserve a trivial portion of the totality of records and, in the current environment, the digital record they do preserve is not the most important part of that record, nor is it acquired in a manner that ensures its authenticity. Nevertheless, by transferring digital records to archives, archivists become responsible for their preservation, yet mostly they are ill-equipped for that task.

Museums and special collections acquire most objects in their collections as gifts from outside. These objects have typically been collected by someone else in the course of research, travel, or connoisseurship, or created by them in the course of their lives and deemed significant as reflections of that life. They are today, still, virtually always analogue, except in the special cases of museums of communication, software or popular culture, though some special collections are now acquiring the private papers of individuals which include digital artifacts. From what they are offered, museums and special collections select that which has meaningful associations with their mission, and like other repositories they then become responsible for its care. For them, preservability is a criterion in selecting, and can affect decisions, such as whether to acquire digital art as a medium of contemporary artistic expression or the papers of a scientist or politician that include large proportions of digital records.¹³

In theory, any of these institutions could select digital holdings in more 'pre-

11 By preserving all that is published within one country, however, we do not preserve all that the citizens of that country publish. And on the Internet, the result can be quite different from a record of national publication once might have been.

12 Many institutions called 'archives' include special collections of unpublished materials from private individuals and organizations. In some cases, the 'archives' may in fact only consist of such collections. In this case, they do indeed select collections from what is offered to them, and most of what is said here about museum selection applies to them.

13 Of course, the documentation that museums create about objects in their collections is now largely digital, but documenting collections is one of the business activities of the museum, and the results are, properly speaking, museum records of the institution that fall in the domain of the museum archives.

servable' formats, or immediately transform their acquisitions to such formats.¹⁴ As a practical matter, this buys them little more than a few years grace, however, even if they succeed. We cannot reasonably expect any software formats that we acquire today to be viable in a decade, and we do not know which formats will be 'better' to convert from ten (or even three) years from now. Libraries are limited in their choices anyway; in theory they are free to select the formats in which they wish to acquire content because, typically, they purchase it. But other consumers with less concern for long-term preservation may have a greater impact on what formats are on offer. Archives have tried to influence the formats in which the institutions they serve create records, and have undertaken large-scale transformations on receipt of records, but they cannot do much about the most difficult formats precisely because the system they belonged to is not standard.¹⁵ Museums and special collections can rarely influence the formats they are offered. We might say they cannot select them, by which we mean they should not acquire them. But this would be to accept that format rather than cultural significance should dictate what heritage is preserved.

The problem faced by all these institutions lies in the link between selecting and collecting. Ideally, if archives, libraries and museums could select what they felt was important to their patrons without being burdened with its care, what would they preserve?

Preserving what?

Some premises are now non-controversial.

- Digital preservation requires media conversion and format migration on an ongoing basis over time.
- Format migrations involve re-presentations of the original bit streams.
- The aim of preservation is to make new presentations that retain as many of the essential characteristics of the original as possible.

But in writing about digital preservation, we have to date failed to acknowledge, and therefore to adequately explore, the fact that professionals in different types of cultural repositories do not agree on what essential characteristics they feel must be preserved. As a consequence, we will probably fail to preserve our heritage even if individual institutions were relieved of the obligation, and the function of preservation was performed for them by a collective agency within the library, museum or archives community according to their common best practice.

¹⁴ The Victoria State Archives in Australia have adopted the most radical and forward-looking policy with respect to archival format transfers to its digital archives, requiring them to conform to the VERS Encapsulated Object (VEO) format. See <http://www.prov.vic.gov.au/vers/digitalarchive/>

¹⁵ This is nowhere more evident than in efforts to conserve digital art, reflected in Refresh: The First International Conference on Histories of Media Art, Science and Technology held at Banff September 28-October 3, 2005 and at this meeting in the paper by Rens Fromme of V2. Artists purposefully push the boundaries of standard practice to create irreproducible effects with novel code. But it is also characteristic of the most important scientific breakthroughs and novelties in popular culture – at first appearance, they are doing highly non-standard things, which is precisely what we want to document for heritage purposes.

Libraries are satisfied to preserve the information content of a historical publication for a future user without being too concerned that over time, through migrations, its presentation may become slightly different than the original, or worried about whether they could prove the precise lineage of their copy of the original digital object. Museums aim largely to preserve the experience that a user would have had at the time the work was originally made available, and are comfortable preserving a simulated system, even if incomplete, or the contextual props that convey the sense of the experience, even if its content is ‘preserved’ only as a highly realistic set accompanied by film and sound. In an archive, the link back to the original form is crucial. Without this connection, the content is useless as evidence, but if the archive maintains that connection to authentic data, it does not seek to preserve the ‘system’ by which the record was generated. In computing environments, the concept of functional migration is important, with the system as the focus, not the particular records which that system once produced. As long as the system operates as it did in the past, preservers of system functions are not concerned that the ‘records’ they have retained with the system reflect a snapshot of that systems data at a moment in time rather than a record in an archival sense.

To illustrate these differences concretely, I offer an anecdote: in the mid 1980s I played a computer game, on an Apple IIe computer, that introduced me to the economic and social world of France in the Middle Ages or early Renaissance.¹⁶ I recall, now somewhat vaguely, learning that the money I had accumulated was of no use in obtaining the bride I desired; it would have been far better for me to hold on to economically worthless lands and obtain a silly position in the court. If we wanted to ‘preserve’ this game, as a library, we could obtain the logical design together with the rules and options of play in ASCII text that will be easy to read even fifty years from now. If we wanted to preserve my game as a museum, a video of me playing the game and an Apple IIe computer would probably best document the experience, but the video would need to be migrated to new media and formats to be usable in fifty years. If we wished to preserve the game as an archive, the record of my actual ‘moves’ and their consequences would need to be preserved in a way that was directly tied to the actual game I played; these would be represented independently of the technology in which I played the game. In contrast, if we are interested in preserving the game as functional software, we might adopt a strategy of emulating its operating environment, permitting me to play the game again, but if I did this, it would certainly be with different moves than when I played then.

I acknowledge that these differences in preservation criteria, reflecting as they do differences in mission, are somewhat exaggerated in the depiction above. Nevertheless, even where moderated by community-developed articulations of best practices, they are real and they translate into quite distinct technical solutions

¹⁶ The game, I recently discovered referenced in an article by Willard McCarty, is called the Would-be Gentleman, see Carolyn Lougee, The Would-Be Gentleman: a Historical Simulation of the France of Louis XIV, *History: A Microcomputer Review* v.4, 1988, p.7-14.

to long-term digital preservation. Indeed, emphasis on content, form, context or function are mutually exclusive.¹⁷ Museums, libraries and archives save very different things because heritage professionals from different institutions within this sector do not agree about what it is that is the record. Hence, multiple strategies for preservation need to be implemented together to yield a complete picture of the past. It may be that such strategies will require more than one representation of a digital object to be created in the preservation transformations over time in order to capture different aspects of its 'recordness' and authenticity.¹⁸

Yet in most cases currently, since preservation is an intervention taken at the level of one individual repository, or at best by a library, archive or museum acting on behalf of other institutions of its type, the choice of how to preserve is grounded in the mission of the institution and its purposes for preservation. Each type of cultural institution could, and does now, acquire a record sufficient to its purposes without satisfying the purposes of the others. And our current best practices in making 'preservation copies' (which are of course not copies at all, but new presentations capturing what we think are the essential elements of the digital representation we are migrating or emulating) have the same limitations.

