Memory of the information society

UNESCO Publications for the World Summit on the Information Society

Authors:
Jean-Michel Rodes, Geneviève Piejut, Emmanuelle Plas

Edited by Isabelle Vinson, Abdelaziz Abid

Memory of the information society

Published in 2003 by the United Nations Educational, Scientific and Cultural Organization 7, place de Fontenoy F-75352 Paris 07 SP

Composed and Printed by Jouve, Paris

Printed in France

General introduction

UNESCO has fully supported the World Summit on the Information Society (WSIS) preparatory process from its beginning, and has succeeded in defining and promoting its positions while setting the ground for its contribution to the Declaration of Principles and the Plan of Action that the Summit is expected to adopt. UNESCO's proposed elements for inclusion in the Declaration of Principles and the Plan of Action are based on its mandate, which leads it to promote the concept of knowledge societies, rather than that of global information society since enhancing information flows alone is not sufficient to grasp the opportunities for development that is offered by knowledge. Therefore, a more complex, holistic and comprehensive vision and a clearly developmental perspective are needed.

The proposals are responses to the main challenges posed by the construction of knowledge societies: first, to narrow the digital divide that accentuates disparities in development, excluding entire groups and countries from the benefits of information and knowledge; second to guarantee the free flow of, and equitable access to, data, information, best practices and knowledge in the information society; and third to build international consensus on newly required norms and principles.

Knowledge societies should be firmly based on a commitment to human rights and fundamental freedoms, including freedom of expression. They should also ensure the full realization of the right to education and of all cultural rights. In knowledge societies, access to the public domain of information and knowledge for educational and cultural purposes should be as broad as possible providing high quality, diversified and reliable information. Particular emphasis should be given to diversity of cultures and languages.

In knowledge societies, the production and dissemination of educational, scientific and cultural materials, the preservation of the digital heri-

tage, the quality of teaching and learning should be regarded as crucial elements. Networks of specialists and of virtual interest groups should be developed, as they are key to efficient and effective exchanges and cooperation in knowledge societies. ICTs should be seen both as educational discipline and as pedagogical tools in developing effective educational services.

Lastly, these technologies are not merely tools, they inform and shape our modes of communication, and also the processes of our thinking and our creativity. How should we act so that this revolution of minds and instruments is not merely the privilege of a small number of economically highly developed countries? How can we ensure access for all to these information and intellectual resources, and overcome the social, cultural and linguistic obstacles? How should we promote the publication on line of increasingly more diversified contents, potentially a source of enrichment for the whole of humanity? What teaching opportunities are offered by these new means of communication?

These are crucial questions to which answers must be found if knowledge societies are to become a reality, and offer a world-wide space for interaction and exchange. They are also questions which the actors of the development of these technologies – States, private enterprise and civil society – must answer together.

On the occasion of the World Summit on the Information Society, UNESCO intends to make available to all participants a series of documents summarizing some of the most worrying questions which have just been mentioned. These will help participants to take the measure of the upheavals brought about by the emergence of the new information and communication technologies (NICTs), and will deal with the potential for development, the difficulties encountered, possible solutions, and the various projects implemented by UNESCO and its many partners.

Abdul Waheed KHAN
UNESCO's Assistant Director-General
for Communication and Information

Table of contents

Introduction

The last fifty years have seen extremely rapid development in the computerization of society. The phenomenon is so massive, and changes so much in the contours of our civilizations, that we now speak about the Information Society.

In four successive waves, the same impetus created all conditions and premises of digital heritage, and then, very quickly, caused its inflation:

- up to the 1970s: archiving of the first computing data from central sites and scientific calculators,
- in the 1980s: very rapid development of digital publishing media: first the audio CD, followed by multimedia CD-ROMs and video games on consoles,
- in the 1990s: the advent of digital television and radio by satellite,
- at the turn of the millennium: the generalized interconnection of networks and the meteoric development of the Internet, especially its Web and mail applications.

Before the Internet developed, it was still possible 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.

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.

To those early humans, computer files would have been the worst possible materials on which to preserve memories: computer storage media are of poor quality, designed for mass distribution, not for being preserved. One machine drives out another to feed consumption cycles, file formats change, each new software version being held hostage to exacerbated competition. Within a few decades we shall have seen more standards for alphabetical characters than since the invention of carved stone.

Unless decision-makers quickly show enough political will, and take measures commensurate with the issues at stake, there is a great risk that our entire information society will explode without leaving any more trace than the Internet bubble. Our information societies would be reduced to societies obsessed by the present, with only tiny amounts of working memory, self-centred in their frenzy to communicate, and turning their backs on the generations to come, breaking the chain of transmission.

And yet, digitization remains a wonderful vehicle for the democratization of culture and the dissemination of cultural works. There are still enormous gaps to fill in terms of equipment, networks and training before everyone on Earth has access to all the world’s knowledge, but the way is already mapped out.

All matter tends to disappear gradually, to dissolve, to disintegrate, to yellow, to age – but not 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 destroyed. 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 regarded as the worst enemy of conservation.

On the contrary, the ability to allow digital information to circulate rapidly on a series of new carriers will be the ultimate guarantee of its permanent existence.

Along the way, places of public reading such as archiving centres and museums will perhaps have become content servers, revolutionizing jobs and qualifications, types of organization, ways of reading, modes of financing, and the conflict between the right to information on the one hand and respect of authors' and publishers' property and individuals' rights on the other.

Moreover, the network offers us new vistas of worldwide knowledge, gathered together and accessible from anywhere. To reproduce this unity, for those who will one day have the task of plotting its historical depth, when individual and collective initiatives across the world have enabled us to preserve fragments of the Web, a powerful effort will be needed to reconstitute the mosaic. An effort which is careful not to crush the legitimate aspirations of each people to control their own memory, but rather an effort which cuts across the whole of the necessary diversity of the world's heritage institutions.

Emmanuel Hoog
Chief Executive
Institut National de l'Audiovisuel

Chapitre 1 The situation is urgent

The wave of computerization which has broken over our societies over the last half-century has today reached a kind of maturity. It will undoubtedly be followed by other waves, but as we look back towards the past which existed before the wave, we cannot escape being struck by the extent of the transformations which will have impacted on all areas of information, creativity and knowledge – in fact on the entirety of our culture.

We suddenly discover, as we are carried by this wave, that various components have been gradually woven together ultimately coming together to form a digital heritage, the extent of which has only become apparent with the emergence of the Internet, which brought this first age of computing to a close.

The topicality of this new form of heritage is all the more urgent inasmuch as computing, up to now, has been noted more for its expansion capacities than for its ability to flourish in the long term. Hitherto, the means of producing and moving information have multiplied in proportions completely unheard of in the history of humanity, but if we are not careful, the information cycle, under pressure from technology cycles, will never have been shorter, and our societies risk witnessing whole areas of memory disappearing as we move further into the new millennium.

The more we communicate, the less we shall transmit to posterity.

1.1. A new model for the memory of society

Computing already contained in embryo certain specific features which enable it:

- to substitute for mankind in a wide variety of fields,
- to constitute a new prosthesis on a hitherto unimaginable level,
- to be the language into which all others can be resolved,
- to act as a generic tool for the production of knowledge and information,

- to emerge in many fields as simultaneously the instrument, the matter and the workbench of creativity,
- to become progressively the dominant medium among peoples,
- and ultimately, to impose itself as the place where a new form of archive arises.

For this stage to be completed, four conditions had to be fulfilled:

- there had to be an automatic machine which made it possible to calculate and execute programs at much higher speeds than in "real time" in human perception;
- the machine had to be capable of communicating and interacting with humans in a wide variety of ways;
- systems had to be technically able to store and index enormous quantities of information;
- there had to be a generalized system of exchange among machines, in such a way as to build an autonomous universe of publication and communication.

All these functions were already present in an embryonic state in the earliest machines – calculation, programming, output peripherals, memory, interconnection. Indeed, the pace of technological innovation conditioned the speed of development of these functions, enabling new uses to be deployed along the way, uses which it authorizes and causes.

And then the last decade of the twentieth century, with its rapid development of a worldwide network, brought the first chapter of computing to a close, raising with unexpected urgency the question of the new kind of heritage which this history reveals.

It is strange, and in retrospect impressive, to note at which point computing pioneers became conscious of the upheaval which computing would bring to collective memory. Vannevar Bush, as early as July 1945 in "As We May Think" described Memex (MEMory Extender) as a kind of vast hypertext memory. J.C.R. Licklider, who in 1962 launched his first work on the ARPANET network, the ancestor of Internet, describes with extraordinary

clarity what, for many, was not much more than a very large calculating machine: a new universe for people to live in[1].

Far beyond this space for the creation of information, computing has also made itself able to absorb a large proportion of pre-digital archives by increasingly fine sampling and the digitization of analogue signals coming from the real world. Hence the entire documentary memory of humanity can be assimilated into digital space (subject, of course, to nuances which we will examine below).

A constant of the history of computing will emerge as we proceed: first, it develops a capacity to synthesize (texts, images, music, sounds, 3D objects, etc.), drawing on the latest technological developments in new interfaces, new forms of representation and new output peripherals, then it develops tools for the capture, analysis and understanding of reality.

With the advent of networks, computing has become a world in itself, one which no longer needs external expression. A virtual world, according to the most frequently expressed fear. Consequently, any human activity which can be expressed in terms of bits and bytes must conceive of the future of its heritage in digital form.

We have come full circle, and the field for this new mode of expression is indeed broad. For as each wave of technological innovation has broken over the world, it has scooped up in passing a significant number of modes of human expression, and set them down on the side of digitization. All that was lacking for it to achieve definitive autonomy was this final territory. We have by no means seen the end of all the changes, but the way forward is now clear, and as technology comes full circle, it reveals something that had remained hidden in this tidal wave of digitization in hardly a half century of our history: the revolution of heritage.

1. “The ARPA theme is that the promise offered by the computer as a communication medium between people, dwarfs into relative insignificance the historical beginnings of the computer as an arithmetic engine.” (ARPA draft, III-24. 1963). This ARPA (Advanced Research Project Agency) document is considered to be the foundation stone of ARPANET, the ancestor of the Internet).

1.2. Society in the grip of digital technology

The digitization of all data produced by human intelligence, whatever their original form – the written word, sounds, fixed or moving images – simultaneously affects the process of creating content, the way in which content is disseminated, and the ways in which 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.

It is related to the widespread improvement in computing performance levels, amplified in recent years by the development of the capacity of communication systems. The effects of these on our modes of production of and access to culture and knowledge cannot yet be fully measured.

From the development of the first computers to the invention of word processing, the design of the first synthesized image software and desktop publication tools, and the manufacture of the first high-capacity digital CDs etc., there has been an uninterrupted succession of technological innovations which in recent decades has penetrated whole areas of cultural, scientific and artistic production, without our realizing or even noticing.

Retrospectively, we may observe that for each function of the computer (calculation, memory, interface, network) there has been an association, a very strong and irreversible correlation between the onset of technological maturity and the computerization of a new continent of the social sphere.

