

# PRESERVING OUR DIGITAL HERITAGE: A NEW PARADIGM

ABDELAZIZ ABID

*A large part of the vast amounts of information produced in the world is born digital, and comes in a wide variety of formats: text, database, audio, film, image. For cultural institutions traditionally entrusted with collecting and preserving cultural heritage, the question has become extremely pressing as to which of these materials should be kept for future generations, and how to go about selecting and preserving them. This enormous trove of digital information may well be lost unless specific techniques and policies are developed to conserve it. 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. UNESCO has developed a strategy for the promotion of digital preservation. This strategy is centred on: a wide consultation process with governments, policy makers, producers of information, heritage institutions and experts, the software industry as well as standard-setting organisations; adoption of an international charter on the preservation of digital heritage; dissemination of technical guidelines and; capacity building and implementation of pilot projects.*

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 ask for new preservation policies and agreements. Traditional preservation policies as we know them in the analogue domain, based as they are on the environment control of the physical conditions of conservation are no more relevant or at least they are no more 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, two are the focus of this paper: the selection of material to be preserved and the sharing of tasks and responsibilities between institutions for preservation purposes. 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.<sup>1</sup>

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:

---

<sup>1</sup> <http://unesdoc.unesco.org/images/0012/001255/125523e.pdf>

- a) a wide consultation process with governments, policy makers, producers of information, heritage institutions and experts, the software industry as well as standard-setting organisations;
- b) dissemination of technical guidelines;
- c) implementation of pilot projects and human capacity building based on the Guidelines;
- d) 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 review of literature, the Library's own experience, and UNESCO's organised 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 aims at **developing the theoretical and methodological knowledge essential to 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 America countries. The countries directly involved in this initiative are Argentina, Brazil, Cuba, Mexico and Peru. Beginning in November 2005 the InterPARES Project was hosting five scholars from the Caribbean and Latin America for a three-week workshop. While in Vancouver, they were learning about the InterPARES research methodology, products and findings in order to bring back to their own countries the new knowledge generated in the context of the project. These same scholars have returned in February 2006 for further elaboration of the knowledge acquired and adaptation to the Caribbean and Latin America environments. A few months later they have had the opportunity to meet once again, in a Latin American environment this time, to further discuss the results achieved and lessons learnt in each country.

Cooperation with CODATA in this field concerns 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 in August 2005 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... 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 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 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, 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 concerned 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 that is 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, or, databases - becomes inaccessible.

Another complex issue concerns copyright, including copyright of 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 for 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 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 keep at 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 whole areas of memory disappearing as we move further into the new millennium.

As long as information used physical media in order to move around, it left traces. Even if one does 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 those 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 transferred 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,

of 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 springboard 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. After purely office activities, 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 on line; 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, very early entered general use. Today, music has become the largest cultural industry by number of files exchanged on the Internet; millions of MP3 files are exchanged every day.

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. Similarly, 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 set.

## **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 affecting 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 finds naturally 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 “.mus” 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 potentialities of the Internet to broaden access to their collections, by putting on line their bibliographical databases online, and gradually also their collections of digitized works. Moreover, they make it possible to access to 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 co-operation 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 web sites 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 gigabyte.

For all that, is all this production 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, 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 Audio-visual

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 irremediable damage in their collective social memory.



United Nations Educational, Scientific and Cultural Organization  
Organisation des Nations Unies pour l'éducation, la science et la culture

## **Charter on the Preservation of the Digital Heritage<sup>1</sup>**

### **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.

---

<sup>1</sup> Adopted at the 32nd session of the General Conference of UNESCO, 17 October 2003

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.

## **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 cooperation 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.