If we are to preserve heritage for all possible uses in the future, we will need to engage in a more self-conscious and catholic enterprise. First, we will be forced to rely on perspectives of authenticity and technical expertise from more than one of our cultural heritage sectors. And secondly, we will need to create and maintain multiple representations of digital objects to satisfy differing aspects of reliability and authenticity.

Representation and authenticity

At the heart of digital preservation discussions today is the assumption that certain representations of informational objects will fare better, or worse, than others over time and that we possess some valid insights today that could lead us to select representations with greater 'preservability'. The first assertion is certainly true, but the second is highly problematic, especially if the period of preservation is intended to be a century or more.

In some cases, it is evident that a specific format will have a longer lifespan than another, because we already see signs that support for one is already waning. Similarly we can identify representations that are today favoured, and are therefore likely to have significant market presence at a time when they would be threatened by obsolescence; the assumption is that this substantial presence will then be sufficient to cause the market to produce a 'migration path'. Yet, despite our widely held belief that we can distinguish fragile and robust representations, we have had little success in correctly identifying either state in the past. And the

17 David Bearman, Reality and Chimeras in the Preservation of Electronic Records, *D-Lib Magazine*, vol5 #4; <http://www.dlib.org/dlib/april99/bearman/04bearman.html>

18 The archival requirements are probably quite close to those the author and his colleagues expressed in 1996 in <http://www.archimuse.com/papers/nhprc/prog1.html> but similar documents do not exist for other perspectives.

future of computing beyond a decade or so is completely unknowable. Even if we did know how to make these distinctions accurately, cultural repositories would be poorly situated to act on the knowledge.

One thing we do know today, however, is that the authenticity of whatever transformations of our heritage we make over time are grounded in their documented genesis and the on-going documentation of their transformations.¹⁹ Regarding their genesis, digital objects must be captured at the time of their creation, and managed under rigorous controls, in order for us to assert that the original bit stream we have selected is indeed authentic.²⁰

Libraries do not have practices to establish the authenticity of their copies. They acquire what is 'published', and assume (reasonably but not provably) that their copy is identical to all others. Technology has advanced to the point that this assumption can be reasonably challenged, and I suspect libraries may, in the future, need to adopt mechanisms to validate the identity of what they acquire.

Archives are in a bureaucratic position to structure the record capture process up-front and satisfy the requirement for authenticity by acquiring records at the moment of creation. This would also ensure authenticity of the content and provide a foundation for trusted stewardship, which are two other major concerns. However, in practice archivists have had little interest in intervening at the moment of the creation of records.²¹ Indeed, digital archival orthodoxy at the moment argues that archival records do not even exist as such until they are purposefully kept by their creators and transferred to archival agencies.²²

Museums do not have the kind of relationships with the source of their acquisitions which would give them the same options to interfere at creation. The original media, rather than its information content, is typically what is collected, often long after it and the data formats stored on it are both obsolete. Museum content is more fungible than archive content, so they can decide not to acquire. But unlike libraries, museums can't just accept a different form of the same thing – if the object was digital in its creation, then the digital form is what a museum needs to acquire and what they will, in any case, be given. If historical artifacts

19 David Bearman and Jennifer Trant, *Authenticity of Digital Resources: Towards a Statement of Requirements in the Research Process*, *D-Lib Magazine*, June 1998 (ISSN 1082-9873) <http://www.dlib.org/dlib/june98/06/bearman.html>

20 David Bearman w/ Ken Sochats, *Metadata Requirements for Evidence*, 1996 <http://www.archimuse.com/papers/nhprc/BACartic.html>

21 As defined in the Society of American Archivists Glossary, <http://www.archivists.org/glossary/>, the University of Pittsburgh Electronic Records Project was 'supported by the National Historical Publications and Records Commission, the University of Pittsburgh School of Information Sciences has conducted a research project to examine variables that affect the integration of recordkeeping requirements in electronic information systems. This project was intended to examine one method to rectify such problems. The major objectives of this research project were to develop a set of well-defined recordkeeping functional requirements – satisfying all the various legal, administrative, and other needs of a particular organization – which can be used in the design and implementation of electronic information systems. The project also proceeded to consider how the recordkeeping functions are affected by organizational policies, culture, and use of information technology standards, systems design, and implementation.'

22 See definitions by INTERPARES <http://www.interpares.org/index.htm>

embody some digital component, they simply won't work after a while. This is typically not a central concern, as most artifacts in museums do not 'work' anymore, and are not 'used'; contextual information collected with them is adequate to their interpretation. In the case of informational artifacts, however, this clearly won't suffice. The few institutions worldwide that have dared to take on the challenge of curating digital artwork are leading the way here. At the time of acquiring the individual works, they have begun to explore with the artists what they consider essential to make decisions about preservation strategies.²³

If the genesis of our digital holdings was provably authentic, we would still need to document the on-going transformation process inherent in preservation. As an absolute minimum, any 'copy' needs to preserve the original bit stream along with all subsequent transformations. Most institutional practices in digital conversion today ignore the requirement of retaining the original bit streams in some encapsulated form after migration to a new format and representation. Few, if any, maintain adequate documentation of the algorithms used in subsequent format transformations. This is particularly problematic as the methods used are also not tested by externally validated standards bodies, deposited in 'methods' escrows, or made reversible.

I believe that too much of our current discussion and investment is going towards trying to define future standards and proactively migrate into those formats; this might buy us a decade of grace at the most. We should be ensuring the authenticity of what we acquire and documenting its technical dependencies so that future migration or emulation strategies can be applied. I believe that libraries, archives and museums would concur that digital content requires authenticity to be worthy of preservation, yet they have done less than is necessary to maintain documented control over digital objects from the moment of their creation, and to be able to retrieve the original bits and the conversion algorithms for all subsequent transformations

The socio-economic model for current digital preservation

Currently, our approach to the selection and preservation of cultural heritage is for each individual repository to take on the responsibility and incur costs itself. This means that each must support the costs of acquiring the digital artifact, possibly with costs in up-front conversion to a 'standard' format that it is better able to manage, and the costs of storage and maintenance of that artifact over time. Costs include determining the technical requirements for the care of the object, the risks incurred by its growing obsolescence, and the strategies for its on-going representation, migrations or emulation. Since there is a very large number of software formats, each institution needs to understand the requirements of each format it has collected and determine when and how to take actions to preserve

23 See, for example, the discussions in Sara Diamond, Participation, Flow, and the Redistribution of Authorship: The Challenges of Collaborative Exchange and New Media Curatorial Practice in *Selected Papers from Museums and the Web 2005*, p.213-232 and, online, at <http://www.archimuse.com/mw2005/papers/diamond/diamond.html>

it (and, of course, when to convert the media on which that data is stored). This model is, of course, precisely the same as the institution has for all its physical collections.

But digital artifacts are not physical, and, in a theoretical sense, there need only be a single ‘copy’ (excluding what is required for data security) preserved anywhere in the world for as many simultaneous users as wish to have networked access in the future. To measure the efficiency of our current system then, we could use as a base unit the costs of preserving such a ‘perfect preservation copy’, embodying all the best practices we know how to employ, so that it is technically and institutionally secure and available for both reuse and representation. There is a vast difference between this base cost and our true costs, because digital preservation decision-making is complex, and because the actual technical actions to conduct representations are subject to extraordinary economies of scale. In addition, the processes are fraught with the danger that they might fail, so the probability of failure by those not exceptionally well-equipped is very high. Finally, failure of the process usually results in total loss, so the social costs of not preserving are very high.