The age of the calculator, or the emergence of scientific and technical computing

At the end of the Second World War, the first electronic calculators, the distant heirs of the abacus and Pascal machines, entered upon their rapid evolution into the modern computer. Thus began an era of profound transformation which has now reached its maturity.

The end of this period of change opens wide the field of a new form of memory for humanity.

The first uses of computers arose from their ability to perform calculations. They were initially built for military purposes – encrypting and decoding, ballistic calculations – then for scientific needs such as astronomical tables, etc.

This first function was to lead to the generalized deployment of computing for science and technology, statistical tasks, and management activities directly requiring calculations.

Since then, the world has witnessed the astonishing progression in the capacity and computing speed of the microprocessor, which according to Moore's Law doubles every 18 months. Each time this speed increases, it will enable computing to enter new fields of human activity.

The first area in which digital technology was massively deployed was science, carried by applications in military and space technology.

Initially in the so-called "hard" sciences – physics and mathematics for example – which constituted both an area for the application of the technological advances developed by computing specialists, and a powerful engine for computing innovations. This cross-fertilization was all the easier in that there existed a real intellectual and theoretical proximity among researchers, who very often worked side by side on the same campuses, or even in the same laboratories.

At the heart of the battle, therefore, control over the scientific information produced by laboratories and collected by multiple recording devices, is exercised by increasingly more powerful systems. These also affect information as it is exchanged, distributed and stored, thus constantly repeating and enriching the chain of the production of knowledge.

But first it was necessary to meet researchers' needs for increasingly extensive computing power to treat masses of increasingly bulky data, in processing times approaching real time. Today, with computing power exceeding 1 gigaflop (1 billion floating point calculations per second), computing has come a long way from its ancestor of 1946, the 30-ton ENIAC supercomputer which was able to carry out 330 multiplications per second.

At the same time, new systems for the production of digital information were developed, for example in the fields of medical imaging and space

research, and new tools for the visualization and representation of data, such as software for simulations or the production of synthesized images in 2D and 3D. Thus there came into being a new vision, a new way of looking at the world.

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, which in many countries is still widely dispersed over a large number of laboratories, cannot be easily interpreted by others, and is thus not readily transmissible.

To respond to this problem, the world's scientific communities created the International Council for Science (ICSU), which manages some 49 world data centres covering all disciplines. These world data centres were set up at the beginning of the 1950s to foster exchanges and create infrastructures to improve access to data and archives.

However, 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 age of dialogue and interface, or the mutation of writing and text

Input and output peripherals have of necessity always existed around central processing units: cards, tapes, various kinds of printers, etc. The screen, in its primitive form, dates back to 1951. In the 1970s, screens and time – sharing became essential, opening up the possibility of truly instantaneous interaction between man and machine.

Gradually, the computer emerged from specialized computing sites and became capable of managing text in day-to-day activities. The end of the 1970s and the beginning of the 1980s saw the dual parentage of office automation, the illegitimate child of the typewriter as it became electronic, and of centralized computing as it made its way into the secretary's office with the transition to micro-computing. The high-speed quality laser printer brought about the ultimate extinction of the first generation.

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. Graphic interfaces accelerated this upheaval in commercial printing and its transition to desktop publishing (DTP).

Gradually, the tools were established of unprecedented change in writing and printed texts. Alongside the revolution in interfaces, digital distribution media began to replace analogue carriers (e.g. the audio CD quickly supplanted the vinyl disc). The CD-ROM goes even further, offering for the first time radically new modes of browsing and reading. Thus were the foundations laid for a heritage exclusively digital in nature.

In 1975, BRAVO, the first WYSIWYG (What You See Is What You Get) word processor, was developed. A few years later, domestic microcomputers began to conquer the general public. Today, there are e-books and people have easy access to the largest libraries in the world from their own homes. In barely a quarter of a century, a digital world has imposed itself, offering a new way of reading, a new vehicle for written culture. After handwritten and later printed texts, it provided text with a new mode of inscription, without replacing the old. And with it, new methods of producing and transmitting knowledge, and new practices for reading.

This revolution, wrought by computing over traditional modes of writing, constitutes a revolution in the history of thought of similar scale to that introduced by the codex, when it replaced the scroll or *volumen* between 1st and the 4th century of our common era.

Later, thanks to the invention of printing, the codex gave way to the book, and in the space of a few decades, printing – the earliest form of mass reproduction and distribution – integrated and consolidated all former lear-

ning in terms of pagination, list of contents and division into chapters, etc., thus contributing to a profound transformation in the uses to which texts and books were put, which lasted several centuries. With the generalization of the book as object, a strong and visible distinction came into being between author and reader, writing and reading, text and book.

Now, it is precisely this allocation of functions which has become blurred by the advent of digital computing. Text on computer media can be subjected to all kinds of processing and revision: readers can annotate it, cut it, copy from it, recompose it, adopt fragments, rewrite it, even becoming in their turn co-authors. And so, since anyone can now become an actor in a kind of pluralist writing exercise, it is the very concept of creation, hitherto defined as a singular and original individual act, which is being called into question[1].

Digital computing has also had a profound impact on the publishing profession. First, publishers' methods and tools have in recent years suddenly gone digital, as techniques such as DTP and on-line distribution have been generalized. This has had the immediate consequence of increasing reliance on the optimization of publication capacities. Secondly, it affects the act of publication itself. Publication, which fixes a text at a given moment, is initially an act of selection performed in an environment of plentiful production. Expressing as it does editorial choice, it represents a promise of quality and result. Nothing similar exists in the digital universe, where texts are nothing more than fragments of a worldwide, anonymous, moving flow.

In the world of print, texts are embodied in material forms which bring about particular expectations, different relationships with the reader according to whether one is reading for example a newspaper article, a letter, a learned periodical or a page of an encyclopaedia. The materiality of the printed word brings about a classification of expression into a hierarchical system, which is erased in the digital universe, where text appears on the surface of a screen without depth, uniform and standardized. As they lose

1. R. Chartier, *Culture écrite et société*, Paris, Albin Michel, 1997.

their physical identity, the various kinds of text become similar in appearance and equivalent in their authority[1].

It was in an attempt to go beyond these limits, which in many respects take us back to the *volumen*, that many recent development efforts have concentrated on the electronic book, with mixed success. This was an attempt to offer texts a medium which was, stable, dated, signed and certified. Supported by on-line services or specialized sites, these developments are particularly promising in the areas of education and training, where new forms of learning and discovery are already taking shape thanks to these new modes of reading[2].

This mode of reading, called "hyper-reading" by some, is no longer based on linear or deductive logic, which supported traditional lines of argument and demonstration. The multiplication of hypertext links causes precisely the opposite, exploding circulation, opening up reasoning, making logic more relational, no longer based on a single thought but on a set of keywords. This profoundly modifies traditional modes of knowledge acquisition and accreditation, all the more so since, for the first time in the history of humanity, the world's entire stock of knowledge, or almost, is virtually accessible.

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. Still, the need remains for our societies to avoid being stricken with amnesia.

The age of memory: images and sounds

One cannot conceive of a computer without memory, i.e. random-access memory and mass storage. Although the computer's random-access memory is only a volatile working memory which is erased when the computer is switched off, its mass storage devices enable it to store data or programs in a permanent form.

1. C. Vanderdorpe, *Du papyrus à l'hypertexte*, Paris, La Découverte, 1999.
2. A. Cordier, report by an appraisal committee on the digital book, initiated in 1999 by the French Ministry of Culture within a programme of government action to prepare France to enter the information society.

Mass storage devices have always existed, and were improvised ones in the earliest days – 35 mm cinema films for Z1 (the first electromechanical relay calculator, invented in 1938 by Konrad Zuse); IBM punched cards and tapes, remote descendants of the Jacquard loom, from 1928 onwards; magnetic tapes from 1949 – but only with the hard disk (1956) did the computer truly achieve what we would recognize as immediacy. In 1975 the Winchester hard disk became a standard for computing sites, but it was not until the 1980s that consumer microcomputers were generally fitted with them.

Since then, storage capacities have grown considerably, in the form of magnetic cartridges (S-DLT, LTO, etc.) which soon broke through the 1 terabyte barrier, optical discs which enable several gigabytes to be stored, perhaps more durably, and which benefit from the extraordinary spread of CDs and DVDs, and hard disks, the per-megabyte price of which has fallen spectacularly, making possible today what even in the recent past was inconceivable.[1]

The extraordinary development of storage capacities, with different access times according to the technologies used, will make it possible for computing to transform itself into digitization, and to enter fields of activity from which its hitherto poor performance barred the way, such as the world of images and sound.

Discrete signals

Ever since it began, the recording of images and sounds has been an analogue activity, and this was also true of each of their components. The term "analogue" really only emerged with the birth of digital technologies, to differentiate an earlier state from what clearly appeared as the result of radical change:

Photographs, films, photochemical processes, all of these are analogue insofar as their recording varies in continuous orders of magnitude according to the quantity of light received. Magnetic tapes (for video and audio)

1. DLT: Digital Linear Tape; S-DLT: Super Digital Linear Tape; LTO: Linear Tape Open; DVD : Digital Versatile Disk.

and vinyl discs (for audio) use a process of magnetization, engraving or pressing to fix the impression of an electrical modulation – itself the translation and continuous amplification of physical orders of magnitude (frequencies of light and sound).

In an analogue recording, the human operator is unable to intervene or to control the process as it reaches each discrete unit, either in terms of the impression (silver halide microcrystals, the metal particles of magnetic tapes, the graininess of vinyl and the width of the groove) nor in terms of the signal (fine-tuning of frequencies).

However, the essential characteristic of digitization is its ability to manipulate the signal, which is made possible by constituting discrete units, i.e. the sampling of continuous physical phenomena according to variable sampling frequencies (the number of samples taken per second) and levels of definition (the size in bits of each sample); thus, the higher the frequency and definition, the higher the quality of the recording, and the closer it will approach the quality of the original. This is in itself an indefinable absolute, dependent on the means of capture (optical components, microphone etc.) and the objective pursued, bearing in mind that all recording media, including analogue ones, and all technical devices have their own limits – their graininess.

Sound and music

Sound recording, a pioneer in the digital field, very early became open to the development of musical computing. Initially for the production of synthesized sounds which contributed to define the contours of a new musical space explored by Pierre Schaeffer, and later in the elaboration and development of tools for computer-assisted composition, with the result that in the 1980s, audio sampling software and other electronic ″instrument factories″ entered general use.

Centred initially on synthesis and the analytical handling of sound for creative purposes, this work generated new applications concerning systems for collecting, recording and broadcasting sound. Today, the digitization of audio signal has become the rule; a new technical system is spreading throughout the world, causing profound upheaval to the world of audio and music in all its forms.

On the creative side, the spread of composition software, its availability to a wide audience, and the generalization of "home studio" systems, have brought about considerable change in the way music is produced, reinstating the continuity between creator and consumer which had been broken by the electromechanical techniques of sound reproduction.