Most libraries, archives and museums will, as our society becomes more digital, select growing amounts of digital heritage to make available to their users. They will lack adequate resources, technical skills and systems to adequately preserve this content on their own. Given the costs of migration, and the uncertainty most institutions have about their migration decisions, it is not surprising that considerable proportions of their digital holdings are not migrated in a timely fashion and some become obsolete to such an extent that recovery is cost-prohibitive. Continued failure to take appropriate, timely, and effective preservation actions results and will continue to result in substantial losses of heritage.

Even that which has been subject to preservation action is suspect. Because multiple institutions are making determinations of when and what to migrate, different choices are made by different repositories. There is no independent verification of the validity of the transformation algorithms employed, or oversight applied to their understanding of what they are aiming to preserve. The representations created tend to reflect one definition of preservation to the exclusion of others. Where multiple copies of the original exist, multiple preservation decisions result in multiple, and probably different, representations, in possibly different formats, but not necessarily preserving content, context, form and function. It is quite likely none of these transformed records will be provably authentic.

Our current methods are wasteful, costly, ineffective and deficient. The proc-

esses we employ are not uniformly applied, not undertaken in a timely fashion, yield inadequately documented and irreversible results, and typically fail to preserve authenticity. Since redundant decisions are made at an institutional level, the inefficiency of this process from an economic perspective is matched only by its futility from a cultural and social perspective.

As the UNESCO Charter on the Preservation of the Digital Heritage says: ‘attitudinal change has fallen behind technological change’ (Article 3). Let me make a few modest proposals that address the root causes of this lag and which I believe could set us on a path to efficient and effective selection and preservation of our digital heritage.

Systemic solutions

The fundamental challenge we face is to move our efforts from the individual repository level to the systemic level.²⁴ Our habit of focusing selection and preservation in individual institutions is a consequence of the characteristics of physical heritage. But this approach fails when applied to the digital heritage. I believe that most of the solutions we have developed to date and envisioned as future solutions to problems of preserving the digital heritage will not succeed because they attempt to solve a systemic problem with fixes applied institutionally. The laudable effort to create trusted digital repositories, promoted by the Research Libraries Group and the Council on Library Resources, and reflected in this conference in the collaborative e-Depot project at the Koninklijke Bibliotheek, are efforts to address the inherent shortcomings in institutional practices by implementing peer review and community-audited practices to support shared preservation.²⁵ These are steps in the right direction, but I feel they fall short of what is required.

We need to pursue new approaches in three areas simultaneously: the way the network is governed, the legal conditions for preservation, and new economic models.

Network services and selection in a systemic reform

The most radical reconceptualization of the current environment will be to move functions that are being performed ineffectively within institutions to the network infrastructure. This will convert the network as we know it from a simple carrier of signals to a regulated service platform, not unlike the shift that takes place when a small company moves from an ISP to an ASP. The means to this end may differ from one nation to the next. Where the network is controlled by a country, as in some totalitarian regimes, it could simply be dictated. Where private

24 Elements of this strategy are reflected in recent reports on the Cyberinfrastructure for science (US National Academy of Sciences) and for the humanities (American Council of Learned Societies). See <http://www.acls.org/cyberinfrastructure/acls-ci-public.pdf>

25 U.S. National Archives and Records Administration (NARA) and the Research Libraries Group (RLG) draft audit checklist for certifying digital repositories (http://www.rlg.org/en/page.php?Page_ID=20769) Presentation by Erik Oltmans on the ‘International E-Depot’ delivered at this conference.

enterprise runs the backbone services and ISPs, and companies or individuals manage their own internal networks, a combination of communications regulations and incentives and liabilities must be created to drive the process. In either case, the result would be the same: in the future, redundancy of heritage materials would be governed by the mathematics of data security requirements rather than by institutional selection and all digital heritage would be managed, migrated and preserved systemically, regardless of where it physically resided and without necessary knowledge or intervention by its custodians.

Elements of the current system are already undergoing change in the case of digital selection. The system of licensing digital versions of journals, for example, means that each institution that selects to offer its patrons access to periodical literature in digital form does not actually acquire a digital copy of that publication, only the right to have access to it. This is a partially successful adjustment of the system for selection, although it fails in respect to preservation because the subscribing institutions are not assured that the publishers will in fact be able to keep the materials to which they have subscribed available to them over time.²⁶ It also fails the publishers and the society in an important respect because it requires the institutions to decide upfront to privilege their patrons by buying a subscription, rather than rewarding the patrons, the society and the publishers by making all content available to any user coming through any institution and returning payments to the publishers based on that use.

In a future model of digital selection, the users, rather than cultural heritage institutions, would select what they wanted to read, view or listen to, and the library, archives or museums would help provide access, context and interpretation.²⁷ Instead of spending money on individual institution-based selection, resulting in many copies of some things and none of others, it makes more sense to acquire every published digital object, thereby eliminating the huge overhead to society of ineffective selection processes. There would only be one logical copy (though for information security reasons this would in fact be a small number of distributed copies backing each other up and for network efficiency reasons might involve some topographic distribution based on demand). Rather than trying to apply traditional archival methods of appraisal of records, archives define algorithmically what records will be retained for how long, after capturing all transactions at the time of transmission.²⁸ In both cases, the systemic answer

26 As I set out for the meeting, the Association of Research Libraries issued a statement 'Urgent Action Needed to Preserve Scholarly Electronic Journals', http://www.arl.org/osc/EjournalPreservation_Final.pdf. Since the conference at which this paper was given, Portico was announced as a joint venture of the Mellon Foundation, Library of Congress, JSTOR and Ithaka Harbours (<http://www.portico.org/>) – it seems to be built on similar principles to those I advanced here, though early announcements were not completely clear.

27 I was delighted to hear at this meeting the idea of users making selections based on universal capture of a digital resource echoed by Catherine Lupovici of the BnF, reporting on conclusions arrived at by the International Internet Archiving Consortium at its most recent meeting.

28 David Bearman, 'Item Level Control and Electronic Recordkeeping', *Archives & Museum Informatics*, vol. 10, #3, p.195-245 <http://www.archimuse.com/papers/nhprc/item-lvl.html>

is more cost-effective from the societal perspective and easier to implement than selection, which is labour intensive and ineffective. Centrally managed network accessible storage would be less expensive than locally managed institutional storage, and would provide much greater guarantees of access. Local institutions would shape their services and support, rather than their holdings, to fit their missions. With the savings, we can then get on with projects to construct a preservation infrastructure.²⁹

Future publications

The mechanism for implementing a systemic solution for future digital publications has long been at hand and is even implemented in some countries, though not as part of the selection solution: it is mandatory legal deposit, covering all publications including digital publications, whether in print, sound, video, still image or other modalities. Correctly organized, legal deposit systems can become the gateway to access all heritage under copyright. Although it is true that this would involve a massive renovation of the existing structures for processing and, of course, using, legal deposit material, in the aggregate it would be hugely more efficient than local acquisition, or even institution-based subscription and licensing. It would also address the issue being raised by ARL libraries just now about continued access to electronic journals, without resorting to contracts to require escrow copies to be deposited.

Digital data that is not deposited, but published anyway would be deemed to lack copyright protection, as is the case, de facto, of much web-based content today. It would be harvested by patrimonial archives or national libraries and stored centrally for purposes of preservation. Access would be enabled by those who chose to make the content publicly available. However, if withdrawn from public distribution by its creators, the national digital archives could make the archival record available, thereby ensuring that all data made public on the network, whether desiring copyright protection or not, could be selected by users.

Existing analogue heritage

We currently curate a massive historical heritage in our many cultural repositories. It is a great pride of our culture, though constantly threatened by the normal processes of decay and by war, fire and flood. We can appreciate that the representation of this heritage in digital form would open it to a vastly broader audience than can now obtain access and could in many cases save our heritage from inevitable destruction.

Very recently, digital scanning technologies and storage costs have evolved to a point that it is well within the capability of the world's national libraries, indeed well within their current budgetary limits, to plan for the digital capture

²⁹ Some savings are also achieved in storage, as the peta-box at the Internet Archive demonstrates with its \$2000/TB storage costs, though distributed systems using unutilized PC storage space could at this point be even cheaper, if exceptionally complex to administer, as the LOCKSS project has demonstrated.

of all print in analogue form. The imagination of many librarians was captured by the Google Print announcement last December that it intended to do just that, and more recently by the Internet Archive's announcement of the Open Content Alliance. The technical means to capture all print in a very short time – within a decade – is here. The costs of massive scale digitization would be easily offset by the savings to individual institutions holding print that they could, in many cases, dispose of without inconveniencing their patrons, but have been holding, and building new storage areas for, because of their sense of institutional responsibility for its preservation. UNESCO has the opportunity at the upcoming WSIS or in other venues to assert the responsibility of each nation to convert its print patrimony to public domain images³⁰ to ensure the preservation of languages and culture and widespread and equal access to educational resources. National libraries could immediately begin the short-term, non-selective, conversion of their analogue print holdings (recognizing copyright issues that are addressed below).

Significant progress is already being made in the conversion of sound archives. For technical reasons, digital conversion makes great sense and is taking place on a large scale, but not systematically. A similar commitment to the short-term conversion of all analogue sound recordings is practical and necessary. It could start in most countries with the creation of union databases of sound recordings which are currently lacking.

Vast stores of photographs and film, and of unpublished manuscript materials, in various formats present a greater challenge, in part because we do not know what they are, in part due to the amount of storage they require and, for film, because we are less comfortable with the current digital formats in which we would make the initial representations. These conditions should change in the coming decade, making the project of capturing analogue images one that could be undertaken systematically when the print and sound have been successfully transformed. In each case – manuscript, archives, still images and motion images – the projects will require union metadata bases that can be constructed in the coming decade as a foundation on which to build the digital collections.

The current repositories of analogue records need not lose their identification with the digital; they can always be associated as the collecting repository and the digital item can be 'branded' as theirs. Nor does capturing a digital representation necessarily have implications for their continued custody of the original. It represents a sort of insurance policy, and it enables broader access, but it may or may not recommend any action for the original. Sometimes it will make it possible to store the original in a more efficient facility, relieving space in the repository.

30 Public domain images of print is a politically charged issue, but an essential component of a digital strategy as it enables value-added services to be built on top of the base level of digital scanning that is done by the National Libraries. The concept has been further developed in David Bearman, 'Converting Scanned Images of the Print History of the World to Knowledge: A Reference Model and Research Strategy', *UNESCO Between two Phases of the World Summit on Information Society*, St. Petersburg Russia, May 2005, Institute of the Information Society, Moscow, 2005, p.423-429; and http://confifap.epic.ru/conf2005/eng/info/pr_s11_en.htm#ms11_2.

Occasionally it may make it possible to dispose of the original as we have always done with nitrate film stock once converted. But in itself, digitization has no necessary consequence for custody of the physical object. However, the digital object should be managed, migrated and preserved centrally.

Future records

If the truth be told, virtually none of the digital record generated by citizens and institutions in our society over the past fifty years is secure. Tiny portions of it have been deposited in archives. These are not necessarily the most important parts of it, and they do not protect needs of the society for records of on-going transactions of government or private business. Insofar as we can speak of a system, it is in a total shambles. Selection is virtually haphazard and infinitesimal. Preservation is largely supported by wishful thinking.

In 1994, I proposed a Reference Model for Business Acceptable Communications, which while flawed is still the best framework I know for automatically archiving all transactions at the point of transmission and maintaining authentic versions over time. Augmented by what we know from the OAIS model and from LOCKSS, it provides a technical framework for thinking about a new structure for acquiring future records.³¹ In this new structure, utilities embedded within the network would be responsible for capturing all records. By establishing networks that support Business Acceptable Communication, the society could ensure that all records were created and kept, though of course the legal system would still need to determine which records had to be revealed for what purposes before they come into the public domain. Alternatively, record creators could be empowered to administer record retention decisions. But since national laws on recordkeeping by private and public entities would then be fully realizable, record creators would incur liabilities for negligence if it were found that they had disposed of records that should have been retained. Today, as no one can provide an adequate solution to electronic recordkeeping, records creators are excused for their failures.

The legal system

Digital preservation 'copying', exemption from copyright

In addition to mandatory legal deposit, only one modification to existing copyright laws need be made to implement the systemic approaches discussed to this point and to enable some of the new economic mechanisms proposed subsequently. A digital preservation copying exemption must be created internationally to permit the conversion of analogue sources, and migrations and representations of digital materials in all formats and modalities. Since digital preservation per se does not threaten any economic interests of copyright holders, and in fact serves

³¹ See Reference Model for an Open Archival Information System (OAIS) <http://www.ccsds.org/documents/650x0b1.pdf> and LOCKSS model <http://lockss.stanford.edu/>

their interests as it ensures that copies will be preserved during the period of copyright protection, this should not be politically impossible in any jurisdiction. But it needs to be introduced now.

Many other proposed modifications to copyright laws might bring more heritage into the public domain, enable a variety of value-added uses, and serve other interests that have social benefits, but since they are not essential, and are highly politicized, it is best not to risk the necessary change in order to make unrelated modifications that will not have as strong a public interest foundation.

Legal deposit requirements for copyright could ensure that published material existed in a national library. Universal web crawling could ensure that a copy of public information on the Internet was saved. By accepting responsibility for eventually putting all legally deposited publications online, national libraries could, in association with web archives, create the conditions for future digital preservation of cultural heritage.

Mandatory format specification to trusted standards bodies

One of the major challenges in managing electronic records, whether in archives or in current records, is to migrate from obsolete formats (media and software) to currently supported representations. Several years ago, I criticized proposals to emulate software as a means of preservation, in part because format specifications on which these emulations depended were proprietary.³² I have since come to believe that this is, to a lesser although real extent, also a problem for successful migrations. If society extended the concept of liability to companies that allowed formats in which clients had previously stored their data to become obsolete, accountants would ensure that companies made the details of their format encodings public after the software that created those formats was retired; the alternative would be a growing weight of liabilities attached to the firm as its products evolved. This extension of liability could be achieved by new legislation or by the application of existing law, through civil cases initiated by other firms, by the government as a consumer, or by class action where available. It could also be implemented through a change in accounting principles that required firms to record risks from software obsolescence as liabilities.

More direct approaches are always possible, but might prove politically unachievable. Using models currently applied to the registration of drugs, national regulators could require specifications to be filed for software in order to approve its sale within their jurisdiction. Alternatively, national legal deposit obligations could be legislated for source code in order to obtain copyright, patent and trademark protection, with assurances that the code would be made accessible only to International Standard Bodies developing algorithms for format migration in the event that the original formats are no longer supported by the owner of the intellectual property.

³² David Bearman, Reality and Chimeras in the Preservation of Electronic Records, *D-Lib Magazine*, vol5 #4; <http://www.dlib.org/dlib/april99/bearman/04bearman.html>

Virtuous viruses

If existing standards bodies established mechanisms to validate transformations from one format to another, they could become the trusted repositories of the proprietary and trade secret information revealed by depositing format specifications. They could use this information for the sole purpose of developing certified migrations or emulations that would be placed in the public domain.