Concerning diffusion, the development of the Internet, combined with the widespread adoption of the compression standards MPEG and MP3, open up new possibilities of on-line access to musical collections, to such an extent that today, music has become the largest cultural industry by number of files exchanged on the Internet (no less than 3 million MP3 files were exchanged every day in 2002)[1]. Destabilizing traditional modes of diffusion such as broadcasting and the record industry, the Internet not only allows its users to access in a couple of clicks the largest sound library ever put together, but also to reach a whole environment of information (for example, names of authors and artistes, titles, themes, etc.), to experience various modes of visual or graphic representation, for example presentation of the score synchronized with the playing of the music, in short, to venture into new landscapes of music.

Synthesized images

As in the case of music, digitization always begins with synthesis (synthesized sound, MIDI, 2D and 3D images) in non-real time, and later spreads to the recording, processing and reproduction of images and natural sounds.

The ability to perform fluid calculations, to control vectorial logic and to provide ergonomic interfaces and vast storage capacities then made it possible to create graphic images on computers. This led to a very broad spectrum of applications, in particular in the areas of computer-aided design (CAD) and artistic creation.

For reasons of processing speed and storage capacity, digitization was initially limited to the processing of fixed images, while the development of scanners and image processing software made this increasingly accessible to

1. MPEG: Moving Pictures Experts Group, the ISO working group in charge of international standards for the coded representation of digital audio and video. MP3 is an MPEG format for audio.

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.

It was by these synthesizing techniques, which had evolved largely from techniques of military simulation and aeronautics, that digitization penetrated the audiovisual world. Rapid extension of these applications in the 1980s to the needs of television and cinema production led to the generalization of special effects – for example, "Tron", the first film made entirely with synthesized images was released in 1982 by Walt Disney Studios – and ultimately techniques of digitally encoding and compressing film and video images made it possible to marry together at the post-production stage synthesized and natural images.

From then on, and with the development of standards for digital compression and image processing – we are reminded here of the standardization work of the JPEG group (Joint Photographic Experts Group) for fixed images and the MPEG group for moving images, from the 1980s onwards – the digitization of the audiovisual system was set to affect all its components, albeit gradually, from production to editing, from editing to transmission control, from transmission control to broadcasting networks, and finally to the private individual's TV set.

Digital audiovisual documents

In a way which clearly seemed to be a radical departure from an earlier state of affairs, the digitization of the audiovisual system began in the production area, or more accurately in the post-production phase. From the very beginning of the 1990s, off-line editing gradually replaced traditional video editing. The same change occurred in the cinematographic industry, where post-production of images (cutting the working copy, adding special effects and colour grading) and sound became an entirely digital process over ten years ago.

As for camera systems, the earliest digital cameras were initially used by television news departments who very soon recognized their advantages (miniaturization, flexibility in use, simplified editing). They also seduced creative programme-makers and directors, and gradually penetrated the

world of cinema production, which during its century of existence had changed neither its medium nor its method of recording. Here, their use contributes to a rejuvenation in approach and in the industry's own way of looking at the world, as seen in the work of some of the Dogma group cineasts. With digital, audiovisual expression and cinematographic creation benefit from new, undreamt-of tools for production, action and photography, without forgetting its remarkable new possibilities for special effects.

Digital broadcasting

A further element of the audiovisual system which was soon to switch to digital was broadcasting, and in particular the routing of programmes to viewers' own TV sets. In 1994, DIRECTV, the first digital television channel aimed at the general public was launched in the United States. In Europe, an organization called DVB Project (Digital Video Broadcasting Project) brought together broadcasters, manufacturers and public administrations, and as early as 1993 established technical specifications and European standards which made possible the launch of terrestrial digital television, and with it, the development of new functionalities and services. As an indication, compression of the digital signal makes it possible to broadcast simultaneously between 4 and 16 different television programmes on a single channel, according to the compression ratio used. Add to this the reduction in operating costs and the possibility of offering associated interactive services, as well as better quality picture and sound, and analogue terrestrial TV appears to be largely condemned to extinction. Already certain countries such as the USA, Great Britain, Sweden and Spain have adopted digital terrestrial broadcasting.

After ten years of continuous technological developments, one can consider that the entire audiovisual chain, from production to broadcasting, is henceforth completely digital and no longer needs at any time to revert to an analogue stage. It is true that physical supports still exist for analogue, as digital audio and video, in particular for reasons of their very high data volumes, use specific digital media (DAT[1], Digital Betacam, etc.) but the time is close when dematerialization will be complete.

1. Digital Audio Tape.

The age of networks: e-commerce, services and public administration

The bases of Internet were established in 1969, and in 1989 Tim Berners-Lee (at CERN) invented the World Wide Web. Computers already communicated with each other, but this communication was a specialist affair, and public networks were still very much marked by the world of telecommunications. The worldwide de facto standardization of communications protocols opened the way to the generalized interconnection of computers all over the world.

It is with this generalized interconnection that the adventure of digital heritage really starts. Before, 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 generalized network rapidly absorbed and extended a number of service activities (private communications, distribution, mail, telephony, trade, financial transactions, in both BtoB and BtoC modes...) and public communications (publication, dissemination, peer to peer, VOD, etc.)[1]*. It thus superimposed itself on the old schemes of things which had taken several centuries to arrive at their present form.*

The Internet, which over the last five years has developed at a meteoric rate, in particular in its Web and electronic mail applications, grew from 1998 to 2002 at a rate of 217% (from 2,851 million sites in 1998 to 9,040 million sites in 2002[2]) and thus has now entered adulthood. After causing the most insane speculations about its economic promises, after giving rise to a seemingly limitless multiplication of applications and experiments, its uses are now tending to stabilize around a platform of large, firmly established functionalities:

- as a tool for communication and dialogue, allowing direct and simultaneous exchange between various participants, and offering the possibility of near-instantaneous interaction

1. BtoB: Business to Business; BtoC: Business to Consumers; VOD: Video On Demand.
2. Source: OCLC (Online Computer Library Center), “Web Characterization”, http://www.oclc.org.

- as a tool for finding information, the most gigantic data reservoir constituted by the dynamic interconnection of multiple databases trawled by powerful search engines
- 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, allowing direct interaction between the product or service and its potential purchaser or user
- 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 formalization of human thought and creativity.

New ways of consuming culture

Drawing its strength from these attributes, the Internet thus naturally finds its place in the field of culture and education, permanently modifying circuits of access to information and knowledge while supporting the emergence of new cultural practices[1].

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, and gradually, their collections of digitized works. Moreover, they make it

1. P. Lévy, *Cyberculture*, Paris, Odile Jacob, 1998.

possible to access to rare and invaluable works, often kept in storerooms so that they are not exposed to the risk of physical damage. Thus areas hitherto reserved for the few are made available for the greatest number to see.

One begins to see here all the benefit that this broader access to cultural sources can bring 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 virtual classes and e-learning programmes have been developed in recent years. As a communication tool, the Internet also contributes via dedicated websites and forums, to cementing communities of interest around a set of themes or areas of knowledge – these virtual communities exchange information, analyses and points of view on the subject which brings them together.

Electronic communications

Today, on-line public services and administration are well on the way to forming part of our landscape. They are however only the visible face of the process of computerization of organizations. In-house, the deployment of e-mail and intranets is 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.

Electronic communications between people cause a kind of smoothing of hierarchical relations, while decision-making mechanisms and the identification of responsibilities, formerly clearly posted in an organization chart or procedures manual, are gradually diluted or at least become less visible. This dilution is not without affecting archiving practices. These, methodologically speaking, are based for example on the origin and hierarchical position of the information producer: the higher the level in the hierarchy of the decision-making unit from which a document emanates, the greater its testimonial value, and consequently the more it deserves to be preserved. Will the weakening of this selection criterion lead archivists to want to preserve everything, for fear of losing the essential? And could they even do so?

The diffusion of contents

The interconnected nature of the networks and the progressive generalization of broadband are disrupting the world of publishing. Publishing products are now either made available on line, or distributed by gigantic worldwide bookshops such as Amazon, considerably reducing the duration of the order process, or, still within the digital universe, distributed by servers to personal computers in the form of files, or finally, are directly distributed by an on-line publisher from the print world and posted in an HTML (HyperText Markup Language) browser.

For audiovisual works, the video-on-demand (VOD) model is now gaining ground, where viewers choose from vast catalogues an audiovisual work which is then posted on their computer screen. Much more than just an electronic distribution system, the Internet makes it possible to accompany the work with an environment of information and complementary documentation. Finally, with the recent multiplication of Web-TV sites, the Internet is not only a vehicle for additional broadcasting: it is a new medium which merges sounds, words and audiovisual productions.

Public services

Around the middle of the 1990s, many governments adopted voluntarist policies to encourage public administrations and services to computerize their activities, sometimes shaking up long-established working practices. One of the effects of modernity, perhaps, but more than that, one detects in these policies a concern for improving relations with citizens, at a time when citizens sharply criticized certain areas of the apparatus of state; one can also see here a real need to increase the effectiveness of the services provided, a way of reducing costs and operating cycles in a measured economy[1].

Powerful, useful and effective though it may be, Internet has its share of snags, not the least of which is its lack of memory. In 50 or 100 years' time, when historians consider this era of triumphant information and communica-

1. Report of the 3rd Global Forum on Governance "Fostering Democracy and Development through E-government", organized in March 2001 by OECD. See http://www1.oecd.org/puma.

tion technologies, they are likely to find themselves faced with a large digital ditch which will have absorbed millions of pieces of information of considerable scientific, cultural, historical or sociological value, and even of commercial and industrial importance. It is thus essential to counter the amnesia of the network and to adopt some means of organizing the memory of the Web.

The digital divide

Digital technologies present considerable advantages as far as the spreading of information and the democratization of culture are concerned. For several decades however, it will continue to create divisions within each society, between those sectors of the population which have financial means to access it and the cultural means to control it, which will be located in places with a heavy density of broadband networks, and those sectors of the population which will be left by the wayside.

Thus, digital has spread very quickly in industrial societies and some niches of the developing countries, but whole geographical areas are excluded from it[1].

As regards digital heritage, studies of the storage market show that purchase of mass storage media (hard disks, magnetic tapes, optical carriers)are concentrated in the industrialized countries. Not all these media are heritage storage media, but one can nevertheless conclude that there are strong correlations between the quantities of information created and stored in these countries and the growth of a digital heritage, either born-digital or digitized.

In addition, no-one can be unaware of the pre-eminence of English on the Internet (72% of sites, according to OCLC 2002). Even if the trend is falling, it is significant of the advance of the English-speaking countries in terms of diffusion of digital culture among all sectors of the population, but also of the tendency towards standardization of the tools of world-wide communication and thus to the risk that the minority heritages will lapse into oblivion.

1. 605 million users in May 2002, barely 10% of the world's population (0.39% in 1995): Europe 31%, Asia Pacific 31%, North America 30%, Latin America 5.5%, Africa 1%, Middle East 0.8% (Source: nua.com/surveys).

Chapitre 2 Fragile heritage

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 computing has one congenital defect: if you do not save something, you erase it.