The cost of on-going format migration only seems prohibitive when viewed from the perspective of one institution. Because each content owner or repository currently keeps their content themselves, they are responsible for determining when the content needs to be migrated and how to do it. Yet most repositories lack the technical expertise to make such a determination or to execute a successful migration, as is evident from the complex requirements recently published for 'trusted repositories'³³. But if content is available online migration could take place, with appropriate permissions, for all content in a given format without further investment by the content owners other than providing adequate storage space and permission. Technically, a unique means to provide for universal registry of digital content such as DOI, a mechanism like OAI-PHM for harvesting, and a format migration strategy built on networked background format conversion, could go a long way towards solving the problem. A repository would, in effect, subscribe to a content migration service. In theory, such services could be commercial offerings, but in practice if there was one such service made available as a public good there would be no commercial market, so format migration might need to be provided as a universal network function. The service would have two roles – protecting the digital archive from accidental destruction by storing multiple copies and protecting it from obsolescence by format migration.

I propose the systematic creation of virtuous viruses, propagated by standards bodies which would certify format migration algorithms and a mechanism. Repositories could 'subscribe' to virtuous viruses that would harvest content in one format and transform it using the certified methods. Such virtuous viruses would be built on a model proposed by Clifford Lynch, in 1999, as 'canonical objects'.³⁴

Expanding the public domain

There is no necessary connection between wider access to digital heritage and the expansion of the public domain, though bringing more content into the public domain makes aspects of the administration of access easier. Yet there is, in some cases, a synergy between institutions being created to extend the public domain and the strategies for digital heritage selection and preservation that I am proposing here.

33 Research Libraries Group and National Archives and Records Administration, An Audit Checklist for the Certification of Trusted Digital Repositories. Draft for Public Comment, August 2005. <http://www.rlg.org/en/pdfs/rlgnara-repositorieschecklist.pdf>

34 Clifford Lynch, Canonicalization: A Fundamental Tool to Facilitate Preservation and Management of Digital Information, *D-Lib Magazine*, September 1999, vol. 5, #9 <http://www.dlib.org/dlib/september99/09lynch.html>. Also Ockerbloom, John, Mediating Among Diverse Data Formats, CMU-CS-98-102, Ph.D. thesis, Carnegie-Mellon University, January 14, 1998 <http://www-2.cs.cmu.edu/People/spok/thesis.ps>.

For example, one means of structuring services that enable institutions to use remote storage and preservation is to put this support function in archives that are in the public domain. Thus obtaining such services would be a further incentive to donate content to open archives. This is one of the attractions of the recently announced Open Content Alliance. Digital heritage that is already in the public domain, or that rights owners are willing to make available on certain terms, is stored and cared for by the Internet Archive, one of the most trusted institutions for preserving and providing access to digital heritage. Libraries can contribute converted analogue resources and thereby build a collective resource that is far greater than what they could have afforded to scan on their own, and which as it grows indicates by its lacunae the works that still need conversion. Publishers and authors can donate materials whose copyright has not expired if they are willing to allow copies to be made under certain circumstances, again in return for broader access and preservation. While many details of governance and operations still remain to be worked out by the charter members of the Open Content Alliance, it seems likely that institutions will benefit by having 'their' collections seen as branded by them, and that value-added information service providers, and therefore clients interested in value-added services, will benefit by having rights to reuse content.

In 1997/8, when I was drafting the business plans for the Knowledge Conservancy, we envisioned the extension of tax benefits for the donation of virtual intellectual property or establishment of 'public rights of way' as the primary incentive for publishers and other rights holders to contribute to a public trust.³⁵ The success of the Creative Commons made it clear that some owners would be willing to contribute rights without such direct benefits. Without amendments to existing copyright laws, the Creative Commons was able to craft licenses that gave limited exploitation rights to users who add value by their representations of objects in the public domain, thereby rewarding that activity, but not giving it additional copyright protection beyond that desired by the original content owner.³⁶ Intellectual Property Conservancies which provided some economic incentives could further advance the public domain.

The recent Association of Research Libraries white paper on E-Journals proposes a contract based mechanism that would require e-journal publishers to deposit copies of their works in a kind of escrow, administered by the subscribers.³⁷ One way to implement this is for the institutional sponsors of the virtuous format migration viruses to simply archive the historical format and reformat the contents of such repositories, rolling the newly reformatted content, and the original versions, forward.

35 David Bearman, Intellectual Property Conservancies, *D-Lib Magazine*, December 2000, vol6 #12; <http://www.dlib.org/dlib/december00/bearman/12bearman.html>

36 Efforts are underway now to craft a Creative Commons-like license that would enable art museums to license for personal use, content in which they wish to reserve their commercial rights. If successful, many major museums have agreed to make significant bodies of their collections available on the public web.

37 Statement of the Association of Research Libraries 'Urgent Action Needed to Preserve Scholarly Electronic Journals', http://www.arl.org/osc/EjournalPreservation_Final.pdf.

Many other efforts to expand the public domain, some already underway – such as e-print archives – and others yet to be imagined, could make use of systemic solutions to selection and preservation proposed here. But we cannot stop with the heritage that is in the public domain, no matter how successful we are at expanding that status.

New economic models

Ultimately we need to develop new economic models that will ensure that all heritage, regardless of its intellectual property status, is preserved.

Systemic preservation solutions are also likely to be built on new socio-economic models. Many of the social constructs we require have already been invented or proposed; a few others will need to be created, but the methods by which they would operate are not novel or untried. For example, those desiring format migration services could subscribe to backup archiving as well, for example, by agreeing to provide locally 150% of the online storage required by their own archive. This storage capacity could then be used to store an encrypted, encapsulated and compressed version of someone else's archives, in return for which theirs would likewise be stored in two or more other locations in the event of a need to recover it. Obviously the precise definition of such new services and economic relations would evolve.

Use-based payment to rights holders

National use-based payments could, and for social reasons probably ought, to replace institutional subscriptions negotiated on a copy-based pricing model and not available to most users. In this way, the library system (and book and video stores printing and writing DVDs on demand) could pay publishers by the numbers of copies in circulation; we don't need additional physical copies in places where they aren't being used and waiting lists where they are wanted. The size of the unit being licensed may change from the book to the chapter, as it has in the digital music realm where the 'tune' rather than the album has become the unit of value. But this doesn't need to matter on the whole as long as a healthy market can be built on whatever granules of intellectual property are being traded.

A model for such national licensing may be the Public Lending Right as it is administered today in the United Kingdom, Australia and Canada. Under this policy, rights holders are paid a small amount for each public library loan of an item of intellectual property in which they have an interest. As currently implemented, the public lending right does not pay for loans from libraries other than the public system, and it makes payments for material that is in print or out of print equally as long as they are covered by copyright. In a revised and extended form of the public lending right, payments could extend to all uses of digital publications under copyright regardless of the lending institution, thereby shifting to use-based payments on a national level and enabling publishers to earn substantial income. At the same time, new public lending rights laws could address the

management of material in the current status 'out of print/in copyright'. This will become an historical oddity as there is no reason to permit anything to go 'out of print' in an era of on-demand copying. Perhaps some lower payments would be established for the out of print material, in return for which governments would make digital copies and keep them available to users.