In other words, the preservation of heritage must henceforth be a deliberate, voluntary act, organized in the present.

Furthermore, computing 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. Multiplying the layers of logical abstraction between matter and semantic content accessible to human beings has resulted in a code which is more transitory than its written expression.

Preserving heritage in digital form could prove to be at once the best of things and the worst of things, according to what one makes of it:

- *at last, a solution which will usher in a new age for all aspects of preservation, and will make it possible to think through the problems of permanency, indeed of eternity, in completely new terms,*
- *or the anarchy of unrestricted information in an eternal present, giving rise to societies devoid of memory, erasing itself at the whim of the obsolescence of machines or formats.*

2.1. The domain of digital heritage

In its traditional sense, heritage can be defined as all data (monuments, museum collections, archives, libraries...) or practices that a society inherits from its past, and which it intends to preserve and transmit to future generations, with the aim of constituting a common foundation of values and references on which there can develop a feeling of membership and sharing of common social values.

These social values are embodied in collective assets, material or immaterial, which can cover an almost unlimited area as soon as a given social

group invests them with a community or identity-giving function. Great works of humanity which deserve to be known and listed among the artisan practices of a particular ethnic group, the field of heritage remains unstable, its borders fuzzy and its criteria inevitably floating[1].

In the broadest sense, the great principles by which these assets are selected, which mark the limits of the domain and guide the actions of cultural and heritage organizations, rest on this fundamental characteristic: "Many of these resources are of lasting value and importance, and consequently constitute a heritage which must be protected and preserved for present and future generations. Heritage can exist in any language, any part of the world and any field of human knowledge or mode of expression."[2] They potentially relate to all manifestations and forms of human intellectual and artistic expression and production, including, as they emerge, any new vectors of this production. Photography, the cinema, audiovisual and multi-media productions henceforth undeniably already belong to the domain of heritage. Today, digital data are about to join it.

Indeed, digitization is on the move, and at work in all spheres of thought and creation. As a result of this process, a significant portion of computer files stored on hundreds of servers, already constitutes new cultural, educational or scientific resources which supplement, or even replace the traditional components of heritage, i.e. books and writings preserved in libraries, works of art and collections of objects preserved in museums, and the memory of societies and public organizations preserved by record-keeping institutions.

Now, this digital heritage is fragile, all the more so since in many cases it exists only in digital form. Faced with this fragility, and in order to be able to transmit this contemporary memory to future generations, archiving devices and preservation strategies need to be put in place. They must provide a response to the needs for perennial conservation of a domain which is still not well perceived as an object of memory, because it poses problems still

1. The domain covered by the UNESCO publication *World Heritage Review* provides a number of examples of the notion of heritage, as does Resolution A/RES/56/8 – 2002 of the General Assembly of the United Nations on cultural heritage.

2. Draft *Charter on the preservation of the digital heritage*. This text will be submitted to the UNESCO General Conference in September 2003 for adoption.

largely unresolved, and because it seems to run into a technological instability which is disarming. Preservation policies have to be deployed in the presence of a double paradox: how can the immaterial be stored, and how can flows be archived.

However, this digital heritage is far from being homogeneous, and its modes of preservation will not be completely comparable, depending on their origin. Thus, it may be a question of preserving contents resulting from the copying, or the digital representation of original, pre-existing material works – the essential questions will concern media and the tools to be used for reading – or of contents which have no other form than their digital configuration. In such cases, the experiments undertaken on the ground show us that it is essential to take into account the requirements of preservation at all stages of production.

The work and its double

A significant part of the 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, etc., recorded on a defined, stable material medium. This digital "double" does not claim to be an identical copy of the initial work, but contents itself with being a representation of it: it is a snapshot, a print, a trace at a given moment in time, and in every case, the result of a voluntarist policy of digitization.

These digitization campaigns, i.e. the encoding in binary data of contents which, by virtue of this process, are freed from their initial carrier, are carried out on the initiative of heritage institutions, museums, libraries, archiving centres or cultural, educational or scientific institutions. Often calling on specific resources on a large scale, these campaigns aim to improve either the safe keeping of the assets held, or conditions of legibility and access.

Indeed, the decision to promote operations of digitization of assets in this way can serve several purposes. First of all, it can be to answer the need to counteract the physical deterioration of the carriers, leading to the loss of

the contents recorded on them. Digitization and transfers on to digital media thus constitute a solution which makes possible:

- protection of the original work, when access to it implies handling, which can lead to deterioration of the unique carrier on which it exists. For example, the Göttingen Library in Germany has undertaken to digitize the Gutenberg Bible, of which it is the custodian, to protect it from any deterioration, while making it available to the greatest number.
- the long term conservation of works existing on media exposed to risks of irremediable physical or chemical damage, for example acidity for paper, vinegar syndrome for celluloid film stock, or videotape: the issue here is to safeguard contents before they are erased for ever from a material carrier in a process of decomposition.
- the maintenance of conditions of readability, when access to the work calls for a technical reading device which is under threat of obsolescence. This is the case for example with collections of analogue audio and video recordings, a world heritage evaluated by UNESCO as representing nearly 200 million hours, which is likely to become definitively inaccessible for want of appropriate playback devices[1].

At the same time, these digitization policies can serve the objectives of the promotion of heritage, extracting greater value from collections and optimizing their diffusion. Thus, what needs to be done is the following:

- produce virtual illustrated catalogues, or establish a digital memory of works, organized in databases or published on digital disks, so that the electronic resources which reproduce the works will make it possible to maintain and make immediately accessible the witness of an artistic or intellectual creation. Many museums have adopted practices which are exemplary initiatives of this kind,
- optimize access, by putting these data on line on dedicated sites: virtual museums, electronic libraries, databases of still or moving images, musical sites etc. These today constitute virtual spaces where these works can be visited and consulted, and which, having been plunged into a digital environment, benefit from new modes of access, reading and analysis.

1. Sources: Sub-committee on technology of the UNESCO Memory of the World programme; Technical committee of IASA (International Association of Sound and Audiovisual Archives).

Born digital

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.

This ″born digital″ heritage results from an ″all-digital″ process of initial production, the message being digitally encoded at the moment of its creation – as is the case, for example, of a collection of digital photographs of the planet Earth; with the spread of the use of computing and the rise in processing capacities, this production is constantly increasing. According to a study by the University of California, Berkeley, it is estimated at approximately 1.5 billion gigabytes, or an annual average of 250 megabytes for every man, woman and child alive on Earth.

To this must be added the products of digitization, as previously mentioned, when all that remains of the original work is a digital file, because the original support has not been preserved, or has only been preserved by accident.

Thus, from a technical point of view, we are confronted with two subsets, each of which calls for the implementation of particular preservation strategies:

- works associated with a physical medium on which the digital file is recorded, for example an educational multi-media CD-ROM; in this case, there is a stabilized, finished, definable object. The appropriate preservation policy for such items will be related to the physical preservation of these media, the material and logical reading or playback tools to give access to contents, and maintenance of the initial digital file to make it possible to guarantee its readability;
- those whose contents are computer files resident on the hard disks of computers or servers, for example a database or pages of an Internet site. In this instance, contents have been definitively separated from a defined, stable physical support. The content in question reconstitutes itself on a computer screen, on request, provided that one possesses the right hardware and software configuration. Only the implementation of true content

gathering strategies will make it possible to fix such flows, and thus to be able to archive them.

Infinite heritage?

A constantly increasing proportion of the information produced today in almost all spheres of human activity is produced in digital form, as we have seen, either stabilized on a physical medium, or accessible on line via the Internet or local intranets. Conversely, the total share of information produced in the world on traditional carriers such as print-on-paper, magnetic tape, or film is decreasing each year. A constantly increasing proportion of the information produced today in almost all spheres of human activity is in digital form, either, as we have seen, stabilized on a support, or accessible on line via the Internet or local intranets Conversely, the total share of the information produced in the world on traditional supports such as printed paper, magnetic tape, or film is decreasing every year.

Thus, from a quantitative point of view, the world's entire annual digital, or potentially digitizable, production, can be estimated at more than 15,000 terabytes[1]. This volume takes account, firstly, of all written materials produced with a view to being published (i.e. books, periodicals, grey literature), totalling 230 TB; CD and DVD publishing, representing 31 TB, cinematographic works, representing approximately 16 TB, and finally radio productions amounting to 800 TB and TV productions representing 14,000 TB.

The Web itself can be estimated at 150 terabytes. The private activity of exchanging e-mails represents a much larger volume than that of the Web, and has been evaluated at between 10,000 and 20,000 terabytes per year.

It should be emphasized that these estimates do not take into account those immense scientific databases of several hundreds of terabytes each, which constitute what is commonly known as ″the deep Web″.

Important as it is, the problem of volume, from a purely technological point of view, is not without a solution, since progress in electronics is such

1. These data come from an assessment carried out in 2002 for INA in France by Dominique Pignon of the theoretical physics laboratory of the Ecole normale supérieure de Paris.

as to allow a constant increase in the capacity of storage media, for a progressively lower cost per stored megabyte: hard disk capacities double every 18 months and the average price per stored megabyte fell from $11.54 in 1988 to $0.20 in 1999.

As an indication, in 1975 magnetic tape offered a capacity of only 30 megabytes, whereas today magnetic cartridges exist with a capacity of more than 200 gigabytes; capacities of the order of 1 terabyte have been announced by manufacturers for within less than 5 years – a capacity which will make it possible to store on one single carrier 1 million written documents of 100 pages each, or 1000 hours of video with average quality compression to (2 megabits per second), or 30,000 to 60,000 hours of music recorded in MP3 format.

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 and the robustness of the tools with which the works were created, and which guarantee the continuance of its existence, or should it be the result of a voluntarist, 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. In fact, digitization operations such as these can be performed by the institution itself, within a specialized department created for that purpose, or sub-contracted to an outside service provider, in particular when the technical equipment required involves heavy investments in a sometimes still unstable technological universe, as it is currently the case for the digital copying of audiovisual collections.

These operations can be carried out systematically as new works are acquired, thus enriching already existing collections, or the priority may be the digitization of old collections, following an appraisal and once a selection strategy has been established.

The selection strategy should make it possible to define priorities based on three types of criteria: technical criteria – for example, the most fragile

funds will be digitized; criteria of content – for example, attention will be focused on entire collections; or criteria of use – for example, those documents the most in demand will be digitized.

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.

The elusive Web

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. Thus, random samples may be taken, the search software then 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 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.

When the archiving of Web is undertaken by a sovereign state, for example within the framework of the application of legal deposit laws, the selection is carried out on the address of the domain name, which amounts to creating subsets of the Web by geographical territory. This is for example the practice of the Royal Library of Sweden, which collects sites emanating from Swedish domains. Selection can also be based on linguistic criteria. Thus, the Royal Library of Sweden supplements its national selection by

archiving sites in the Swedish language. The principles of collection will guide the collection strategies adopted by those libraries which intend, in a logic of continuity of their collections, to preserve the electronic extensions of their existing collections; for example, this could involve supplementing collections of periodicals by archiving their on-line editions; national archive services also comply with this logic of continuity of collections by taking account of the Web sites of ministries and public institutions.

Other collection strategies can be based on criteria of content or theme, which makes it possible to constitute specific archives; for example, an initiative by researchers at the State University of New York, with the support of the Library of Congress of the United States and the Internet Archive Foundation, enabled a Web archive of September 11, 2001 to be constituted. Another example is provided by France's Institute of social history, which in 1994 undertook to collect documents published on the Internet within discussion forums relating to politics, social affairs and ecology. The Institute of social history has thus gathered 900,000 messages emanating from 974 forums accessible on the Internet.

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), plans to concentrate 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 or formal categories, and organizing this prolific and polymorphous data source.

Throwaway media

Digital data, whether manufactured or collected, become heritage only when they are stabilized, authenticated, referenced and kept accessible within the framework of a permanent archiving system.

However, one of the aspects of their vulnerability resides in the extraordinary technological dispersal governing their constitution. This dispersal

simultaneously concerns both the media which will become receptacles of the data at some time or other in the process, and content, coded in the form of bits, according to a very wide range of encoding norms and standards. Indeed, these vary on the one hand according to the form of the original document – still image, text, graphic, Web page, audio file, 2D or 3D animation, audiovisual streaming – and on the other hand, within each category, according to technical and industrial developments which launch on to the market applications in proprietary formats with increasingly short lifespans and which are generally mutually incompatible.

From time immemorial, the methods and practices of 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.

With the advent of digital, the family has grown in size, in particular with the arrival at the beginning of the 1980s of the audio CD, developed by Philips and Sony, followed by the computer CD (known generally as the CD-ROM and the CD-WORM, then the general public multimedia CDI, photo CDs and video CDs, and finally, the arrival in 1996 of the DVD, over which a formats battle is currently raging. Heritage institutions have had to integrate these new publishing and storage media within their collections.

But everyone knows that storage media are not eternal. According to studies by heritage institutions such as the Library of Congress and the BNF[1], it appears that the plastic used to manufacture audio CDs, CDIs, photo CDs and CD-ROMs pressed and duplicated from a master disk, may only have a lifespan of about 10 to 25 years under average conditions of conservation and use. Rewritable disks may only have a lifespan of about 3 years before being burned (because of ageing of the sensitive layer, comparable to that of a photographic film before exposure and development) and a lifespan of about 5 to 10 years once burned, before deterioration sets in. These findings come from live tests in laboratories, and are more pessimistic than the lifespans announced by manufacturers. Tape- or disk-based magnetic media, for

1. C. Lupovici, *Technical metadata and preservation needs*, presented at the 67th IFLA General Conference, Boston 2001.

their part, offer the disadvantage of natural magnetic migration due to magnetic fields inside the sensitive layer.

Thus the new media used for electronic documents are more fragile than the media used hitherto, have even shorter expected lifespans, and therefore require actions to conserve, restore and periodically refresh them.

Ephemeral formats

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 its survival. Indeed, it exposes the content to the risk of becoming "dull", even though its supporting medium may have remained intact, because it may have become impossible to decode it, given the considerable instability in encoding standards and formats when the document was born.

Since content consists of an encoding of the whole of the information, any loss of even a portion of the code can ultimate make it impossible to read that content. It is thus necessary to preserve not only the storage medium, but also the format in which the data are physically organized.

However, code alone is not enough to guarantee long-term access. Coding brings into play a whole technical environment which intervenes not only in the creation phase but also during use, i.e. reading, by treating the code to make it understandable to the user, via for example a simple viewing interface. All these ″technical layers″ which treat the flow of bits to make readable content out of them are particularly vulnerable to technological developments which limit their lifespan, rapidly rendering unusable certain digital contents as soon as these are transposed into a new environment. Thus, the lifespan of both software applications and technical platforms is often much shorter than that of the medium itself.

Moreover, this technological ageing is not identical for all types of objects; it varies according to the use made of standards and open applications, i.e. those for which technical profiles are available, so that it will be relatively easy to upgrade them, and according to the use made of proprie-

tary formats or applications inseparable from a particular technical platform – these are known as proprietary applications. In this particular category we find most production of digital content for mass publication; it should be further noted that publishers very often integrate anti-pirate devices into their products which further exacerbate the difficulties of their preservation.

Archiving will always lag behind technological development, which is little concerned with permanence, and which it only comes across it by chance.

2.2 Memory problems

In the explosion of the Information Society, emphasis is often placed on its characteristic of enabling generalized communication among peoples and of interconnecting networks, and on its aspect of democratic progress, which enables access for all to information, any type of information, constantly, and even to create information.

On the other hand, we are concerned about the security of transactions, the protection of privacy, the porosity of the network to illicit trafficking and messages (pornography, Nazism, espionage, cyber-terrorism); Internet throws the social sphere into confusion in all its component parts.

The upheavals that this explosion causes to traditional modes of memory in our societies are less well measured. However, these upheavals, caused by a narrow array of multiple factors, are without precedent, and call into question several centuries of patient organization of one of the fundamental components of social memory, that memory which for such a long time had been organized around an archiving system using writing on paper as its medium.

Dematerialization

One of the most remarkable characteristics, and one which is evident to all, is the dematerialization of carriers. Indeed, can we really still speak about carriers, when they no longer have palpable material substance?

Contents and containers

This evolution comes from a long way off. With the advent of printing, ideas were expressed by means of artistic forms which could be reproduced mechanically, were designed to be distributed in multiple copies, and whose form was inseparable from a single carrier[1]. The value resulting from this distinction, which brings in the concept of original and copy, meant that these two artistic genres were fundamentally different from each other.

Audiovisual creation adds a new dimension to this volatility of expression, since the form of the message, and not only its content, becomes reproducible mechanically.

But the great revolution in this respect was brought about by digital technology. The distinction between original and copy disappears completely insofar as reproductions are identical to the original, with no losses.

The fragility of carriers

Carriers have development in a parallel manner to the works themselves. Signals are recorded in increasing densities, and the carriers themselves have become increasingly fragile: from tablets of clay to papyrus, parchment and then paper, the loss of durability amounts to hundreds or even thousands of years, not to mention manufacturing techniques which are insufficiently mastered and accelerate deterioration (acidity, the vinegar syndrome). And today, nobody can predict the lifespan of computer diskettes or CDs. For the latter, and according to the experts and manufacturers, the range varies from 5 to 270 years!

The encoding process is becoming increasingly opaque: written media can be read without any intermediary, but even if images on film are still directly perceptible, this is not the case with video or audio recordings on magnetic tape, and even less so when they are digitized. Reading requires a technical intermediary. This intermediation is a new point of vulnerability for our preservation efforts, as it implies that we should preserve the machine as well as the medium. But it is less and less practicable to pre-

1. Semiologists speak of autographic arts and allographic arts (Nelson Goodman, Gérard Genette).

serve technical devices which cost considerable sums to develop and, contrary to the requirements for permanent safe-keeping, are designed under the rules of technico-economic cycles for investment goods, i.e. with short lifespans and for early replacement.

The nomadic archive

All these technical considerations lead to a situation in which the dematerialization of the message is accentuated, and in which its survival depends on increasingly rapid replacement of its host carriers. Some[1] even believe that storage media no longer exist, and that there now merely exists what one might almost call a parking process, while messages live and are given permanence in the ether of the network, bouncing from machine to machine, at the whim of the caches of different servers.

Delocalization: memories with no fixed abode

The system of interconnected servers known as the Internet exacerbates this dematerialization of messages. We now speak in general of *contents*, and no doubt we should understand that if we are now living in an era of *contents,* we are no longer in the era of *containers*, and that these two aspects of the message have been definitively separated. In fact, what has been lost is the body, as if the electronic era had intensified, and by new means, an old way of thinking.

Digital systems do not of themselves impose delocalization. It is possible to imagine digital technology being used in the same way as analogue, where the message resides on a support, domiciled in a fixed, linear way. In the audiovisual world, this is indeed how things began, with digital tape reproducing analogue operations and offering the possibility of precisely locating a signal on a particular carrier. Operations of this kind are now giving way to systems of management which decouple a logical view of the information from its physical whereabouts.

1. FIAT (International Federation of Television Archives) /IASA (International Association of Sound and Audiovisual Archives) /FIAF (International Federation of Film Archives) Conference, Santiago, 1999, www-.fiatifta.org

The controlled dispersal of basic units of information to distant physical sectors is a pledge of resistance on the part of the message to the aggressions to which fragile media are susceptible. This operation imposes considerable redundancy of information, in particular for data checking and addressing systems, but increases in network flow rates, processor speeds and mass storage capacities make it possible to accept this inflation of the internal layers of information management.

Robotic systems make it possible to manage enormous masses of data (described in terms of petabytes, 1 PB being one million gigabytes) seen as one logical unit. Only the system[1] knows where such-and-such byte is stored, and takes charge of its regeneration on another carrier if it considers it necessary.

Deterritorialization

The end of archival territories

Generalized networking allows an additional leap in the delocalization process, by making it possible to manage whole of the planet's servers as a single system of integrated storage. This is what is offered by systems for the exchange of music in MP3 format such as Napster, which connects Web users wishing to publish a piece of music and those wishing to acquire it (a process known technically as *peer to peer*). Systems can even be freed from central servers, since metadata (i.e. associated data on the author, composer, title of the work, etc.) self-referencing the MP3 format are reconstituted without central intervention, the stated objective being to enable entirely anonymous exchanges.

The role of States

What is at issue where the Internet is concerned is thus the possibility of controlling the means of information. At all times in history, power has developed from such control. The control of transportation routes was

1. In computing, one speaks of HSM, or Hierarchical System Management, whereby a software system manages multiple types of storage media with decreasing access times (hard disks, optical disks, tapes) in the same logical unit, transparently to the systems administrator and to users.

the symbol of power over a territory, controlling movements for public safety purposes, to safeguard the integrity of one's territory or domestic market, but also, and for the same reasons, to control the means of communication. Freedom of expression, affirmed in the Universal Declaration of Human Rights in 1948, has always had to compromise with systems endeavouring to restrict it. The reflexes that lead States to centralize, whether through laws on the press, broadcasting, or royally regulating the frequency spectrum, whether justified by the requirements of national security or to protect the people (privacy, child welfare, the repression of slander, etc.), have driven home solid bolts ensuring control of the territories of information, the ability to censor or to prohibit publications, to constrain information, to spread it, but also to protect authors, private individuals and pluralism of currents of thought and opinions...

As has been emphasized many times, the way the Internet burst on to the scene outside any traditional official capacity, profoundly upset the exercise of public power, which was led to substitute in place of its capacity to regulate a role as a sort of co-regulator within an increasingly internationalized framework.

The destructuring of companies

What is true for sovereign states is also true for each subdivision of power. In every company, e-mail systems shake up established, proven circuits. Pyramid-shaped hierarchical organizations are being destructured: intermediate layers of management, which drew their legitimacy from micro-strategies of partial and temporary retention of information, are being destabilized by this new, horizontal, informal circulation of information. E-mail has borrowed both from traditional circuit of administrative mail and from the private telephone communication system. The second model, much more labile, is winning out in practice, and at the same time endangering all the certification and archiving systems of the company.