Funding systemic mechanisms from existing budgets

Moving towards solutions that break the link between institutional selection and institutional custody and control will not be easy. It will require new sources of funding, and will require existing sources of funding to operate in new ways. In the 21st century we still need to craft solutions within a nation state context that we inherited from the 19th century, so the answers we arrive at may take a different form in different countries, reflecting the ways that the heritage sector, and its individual types of institutions, are funded, regulated and administered.

The opportunities for shifting funds to new ways of doing things are all around us. In the past year, I have been exploring digitization of Canadian publications, and discovered that the National Library of Canada currently pays the Canadian Postal Service a multi-million dollar library book rate subsidy which would be partly offset if digital loans were enabled. Interlibrary loans in Canada, according to recent studies by the Canadian Association of Research Libraries (CARL), cost the lending libraries nearly CDN\$80 per loan – sums that could be offset by digital lending with potentially sizable payments to publishers included!³⁸ Even such routine activities as reshelving books, estimated at the University of Toronto to cost over \$2 per book, might be significantly reduced by having digital copies available for those simply wanting to quickly browse for information, again liberating money that might pay for such privileges. Similar existing funding that could be saved in a digital system can probably be found in other national contexts.

We can succeed in finding both funding and mechanisms to achieve reliable digital preservation if the community of heritage professional understands the need and commits itself to finding solutions. The 'top-slicing' tradition of British sectoral bureaucracies has given rise to the Joint Information Systems Committee of the Higher Education Funding Authority, which has emerged over the past decade as one of the most powerful forces for collective change. On occasion, the NEH, NSF, IMLS and NIH in the United States have been able to exert similar influence, an excellent example being the current NIH policy requiring recipients of government medical research funding to deposit their research results in public domain archives.³⁹ By fostering collective mechanisms for retrospective digiti-

38 CARL report on ILL, NAC budget for postal subsidy, and Carole Moore report on U of T reshelving costs are among a number of such measures cited in the strategic recommendations by Archives & Museum Informatics for the Library and Archives of Canada, summer 2004.

39 In the United States, the NEH, IMLS, and NSF funding guidelines have often established expectations for applicants for Federal grants. Most recently, the National Institutes of Health published draft rules (Sept. 3, 2004) that would require any grant recipient to make copies of their articles available for free to the public.

zation, such as HEDS or taking up the recent Internet Archive offer to digitize books at US\$0.10 per page, the heritage sector could introduce standards for the deposit and care of such assets that moved responsibility out of the hands of individual repositories towards collectives more able to support the requirement effectively and cost efficiently.⁴⁰

Conclusions

The original question, ‘why are selection and preservation an issue?’ has an unfortunately simple answer: the link between selection and custody, and between preservation and institutional action has put our digital heritage at great risk and prevents us from implementing effective solutions to its long-term care. We need to adopt a systemic approach to heritage preservation.

On a practical level, the changes that are proposed are quite radical and conceived on an international level, though they can be implemented on a national level, and will largely be realized through shifting the direction of funds currently administered at a local level. If adopted, these solutions could generate vast societal cost savings and achieve higher levels of heritage preservation. In addition, except for the modest exemption for digital preservation copying, the proposed enhancements in network services and the development of new economic models can go ahead without any changes to existing law.

⁴⁰ The Internet Archive has offered to digitize books for US\$0.10 per page. This figure is being substantiated by a University of Toronto pilot project undertaken in 2004/5 which, though it has not yet actually achieved this rate, is progressing towards it and is expected to deliver at this cost soon.

Preserving the digital heritage: roles and responsibilities for heritage repositories

John Mackenzie Owen

Keys aspects of preservation

Preserving the digital heritage is a complex process involving many different actors. In order to clarify this issue with respect to the current situation, it is useful to distinguish between three important dimensions of preservation: functions, responsibilities and funding (table 1).

Functions are the various activities that need to be carried out in order to preserve digital heritage materials. The main functional categories are the selection and storage of data (e.g. in repositories and archives), maintenance (e.g. conversion, migration and emulation), and services provision (e.g. indexing and access services).

The responsibility for performing one or more of these functions may be held by different actors, notably the originators of heritage materials (e.g. authors and publishers), stakeholders, i.e. those who have an interest in using heritage materials (e.g. professional societies, universities), heritage institutions that have a specific mission to preserve heritage materials (e.g. libraries, archives and museums), and finally, service providers within the information industry who may wish to perform functions on a commercial, for-profit basis.

The third dimension is related to the funding of preservation functions as performed by the various actors. This points to the important distinction between actors responsible for preservation activities, and funding agencies. Funding digital heritage preservation may be seen as a responsibility of the original 'owners' of the materials, i.e. the copyright holders. However, it is usually seen as a public, government responsibility

These three dimensions of digital heritage preservation allow us to bring into focus the responsibilities for performing specific functions and for funding these functions (table 2).

At the present moment, the selection and storage of heritage materials is the primary responsibility of government-funded heritage institutions. There has been some discussion about leaving this responsibility with the originators and copyright holders. However, many arguments can be brought against this, including the fact that originators may cease to exist, that copyright expires, that it leads to fragmentation of the canonical collection, and that originators would have little financial incentive to guarantee continued maintenance and access. Where data collection and maintenance therefore would have to remain the responsibility of government-funded institutions, a case could be made for shifting the respon-

Table 1: three dimensions of digital heritage preservation

| | |
|--------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Functions | Data: selection and storage (repositories, archives) Maintenance: conversion / migration / emulation Services: access and indexing |
| Actors | Originator (publisher, author) Stakeholders (e.g. professional societies, universities) Heritage institutions (museums, archives, libraries) Service providers |
| Funding responsibilities | Copyright holders Government (via heritage institutions) Users |

sibility for indexing and access more to the private sector. This would allow the information industry to develop products and services based on heritage collections. This would also imply that funding of indexing, access and other services would be shifted towards the user through payments for specific products and services. Finally, a case could even be made for shifting the funding responsibility from the collective, government level to a more market-driven model where users and stakeholders are made more responsible for funding all aspects of heritage preservation.

Digital heritage objects

The previous analysis is based on the current situation, where a limited number of types of institutions are responsible for heritage preservation, notably libraries, archives and museums. These are institutions that have held this responsibility in the pre-digital era, and are now adapting to the preservation of digital materials. However, the main point to be made here is that digital heritage materials are not just traditional heritage materials in digital form. Many traditional materials, such as publications, archival documents and artworks are now being produced in digital form. Traditional heritage institutions have little or no problems in dealing with the transition to digital versions of these materials. But if we look at the digital world around us, we see many more types of objects than those that are traditionally preserved in museums, archives and libraries. To give a few examples:

- Personal and institutional websites
- Discussion lists
- Email, SMS, MMS, IM (= Instant Messaging)
- Blogs
- Spam
- Private and professional correspondence
- Photographs and films
- Music
- Radio and TV
- Podcasts and videoblogs
- Social software applications
- Games

Modern society is characterized by a multitude of digital forms that far ex-

Table 2: functions – actors and funding

| | | FUNDING | | |
|----------------------------|-----------------------|-------------------|--------------------|------------------------------------|
| | | Copyright Holders | Government | Users |
| A C T O R S | Originators | Data(?) | - | - |
| | Stakeholders | - | - | Data, Maintenance, Indexing/Access |
| | Heritage institutions | - | Data & Maintenance | - |
| | Service providers | - | - | Indexing & Access |

ceeds what is preserved in the majority our heritage institutions, even when they make the shift to digital materials. What is also interesting about these materials is the originators of all these digital objects are not only publishers, commercial companies, government and educational institutions, but also, and increasingly, private individuals. Creating digital objects in cyberspace can be considered as the normal way in which society expresses itself both at the institutional and at the individual level. This is currently true for advanced societies, and could increasingly become so in less advanced societies as well.