The principal components of the old ways are in the process of recombining in the form of what is known as workflow, at least for the productive functions of the company. No-one knows where the change will ultimately lead, but while it is happening whole sections of company information are disappearing.

New resistance to the appropriation of works

In 1777, Beaumarchais drew up the bases of an organization to defend the rights of authors, while in 1776 Condorcet, in his Fragments on the freedom of the press, had denounced copyright as contrary to the freedom of information. The technological escalation of networks has taken this debate between Beaumarchais and Condorcet further, but with perhaps unprecedented violence. Copyright for a long time had been seen as the victim of history, crushed between authoritative and centralized regimes on the one hand which reduced the individual appropriation of thoughts to a system for collective benefit, and liberal systems on the other which allowed the private alienability of intellectual or artistic works to develop as for any other goods.

But today copyright is in the dock, and we see an entire fringe of civil society challenging any obstacle to the free use of works, and deploying a wealth of inventiveness for the purpose. The heart of this inventiveness is the anonymization of individual movements on the network, and the dilution of places for information storage. This is additional proof of the close connection between information storage and modes of appropriation.

The new regime of copying

The disappearance of the difference between original and copy, and the facility with which copies can be made, have led to thorough revisions in many legislations[1]. Publishers, producers and authors are all increasingly mobilized around efforts to stop piracy, in general by means of encryption and anti-copying techniques and devices, in particular on DVDs. We are thus moving towards a return to the concept of the original, but an artificial original created by the voluntary downgrading of the possibility of copying. These protection systems are however likely to discontinue the exceptions traditionally made for private copying, and to make it more difficult for heritage preservation institutions to accomplish their mission.

1. For example, the *Digital Millennium Copyright Act* of 1998, enacted by the Clinton administration in the U.S. or the *European Directive on the harmonization of certain aspects of copyright and related rights in the information society,* adopted in 2001.

Delinearization

There seems to be no end to the delinearization of the message, as if a law were at work generating vectors of communication which are initially linear, but over time develop in the form of a flow, then in a second phase are spatialized, so that they can be represented in space. Discourse is spatialized in strategies of memory, in rhetorical figures, and is spread in the written form. Writing was spatialized as it evolved from the scroll to the codex, and later through he invention of space markers such as pagination, tables of contents, etc. Even the text on computer screens has moved from on-screen scrolling to the graphic interface.

Further, text hypertextualizes itself to offer other reading itineraries than its initial linear development, unfolding in ways which are specific to each reader. What once was only avant-garde poetry[1] becomes a usual mode of writing and reading.

Digitization has also allowed the spatialization of audiovisual flows. It is now possible to stop, capture, annotate and structure them, and to read them again at speeds other than that dictated by the machine.

Moreover, the general organization of networked servers makes it possible to transform gradually all these flows into a gigantic, single mass. Radio and TV messages, but also the whole of the written system will be able to live in time which has stopped, offering the possibility of going back in time, of instantaneous access to all the historical density of time which used simply to pass by. Formerly, today's news used to drive out yesterday's: now, newspapers put their archives on line. And all this information is indexed on the same basis, from the most recent to the oldest, by search engines which constantly crawl over the Web.

Information will only disappear if there is no more room on the server. Which is not without posing its own problems: some people call for a right to memory lapse vis-a-vis this open, public, electronic eternity. The Internet multiplies this phenomenon, insofar as a text can be fixed, edited and copied

1. R. Queneau, *Cent Mille milliards de poèmes,* Paris, Gallimard, 1961 (100,000,000,000,000 poems, tr. Stanley Chapman).

several times, and then resound like an echo from one end of the Web to the other, only to die out according to the laws of a new kind of physics which has yet to be invented.

The protection of personal data becomes a particular concern in this context, since it is so easy for information to be appropriated by individuals, or for occult databases containing personal data to be set up and put to uses which are not always lawful.

The opening up of text, a memory with neither beginning nor end.

Since the invention of printing, we have been living in a regime in which texts were published in their final state. Later versions may revisit this final state, and authorize modifications or corrections, but the weight of the publishing system makes this difficult in practice.

Computing has reopened text in its state of innate incompleteness.

We all know how difficult it is to correct a manuscript: overwriting words quickly makes the text illegible, and methods learned in school make the passage from draft to fair copy one of its fundamental values. Typed text has benefited from successive generations of ″Tippex″, but its use was necessarily limited, and one generally lacked the courage to retype a whole page. Beyond a certain stage, only essential corrections could be allowed. Word processing radically reversed the proposition: now, a text is never finished, so easy and tempting has it become to add a final touch here or there. Only printing out the document on paper still obliges one to recognize that the document's development has reached its final stage.

With networked electronic publishing, nothing further stands in the way of the continuing evolution of a given text, about which all that can only be known with certainty is its state at a given point in time. The fundamental regime under which a text was published in a final state and on a particular date no longer has any support from technology. If it can survive, it will be for other reasons.

Moreover, text no longer has borders. It lives in an ocean of texts, with which it maintains a number of relationships – incoming and outgoing

hypertext links, indexing in massive resource location tools. What constitutes a work, if it has no borders either in time or space, and how much of it should we try to preserve?

An entirely new mode of quotation may emerge from this. Formerly, a work incorporated quoted extracts into its own self. What need is there still to proceed thus? Footnotes can already be replaced by URLs (Uniform Resource Locators), and the text itself could be simply a virtual patchwork of various quotations. But what will become of this text if one of the addresses referred to ceases to be active, or if the text is changed? One can imagine, as if in some Borgès-like tale, that through a cascade effect one change to a text could instantaneously modify thousands of other texts, or even the entire Web.

The ease with which others may be quoted is a source of the conflict mentioned earlier between free use and copyright. For many, the network has become a gigantic quarry, where a work is built up through a number of loans from other works used as raw material. This phenomenon is particularly sensitive today in music, and will develop quickly into the audiovisual arena as soon as networking flow capacities allow it to.

All these phenomena have always existed: works were quoted, they could have several versions, they could be read diagonally, from back to front, they could be compared with others, and thus their authors influenced the authors of other works... but all this occurred on an infinite scale, outside time. Texts could be ascribed and dated.

Technological instability

As far as we have been able to discern until now, the Internet is in a state of permanent technological evolution. The price to pay for dematerialization, for an ever greater abstraction of code from carrier, for spatial as well as temporal apprehension of the messages, is this technological race. The physical effort to fix information on a carrier becomes negligible if one compares it to ancient steles, or even the heavy processes used in the chemical industry to manufacture photographic film, rotary printing-presses or the pressing of vinyl discs. The turning of the technological wheel, and the rapid obsolescence of machines which ensues, is inversely proportional:

megaliths lasted a few thousand years, books a few hundred years, audiovisual productions a few decades; and the network of networks has existed at most for just a few years. Such, approximately, are the lifespans of technological innovation.

In other words, a given content published today on the Internet has no chance of being readable in ten years, unless regular format migrations are organized in the meantime.

Computing is sufficiently widespread today for us all to have learned about these problems of formats, software versions and incompatibility. Even if hardware (computers) were to remain stable, even if storage media were permanent, even if basic computing languages (ASCII etc.) were stabilized, a text can still become illegible, and even more so its form unmodifiable, if one does not have the right version of the software which was used to create it.

Information invaded by noise

The volumes of information currently being produced are on an altogether different scale from those produced before the advent of the electronic systems. There has been a radical change of scale, contemporaneously with a reorganization of document management.

Current systems of archiving were inherited from the nineteenth century, and many research techniques, especially for historians, are in line with this. Already, as the paper era draws to a close, we can wonder at the kind of relationship the researcher can have with these masses of documentation: more than 50,000 books are deposited each year with the BNF (National Library of France), several tens of thousands of collections of periodicals, nearly 110 kilometres of shelves of administrative archives are lodged annually with the French national archives... as these enormous masses grew, until the 1980s referencing instruments, catalogues, inventories and archives remained very modest.

The Internet has ushered in a new dimension: it catalogues itself. Directories, search engines indexing the Web, indexes of indexes are emerging everywhere. After the great age of silent catalogues, a prolific, noisy era of indexes bulging with information has arrived.

Full text engines and indexing systems were until recently reserved for the tiny world of documentation. From a computing point of view, this was a niche market compared with the world of relational databases for management. And then the development of the Web changed the deal. Today it is one of the most promising sectors of computing, and one in which research is most flourishing. It certainly needs to be, given the immensity of the Web, at least in terms of the number of documents it contains. If the user is not to be met with a deluge of responses to the slightest query, search engines will have to develop relevance strategies, using for example semantic weighting techniques based on the hypertext environment of a group of words, for example[1]. All texts are no longer equal on the Internet: just as in a galaxy, some are close to the nucleus, where hypertext links come and go, while others are in cold regions, or even in the unknown galaxies of the centre. Apparently, this galaxy rather resembles a bow tie in shape[2].

That a topological science should model the moving distribution of information within cyberspace shows clearly that we have changed scale in our way of apprehending human knowledge. The advances in the social sciences of the 1960s, which moved away from interpreting texts towards building a science of knowledge, required hours of meticulous study of archives and libraries, and all things considered apprehended only a narrow corpus. Today the Internet makes available to research of this kind an immediate and total presence of knowledge.

The end of exhaustivity

Earlier methods of management of memory were founded on exhaustive inventories, the painstaking, fine cataloguing of information. The dual origins of archiving systems, for the preservation of heritage on the one hand, and policing on the other, is to a great extent responsible for the compulsion to be exhaustive. But the very tension inherent in research tended towards exhaustiveness. All serious researchers owed it to themselves to investigate their subject punctiliously.

1. For example, the PageRank system used by Google.
2. Mapping of Web links demonstrated in 2000 by a study conducted by researchers from IBM, Compaq and Altavista entitled *Graph structure in the Web* , http://www.almaden.ibm.com/cs/k53/www9.final/temp.gif

The explosion of easily accessible information reduces the dream of exhaustiveness to nothing. The essential problem becomes, somewhat obviously, how to apprehend these masses of information. New tools and concepts will have to be developed to enable us to work on these great masses, and, according to the approach required, to grasp them using fuzzy logic, rather than controlled logic on the level of the documentary atom. This is undoubtedly not new: inflation in quantities of paper already precluded an approach on the level of the unit, but this had not yet been radically appreciated. Current tools for indexing of Web open the way for a quantum revolution in the management of sources.

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. The paradigm has already been transformed, and the devices in place are unable to deal with this brutal advance of information technologies, its strangeness, and the quantitative inflation which it causes.

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 revolutionizing old memory and archiving devices, for, without our realizing it, our societies today are making great holes in their collective social memory.