In fact, these digital artifacts constitute what may be described as the digital fabric of society. It is through this digital fabric that our culture expresses itself, and it is therefore this fabric that constitutes the cultural heritage that needs to be preserved.

Characteristics of digital culture

The digital objects that represent and preserve contemporary culture have a number of special characteristics that distinguish them from traditional heritage materials. These include:

- Global provenance: any digital object may contain components from any number of locations worldwide. The question of an object's 'nationality' often makes no sense.
- Heterogeneous: it is no longer possible to classify heritage materials in a limited number of distinct media, types and genres: in the digital world these exist in multiple and changing combinations; distinguishing heritage institutions along traditional lines is no longer feasible.

- Fluid / dynamic: digital materials are often not static objects that can be stored ‘as-is’, but dynamic documents that change over time and therefore have a ‘history’ that is also part of the digital heritage.
- Interactive / collaborative: digital materials are often interactive and are constructed in a collaborative fashion; interaction and collaboration are therefore essential aspects of the materials themselves; these social aspects, too, need to be preserved.
- Interrelated: meaningful and multiple links exist between data objects; these links are integrated components of the heritage materials.
- Fragmented: the user experience is not confined to a single object at a time, but is constructed out of multiple fragments from different sources and of different types.
- Embedded metadata: digital materials include descriptive and contextual metadata that are embedded within the materials themselves, and that form the basis for future services (e.g. indexing and access) and interpretations.

Perhaps the most significant consequence of these characteristics is that modern culture is represented by the use of digital materials and the social and cultural processes they invoke, rather than by the materials themselves. Heritage preservation, therefore, implies not just storage and maintenance of digital artifacts, but the capturing of dynamic processes and patterns of use.

Digital heritage repositories

It would seem, therefore, that there is a large gap between what traditional heritage institutions preserve, and what should be preserved in view of the characteristics of our present-day digitized society. This is not just a case of what and how these institutions preserve, it is also more fundamentally related to the structure of the entire heritage preservation field. Digital objects of a specific type are embedded in a networked environment and cannot be separated from other object types. It is no longer possible to share the responsibility for preservation amongst specific institutions such as libraries, archives and museums. Also, the digital context is not defined by national production, but what individuals select from global resources. Preserving materials on a national basis, as is the case with most ‘national’ libraries, archives and museums, will preserve nothing of the information environment in which the information actually exists within a society.

The culture, standards and values of traditional heritage institutes are not suited for preserving the ‘digital fabric of society’. They are too much focused on local production, high culture, official documents, traditional materials and ‘business as usual’. Traditional heritage institutes are too inward-looking, defining the digital world in terms of the institution instead of defining the institution in terms of the digital world. They also lack the required understanding of digital culture and the skills that would allow them to handle the new digital materials. The result is that future generations will see the 21st century through the 20th century eyes of the heritage institutions that have failed to make the transition to an entirely new definition of heritage materials required by the digitization of society.

Finally, close cooperation between public sector (data collection & maintenance) and private sector (indexing and access) is inevitable in order to maintain the fundability of heritage preservation. This too points to the necessity of a fundamental restructuring of the heritage preservation field.

A solution to this problem could be the establishment of a new type of heritage institution for digital materials, together with the recognition that existing heritage institutes should continue the important, but limited task of preserving our pre-digital heritage, especially with respect to 'high' and 'official' culture. The new type of institution, that tentatively could be called the Digital Heritage Repository, can be characterized as follows:

Memory institution for the digital society, encompassing the full range of digital materials (and their underlying processes of creation and use) through which society expresses itself (the 'digital fabric of society')

- New institution alongside archives, libraries and museums
- Based on usage and social processes rather than on national production
- Based on publicly debated selection criteria
- Based on public funding of collection, storage and maintenance
- Closely cooperating with the private sector for indexing and access services

Conclusions

We currently have specific and well-developed heritage institutions for 'high' culture, government and scientific information: museums, archives and libraries. These institutions are well on their way towards digital preservation of the types of materials they have always preserved in non-digital form. However, digital is a characteristic of almost all forms of cultural expression, well beyond the boundaries of 'high' culture: digital objects are the fabric of contemporary society. To preserve this aspect of the modern world, we need a new type of heritage institution that will focus on the wide variety of digital materials and their contexts as they are actually used in society. Creating this new institution requires an 'outsider' approach, and should not be left to the traditional heritage organizations. Since preserving the digital heritage will have to be a joint public-private responsibility, involvement from the information industry has to be called for. Eventually, this should lead to the establishment of what I have called here 'Digital Heritage Repositories'.

Appendix A: Charter on the preservation of the digital heritage⁴¹

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 access to this heritage will offer broadened opportunities for creation, communication and sharing of knowledge among all peoples,

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,

Proclaims the following principles and adopts the present Charter.

THE DIGITAL HERITAGE AS A COMMON HERITAGE

Article 1 – Scope

The digital heritage consists of unique resources of human knowledge and expression. It embraces cultural, educational, scientific and administrative resources, as well as technical, legal, medical and other kinds of information created digitally, or converted into digital form from existing analogue resources. Where resources are 'born digital', there is no other format but the digital object.

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 ever-growing heritage may exist in any language, in any part of the world, and in any area of human knowledge or expression.

⁴¹ Adopted at the 32nd session of the General Conference of UNESCO, 17 October 2003

Article 2 – Access to the digital heritage

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

Member States may wish to cooperate with relevant organizations and institutions in encouraging a legal and practical environment which will maximize accessibility of the digital heritage. A fair balance between the legitimate rights of creators and other rights holders and the interests of the public to access digital heritage materials should be reaffirmed and promoted, in accordance with international norms and agreements.

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. Member States will benefit by encouraging legal, economic and technical measures to safeguard the heritage. 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.

Article 5 – Digital continuity

Continuity of the digital heritage is fundamental. To preserve digital heritage, measures will need to be taken throughout the digital information life cycle, from creation to access. Long-term preservation of digital heritage begins with the design of reliable systems and procedures which will produce authentic and stable digital objects.

MEASURES REQUIRED

Article 6 – Developing strategies and policies

Strategies and policies to preserve the digital heritage need to be developed, taking into account the level of urgency, local circumstances, available means and future projections. The co-operation of holders of copyright and related rights, and other stakeholders, in setting common standards and compatibilities, and resource sharing, will facilitate this.

Article 7 – Selecting 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. 'Born digital' materials should clearly be

given priority. 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 legal and institutional frameworks to secure the protection of their digital heritage.

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.

Access to legally deposited digital heritage materials, within reasonable restrictions, should be assured without causing prejudice to their normal exploitation.

Legal and technical 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 – Preserving cultural heritage

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, so as to assure over time representation of all peoples, nations, cultures and languages.

RESPONSIBILITIES

Article 10 – Roles and responsibilities

Member States may wish to designate one or more agencies to take coordinating responsibility for the preservation of the digital heritage, and to make available necessary 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, both public and private, 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 propose standard ethical, legal and technical guidelines, to support the preservation of the digital heritage;
- (d) determine, on the basis of the experience gained over the next six years in implementing the present Charter and the Guidelines, whether there is a need for further standard-setting instruments for the promotion and preservation of the digital heritage.