Chapitre 3 Institutions of memory faced with the state of practice in the field

It is often said that the Museion of Alexandria, the temple of the Muses, was the ancestor of our museums, and that Pisistratus created the first public library, in Athens in the sixth century BCE. The archive is probably contemporary with the first states based on writing, and is simultaneously that which is old, and that which resides in places of power[1], the trace of the acts of authority that were kept by the archons.

This tripartite division of places of conservation of the collective memory lived on across the centuries, and, expanding to the whole of civilization, crossing the great divisions between public action and private initiative, took on a universal character. We are undoubtedly concerned here with constants which define the limits of the field of possible methods of preservation of the collective memory[2].

Even if we have stressed the invasive character of digital technology throughout the social sphere, and to a large extent throughout most modes of human expression, each of these great functions of memory is not affected to the same extent by the preservation of digital heritage. The fundamental differences relate to the nature of the document at the heart of each institution, and the degree of digitization at each phase of circulation of the document.

There exists a whole range of nuances, from documents whose digital copies may be considered to be so much complete substitutes for the originals that original and copy can no longer be distinguished, to documents which undergo more or less serious deterioration, and documents of a radically foreign nature to digital technology, copies of which will never be anything more than simplistic images of the originals.

---

1. J. Derrida, *Mal d'archive*, Paris, Galilée, 1995 (tr. Eric Prenowitz, Archive Fever, University of Chicago Press, 1996).
2. One ought to add here monumental heritage, but that can be considered here as another kind of museum, whereas oral transmission represents another kind of library. "When an old person dies, it's a library burning down", as Amadou Hampaté Bâ has said.

3.1. Collecting digital items

Virtual museums and museums of the virtual

Museums often preserve unique or rare objects (until the XVIIIth century, we used to call them chambers of curiosities or rarities) and which are always of value. The entire art market is based on the valuation of physical objects. However, we know little about an economic model potentially underlying the valuation of immaterial objects. Already, mechanically reproducible objects, such as photographs, somewhat artificially recreated the concept of the dated and/or marked original, and thus found their place in the art market.

In general, museum exhibits are more than two-dimensional, and the notion of originals is predominant. Digital technology does not yet offer a way of making it possible to imagine a virtual digital object completely replacing a real, three-dimensional physical object. This all the truer for monumental heritage.

Digital technology will thus concern museums only at the margin, primarily in their functions of diffusing, and sometimes of safeguarding contents after the death of the objects concerned, with relatively little preservation work. Digital technology is an extraordinary asset for diffusion and democratization, because it allows very widespread development of the representational function of a work, either on line or on multi-media carriers including a critical apparatus, or even locally when the work has become too fragile to be presented, but this representation is not the work.

Moreover, the fact remains that digital technology is increasingly present at the source of various types of artistic creation: authors' manuscripts, the design archives of architects and designers, the works of photographers, sketches for stage or film sets, costumes, archives of drawings for 2D or 3D animated cartoons, models used in town planning, graphics and other artistic creations using the entire range of image processing possibilities...

Generally, these are the preliminary or intermediate phases of a work which will, according to case, become monuments, print-on-paper documents, digital films or videos, industrial objects... or websites, all of which

are forms of publication whose preservation is possible according to known methods.

But what is to be done about all the original materials, the traces of the genetics of the work, about which our age is so concerned? Their situation falls between two stools: as long as they remain within the closed loop of creation, they are single objects, but because of their digital nature they are infinitely multipliable products. Their origins as private property, moreover, will have led to the greatest variety in the computing formats used, making their preservation for posterity particularly problematic.

The e-archive

Archives, in as much as they result from the activities of public or private organizations, constitute social and historical heritage of the first order, which makes it possible to transmit to future generations traces both of the exercise of the authority of the state, and of the activities of peoples and groups. At once instruments of proof essential to the operation of a modern society, and privileged carriers of the collective memory, archives are generally entrusted to the care of a specialized institution or service whose mission is to gather and preserve a whole set of documents, studies, reports, statistical data and other items... all carriers of written deeds or facts able to inform the people of today and tomorrow about their activities and those of preceding generations.

These collections, once they have been duly selected, classified, catalogued, stored and treated according to methods which have been tried and tested over several centuries – was the archive not born with the invention of writing, in Mesopotamia? – will quietly occupy metre upon metre of shelving in preservation storerooms, waiting for the moment when a reader comes to consult them.

Today, this processing sequence, hitherto perfectly controlled, is now wavering. The premises of these upheavals have been observable since the 1960s, when large organizations started to use computing resources to process information, especially great masses of data such as demographic data. This movement has accelerated over the last ten years: not only has computing penetrated all areas of organizational activity, but organizations have

adopted it for the purposes of internal communications and interface with the outside world.

Indeed, if one considers the increase in quantity of the information produced, brought about both by increases in processing capabilities – for example, the multiplication of economic indicators – and by the ease of writing and reproduction which generate high degrees of both redundancy and distortion of the original, and if one considers at the same time the extreme volatility of this information – e-mails, notes, reports are fixed on a material medium only insofar as the sender or recipient takes the initiative of doing so – one can easily perceive the disorders to which the archive is subjected by digital technology[1].

We are still far from the paperless society. That idea even seems to have moved away from us in inverse proportion to the extent to which society has been computerized. For archives today, paradoxically the most concrete effect that computing will have had will have been to increase considerably the consumption of paper, and thus to enable its preservation, because of the exceptional expansion in the production of texts which has been made possible by office automation.

But, as we have seen (cf. chapter 1), this phase of computing is coming to a close, and in the last few years the e-mail with its attached files has tended to invade administration and management systems. Moreover, most legislatures have legalized authentification systems for electronic documents, removing one of the last necessary functions hitherto reserved to paper documents.

Although the movement towards digitization is irreversible, very few institutions today raise the question of how their memory should be organized for future generations. In the past, this responsibility was delegated to preservation institutions which, by virtue of their closeness to the originating organizations, collected documents once their practical value (Administrative Utilisation Period) was exhausted, selected them according to well-established criteria, and organized them into perfectly identified collections

1. Reportedly, many tens of millions of e-mails were deposited with NARA (the U.S. National Archive and Records Administration) at the end of the Clinton Administration.

which respected their organic composition. These interventions, external to their production contexts, enabled documents to acquire the status of archives[1]. However, with digital technology, it appears more and more necessary for preservation levels to be integrated into the production chain at the time the document is created, not only in terms of classification of contents, as at present, but also in terms of carrier and format. Otherwise, if no provision is made, nothing will remain of it, and there will be no archive.

It is thus for institutions to take the necessary measures for the safeguard of the digital information which they produce, in close co-operation with the heritage institutions which will have the task of preserving it for posterity.

Electronic libraries

Libraries will be deeply concerned by the revolution in writing, and indeed by all phases of this revolution from the act of creation to digital dissemination of the work over networks. We are of course speaking here about libraries in their function of preserving and communicating objects published in quantity, in whatever medium, not in their museological function which is often also fulfilled by libraries (incunabula, seals and coins, original manuscripts, etc.)

Large libraries have already begun to implement digitization programmes of their old collections, but, beyond that, the existence of "born digital" books and newspapers which are read on the Web profoundly modifies library operations (collection, conservation, communication).

Many countries' legal deposit laws include the deposit of databases and multimedia works on physical media. The majority have not included, until recently, the deposit of on-line documents[2].

Some digital publishing products preserved in libraries do not pose particular conservation problems: audio CDs in particular can be easily duplicated without loss, thanks to their digital output.

1. C. Dhérent, *Les archives électroniques,* Paris, Direction des Archives de France (French National Archives), 2002.
2. Denmark, Finland, Norway and Sweden have integrated this into their legislation. France is about to do so.

Video DVDs on the other hand will gradually present increasing difficulties because of their in-built anti-copying devices, which are beginning to be legalized under authors' rights or copyright laws.

The complexity of multi-media products is greater still. They do not exist in the absence of a hardware platform (microcomputer or game console) and software layers specific to that platform. The risk is thus very great that the game will no longer be playable once the platform, or even the version of the operating system installed on the platform changes. This is already the case for a number of the earliest video games, whose consoles have disappeared or are no longer in working order (Amiga, Atari, etc).

One solution to this problem is to emulate old operating systems on a contemporary machine. There are still some nostalgic independent developers who voluntarily develop emulators to make these old games or programs work. It is doubtful however whether this solution can be viable in the long term, and moreover, if libraries have to cope with this function on a permanent basis the cost will probably be exorbitant.

In a more general way, the preservation of computing languages "in working order", with their source code editors, libraries and compilers etc., is a question addressed today only by software publishers – but many software publishers disappear quickly – and a few amateur collectors.

While collecting distributed multi-media products can be brought down to known content categories related to mass-produced carriers, the development of the electronic on-line library entirely transforms collection and conservation methods.

On the one hand, this mode of collection addresses "born digital" objects, and thus avoids the task of digitization and reduces the storage capacity required, since the documents exist in native mode and not in image mode. For structured textual documents, this mode of collection offers the possibility of much finer, and automated cataloguing. In addition, here too, publishers tend to develop anti-copying devices (downloadable music in streaming mode, etc.), hence the burden of collection is reversed, and libraries must become pro-active where previously they simply waited for works to be deposited. Finally, libraries are forced into managing an on-line digital stock, with all that that requires in terms of computing logistics.

While this latter workload can easily be absorbed by large libraries, where the collection of on-line products can be dealt with alongside the digitization of old documents in large databases, on the other hand it radically changes the nature of the investment necessary for small libraries, and calls into question concepts of lending and methods of consultation.

Initially, the consultation of digital collections can be organized in the form of computer networks internal to the library, but this should naturally develop towards the virtual library. The principal constraint, for documents outside the public domain, is managing publishers' and authors' rights. Ultimately, the very concept of a reading library is diluted into what could become new methods of financing of public reading, based on the one hand directly on publishers' sites, and on the other hand on the sites of some large central libraries. Thus, a whole new economy of public intervention would have to be rebuilt, with a view to democratizing reading.

As regards preservation, the audiovisual sector has always been a real headache for heritage institutions obliged to preserve in optimal conditions not only a multitude of media threatened with physical deterioration in the short term – films and magnetic tapes in a variety of formats and standards – but also a whole range of reading devices in order to continue to be able to read them, not to mention the irremediable, progressive loss of signal quality with every copying cycle. In response to all this fragility with respect to preservation, the digital way lies precisely in dematerialization: once the message is detached from its material carrier, it can lend itself to any kind of transfer on to a homogeneous host carrier, the format of which will have been defined by the preservation institution itself. If we add that the transfer speeds of computer archives bear no comparison with real-time analogue copying times, for the first time perhaps in its history the audiovisual sector has some hope of controlling its memory. But the price of getting there will be high.

Web and flow

Immaterial, unstable, volatile, multi- and hypermedia, unfinished and infinite, the Web raises frightening questions concerning the sciences of heritage, whose methods and practices have been patiently constituted on the basis of collection and preservation of stable material carriers.

Nevertheless, preservation initiatives are today multiplying, and even if all the problems are not yet solved, outlines of answers are showing us the right way to go in order to organize, at long last, the memory of the Web.