About the authors

Abdelaziz Abid is in charge of the ‘Memory of the World’ Programme of UNESCO and projects relating to libraries and access to information, digital heritage preservation, and information literacy. He also coordinates joint IFLA/ UNESCO library information projects. He has published widely on a range of preservation issues.

David Bearman is president of Archives & Museums Informatics. He consults on issues relating to museum, archives and library collaborations, cultural knowledge management, interactive multimedia, and electronic recordkeeping and archives. In 1997, Bearman co-founded the annual Museums and Web conferences in North America. He is the author of over 125 books and articles on museum and archives information management issues.

Annemieke de Jong is head of information policy at the Institute for Sound and Vision, the national audiovisual archive of the Netherlands. She is involved in several projects and expert groups in the field of media asset management and digital heritage, on a national and international level. Annemieke de Jong holds an MA degree in New media and digital culture.

John Mackenzie Owen is professor of information science at the University of Amsterdam. He has been director of the R&D company of Cap Gemini Innovation in the Netherlands. He has conducted strategic studies for Dutch government library organizations and major international companies. Mackenzie Owen was also involved in programme development and research activities for the European Commission.

William Uricchio is professor and director of Comparative Media Studies at MIT and professor of Comparative Media History at Utrecht University. He has written extensively on ‘old’ and new media, popular cultures and their audiences. His current work takes up these issues through topics ranging from media historiography to peer-to-peer communities and cultural citizenship.

Vincent Wintermans is project coordinator at the Netherlands National Commission for UNESCO. His main professional interests are documentary heritage, cultural diversity, communication and development, and human rights in the digital domain.

List of participants

A. Abid, UNESCO Information and Informatics Division, Tunisia
G. Alberts, Radboud Universiteit Nijmegen, Netherlands
I. Androvic, University Library in Bratislava, Slovak Republic
V. Arden-Wood, TMR Digital, United Kingdom
N. Beagrie, British Library, United Kingdom
D. Bearman, Archives & Museum Informatics, Canada
K. van Bezu, Koninklijke Bibliotheek, Netherlands
I. Bjorlin, Regional Centre for Museum and Archives, Norway
A. Borsboom, Nationaal Archief, Netherlands
T. Bos, Strata Preservation NV, Netherlands
M. van den Bosch, University Library Utrecht, Netherlands
A. Bramwel, African Intellectual Property Association, South Africa
E. Caneda Cabrera, Institute for Information Law (University of Amsterdam), Netherlands
M. Chabin, Archive 17, France
D. Clarke, TMR Digital, United Kingdom
G. Clavel-Merrin, Institution Swiss National Library, Switzerland
M. Douze, Int. Information Centre and Archives for the Women's Movement, Netherlands
R. Egeter van Kuyk, AV section of the National Organization for Records Management & Archives DIVA, Netherlands
A. Egli, Swiss Federal Archives, Switzerland
A. Fontana, National Central Library of Florence, Italy
G. Fossati, Nederlands Filmmuseum, Netherlands
R. Frommé, V2, Netherlands
I. Garcia-Monge, Ministry of Culture, Spain
A. Goossens, INWA, Belgium
S. Gorter, De Nederlandsche Bank NV, Netherlands
I. Guicherit, City of Utrecht, Netherlands
V. Hauge, Regional Centre for Museum and Archives, Norway
R. van der Heide, The States General, Netherlands
T. Hessling, De Nederlandsche Bank NV, Netherlands
H. Hofman, Nationaal Archief, Netherlands
B. Hoomans, Netherlands Institute of Sound and Vision, Netherlands
W. Hudson, National Library of Jamaica, Jamaica
V. Huong, The State Records and Archives Department, Vietnam
J. Hutar, Czech National Library in Prague, Czech Republic
E. Jacobs, Netherlands National Commission for UNESCO, Netherlands
M. Jones, Digital Preservation Coalition, United Kingdom
A. de Jong, Netherlands Institute for Sound and Vision, Netherlands
C. Karp, Swedish Museum of Natural History, Sweden
D. Katuscak, Slovak National Library, Slovak Republic
R. Katuscak, IBM Slovakia Ltd., Slovak Republic
A. Kavcic-Colic, National/University Library of Slovenia, Slovenia
C. Keijsper, Koninklijke Bibliotheek, Netherlands
D. Kitsiou, General Secretariat of Communication/Information, Greece
S. Klarin, National and University Library Zagreb, Croatia
C. Konings, Delft University of Technology Library, Netherlands
E. Kwant, Koninklijke Bibliotheek, Netherlands
H. van der Linden, Ministerie van de Vlaamse Gemeenschap, Belgium

C. Lupovici, Bibliothèque nationale de France / International Internet Preservation Consortium, France
Y. de Lusenet, European Commission on Preservation and Access , Netherlands
J. Mackenzie Owen, University of Amsterdam, United Kingdom
L. Maurizio, Digital Renaissance Foundation, Italy
T. McSean, Elsevier, United Kingdom
M. Mihaylov, National Center for Museums, Galleries and Arts, Bulgaria
S. Milonig, Landesmuseum Joanneum, Austria
N. Mkhwanazi, Swaziland National Library Service, Swaziland
G. Mulindwa, National Library of Uganda, Uganda
M. Newton, Caribbean Community Secretariat, Guyana
G. Nilsen, Regional Centre for Museum and Archives, Norway
M. Novakovska, National/University Library St. Clement of Ohrid - Skopje, Republic of Macedonia
A. O hAonghusa, National Library of Ireland, Ireland
E. Oltmans, Koninklijke Bibliotheek – National Library of the Netherlands, Netherlands
C. Pawirosoerono, Koninklijke Bibliotheek – National Library of the Netherlands, Netherlands
M. Pe, Nationaal Archief, Netherlands
M. Pennock, UKOLN University of Bath, United Kingdom
D. Peters, University of Kwazulu Natal, South Africa
M. Pinter, National Archives of Hungary, Hungary
J. Pizza, Institution Greenpeace, United States of America
K. Pollmeier, Staatliche Akademie der Bildenden Künste Stuttgart, Germany
M. Ras, Digitaal Erfgoed Nederland, Netherlands
E. Sachev, National Center for Museums, Galleries and Arts, Bulgaria
F. Salvador, National Library of Spain, Spain
P. Sarens, Library of the Belgian Federal Parliament, Belgium
U. Schwens, Die Deutsche Bibliothek / NESTOR network of expertise, Germany
D. Seiter-Sverko, Ministry of Culture, Croatia
K. Shawul, African Union Commission, Ethiopia
G. Shillingford, Ministry of Education, Dominica, Commonwealth of Dominica, WI
B. Sierman, Koninklijke Bibliotheek, Netherlands
Skultety, Tempest, Slovak Republic
J. Stapel, Koninklijke Bibliotheek, Netherlands
B. Stie, Regional Centre for Museum and Archives, Norway
M. van Stiphout, Q-Info, Netherlands
Z. Szatucsek, National Archives of Hungary, Hungary
J. Szczygiel, The Head Office of State Archives of the Republic Poland, Poland
K. Takehana, National Diet Library, Japan
G. van Trier, Koninklijke Bibliotheek, Netherlands
W. Uricchio, MIT / Utrecht University, United States of America
G. Verhoeven, Amsterdam University Library, Netherlands
L. van Vucht Tijssen, Netherlands National Commission for UNESCO, Netherlands
K. Westerink, Netherlands Institute of Sound and Vision, Netherlands
V. Wintermans, Netherlands National Commission for UNESCO, Netherlands
E. Zalme, The States General, Netherlands
M. Zwenne, Koninklijke Bibliotheek, Netherlands