Managing mass

One of the obstacles most often emphasized is that of mass, of a volume of data believed to be so vast that it would be impossible to control. In fact, and while it is true that the Web is growing fast, it is stabilizing. The response of the heritage institutions to this mass of data, according to their sector of competence or centre of interest, consists in partitioning the Web, constituting subsets by territories, themes or languages. The large national libraries and archive institutions, for example, set out to preserve that portion of the Web which falls within the limits of their national heritage.

Constituting a hyperarchive

Practices of partitioning the Web must have as their corollary the interoperability of systems – i.e. the capacity to access in a standardized way all the collections in existence; this step is decisive if the Web is to be reconstituted in its full dimension. It is all the more essential in that the universe of the Web is characterized by hypertextuality. This pushes back the borders of documents to open on to an infinite number of reading paths, so that the direction of the reader's progress is determined by a series of moves from link to link, from tag to tag, either within a site, or from one site to another. In response to the hypertextuality of the Web, we must consider the generalized constitution of a hyperarchive.

Containing the volatility of contents

Another difficulty faced is that of the volatility of data: the Internet is a medium of flows, characterized by their transitory nature. The Web exhibits extraordinary instability in terms of contents: 70 % of Web pages have a life of less than 4 months, and only 10 % of .com sites remain stable for more than one month. Responses to this instability of content can be found in the procedures and tools used for collection. Thus, during sweep campaigns, search robots can detect any pages of the sites visited which have been modified or refreshed. This information feeds an sweep plan

which controls the collection process according to the page modification rate: the more unstable a site is, the higher the frequency of collection will be set, and conversely, the more stable the site is, the lower the frequency of visits.

Exploding convergence

Media convergence is another characteristic of the Internet which is not without raising serious difficulties. Since all types of information, be they texts, images, sounds, or audiovisual productions, can exist in the form of digital data which can be manipulated, they naturally find their way to being broadcast on the Web, which accommodates all forms of expression of earlier media, in some cases live on Web-television or Web-radio. Media convergence is not simply the coexistence of earlier media, but more their fusion into a regime of interdependence relationships where sense, completeness of information, is worked out in an interactive way by one form of expression rebounding on the other. This convergence, cemented by hypertext links, makes the Internet a medium in and of itself, not just a simple carrier of earlier media.

The state of current practices and thinking provide the beginning of a solution: in particular, the need to deal with documents on a differentiated basis according to media type. Thus, during the collection process, autonomous collecting windows will be devoted to the uninterrupted recording of streamed audio or audiovisual data according to principles altogether rather similar to those used in the collecting for heritage purposes of radio and TV programmes.

But to make it possible to reproduce the dynamic nature of these contents during consultation, while avoiding enormous data redundancy, each object will be treated from two points of view, distinguishing content from structure. Thus, during collection, a distinction will have to be made between on the one hand the informational content, which will give rise to a documentary description, and on the other hand its organizational structure, i.e. how it was written and formatted and how it relates to its environment, which will give rise to a description of structure. This route remains the subject of experimentation on large volumes of data, and should facilitate restitution of the dynamic reality of the Web.

From the non-finishedness of objects to the unity of the document

In a non-finished universe, the question of the unity of the document is posed with particular acuteness.

From a technical point of view, some heritage institutions consider that an archive object is defined by that page or file which can be located using URL data and is dated in chronological order from first publication. Other institutions consider the documentary unit to be the site or group of sites, a portal for example, which will then be described and dated. It should however be borne in mind that, from the point of view of restitution, new levels of granularity constituted by users themselves will superimpose themselves on the objects thus defined, for example according to the grouping of themes around key words or relationships of proximity. It is thus necessary to set up restitution systems able to reproduce all the original meaning of the archive.

Remedying technological instability

In another area, technological instability is often advanced as a heavy handicap to constituting a memory of the Web. Indeed, the very vitality of the network, where one new piece of software drives out another, where a new application pushes back the limits of the possible, brings batches of perpetual evolutions in standards of encoding and writing formats, all of which weaken still further the possibilities of preserving digital archives. Here too, those heritage institutions which have taken the measure of the difficulty have suggested solutions – migration, emulation and encapsulation – which make it possible to envisage long-term data preservation. But it will no doubt be necessary to advance resolutely along the road towards defining a standardized preservation format for all archiving institutions.

Preserving and communicating the immaterial

Finally, constituting an archive of the Web means fixing its contents on mass storage media, and consequently facing up to the last challenge of the Web, that of its immateriality. However – and this question brings us back to that well-known territory of the respective advantages and disadvantages of mass storage media – the choice of storage media and architecture

runs into new constraints, in particular those of access and restitution methods. Indeed, we shall need to conceive of an organization of the archive which gives access both to its historical depth and to its spatial dimension, making it possible to surf dynamically among mutually interconnected objects of the past. Integrating this space-time dimension into the organization of the archive will be no easy matter, and still deserves research and development efforts which will have to be based largely on the uses of this new heritage.

The immemorial aspects of digital objects

Digital society brings together, depicted in a succession of small strokes, a whole set of knowledge about people. At the beginning, this mainly consisted in the computerization of data produced by administrations – personal details, incomes and taxes, criminal records, social security, courses taken at school, etc. – to which were then added all the data resulting from the relations which individuals maintain as customers with various suppliers – consumption and invoicing patterns, bank accounts, insurance policies, etc.

The first wave was thus limited to the collection of management data. To preserve personal freedoms and privacy, many States raised barriers to the interconnection of these various archives because of the fear of a ″Big Brother State ″ which haunted their populations.

Today digital data go much further: they may relate to all e-mail exchanges, personal web pages, the recording of surfing paths from site to site, telephone numbers called and cost and duration of calls, and in future even the recording of telephone communications themselves, video surveillance data, etc. as the digital society becomes capable of logging the traces of everything any individual does.

But digital data can be even more intimate, and relate to persons in their own bodies. Think here of the increasing use of digital scanning of the body for medical reasons. The developments of genetic engineering and medical imagery will multiply in considerable proportions the digital traces left by individuals.

Today, most of these data are not preserved, or are dispersed. They are neither consolidated nor capitalized on, and do not yet constitute the digital archives of individuals. But technically, nothing prevents it. What will be the limitations on the digitization of the individual, covering all components of the body, constituting a digital double of oneself throughout life? From one's genetic print to love letters, holiday photographs and Web videos, one's whole life could be written down in the same digital medium, infinitely and indefinitely easy to handle and to restore.

Here we have potentially all the elements of an absolute memorial of Humanity, from our common genetic heritage to each of its concrete occurrences. To prevent the dead from seizing the living, to avoid being buried under an avalanche of memory overload, it will be very necessary to write new rules for a work of mourning, new funereal rites for the information age.

3.2. Digitization

A certain number of institutions – museums, libraries, archive centres, scientific centres or cultural institutions –have already adopted, to varying degrees, the route of long-term preservation.

Some have published guides and recommendations[1], and we will discuss the broad outline of these below. Others have participated in experimentation and research programmes. All agree to recognize the need to take into account the purposes of preservation from the moment a digital document is originated – preventive safeguarding – and to maintain this concern at each phase of the life of the document.

Elaborating a strategy

Whenever digital information is created, updated or provided, a total long-term strategy must be defined, taking account of its needs for permanent preservation, principally for the following reasons:

1. See for example, *Guidelines on best practices for using electronic information*, ttp://europa.eu.int/information_society et *Principles for the preservation of digital heritage*, UNESCO, 2003.

- a variety of hardware and software resources must be maintained to provide access to contents, enabling the latter to be consulted long after the end of their normal period of current use (for example for legal or historical reasons)
- preservation devices must guarantee readability of data, whatever material treatments the latter may undergo
- consultation or updates may be carried out by other people than those who originated the information
- multiple consultations, in connection with other sources, must be enabled.

In developing this strategy, all the actors concerned must be closely involved: curators, archivists, librarians, computing engineers, technicians, etc. One possible solution consists in setting up from the outset a multidisciplinary group to define and maintain this strategy. The following elements must be taken into account:

- selection and identification of the digital data to be preserved
- establishment of rules of organization and classification, in particular the definition of the documentary unit and the categories corresponding to that definition
- definition of the standards and specifications to be used to ensure the independence of data from media and to guarantee permanence
- identification and attribution of responsibilities for each stage
- users' needs
- legal opinions on the property of collections, their diffusion and conditions of use
- implementation of an evaluation plan to control and correct the programme
- implementation of a technological watch centre to survey and follow up new developments in systems
- definition of policies for training and internal information
- the exchange of information and monitoring of good practices of related institutions.

At this stage, it is fundamental to define procedures and rules for the material and documentary treatment of digital objects, in order to be able, thereafter, to accompany them through each stage of their life.

Each document will have to be clearly identified and described in terms of all the fundamental data which constitute it.

Digitization techniques

To turn a paper document into a digital format, there are in general two principal solutions.

First of all, image mode, which consists in scanning the document to obtain a more or less high-resolution image of it; current scanners enable resolutions from 300 to 600 dpi (dots per inch). Second, alphanumeric mode, which consists of encoding an already scanned document into an electronic format with the assistance of an optical character recognition (OCR) tool.

There also exist vectorization tools used mainly for the treatment of graphics.

In image mode, the size of the file to be preserved is larger, of the order of 50 kilobytes per page, compared with a few kilobytes in the case of alphanumeric coding. One possibility consists in combining the two solutions: preserving images in image mode and treating texts in an OCR system.

In the case of OCR, a check must always be made by an operator using correction software.

Still images are also treated by scanning, using an appropriate standard of resolution according to quality and intended use; the file can then be compressed in order to reduce its size and thus to facilitate storage or diffusion over a network.

In the case of still images, animated images and audio, a problem arises concerning the compression of the digital signal directly resulting from digitization (270 megabits per second in the case of video). Compression is

regarded as non-destructive only if it is completely reversible, i.e. if one can restore the file identically to its non-compressed state from the compressed archive.

Whatever the medium, but it is particularly true of audio-visual recordings, before undertaking any digitization operation it is necessary to raise the question of the purpose to which the recording is to be put, so as to define the degree of resolution needed for the digitization process and the compression ratio, with the aim of:

- making a backup copy of an old analogue carrier which will disappear in the short or medium term. The ideal solution in such cases is to choose maximum possible resolution and zero compression, or at least zero destructive compression (in video, this is the case for the Digital Beta standard, and for audio, for flows compressed from 50 megabits/second to 1,5 kilobits/second). From this position of principle, it will often be necessary for financial reasons to make compromises between quantity and quality of the original, which may not deserve such an investment. Indeed, MPEG 2 above 8 MB/s approaches broadcast quality.

This stage can provide an opportunity to proceed to corrections or restoration of the original medium.

- making a publication copy, i.e. a copy intended for re-use (print-on-paper publishing, video, audio, television, radio, etc.). The quality selected will depend on the usual quality range in use in the particular publishing universe.
- making a copy for consultation purposes. The choice can be broad, according to whether one is in a local environment where the volumes can be high (up to several megabits/second in MPEG1 or MPEG 2 video) or on the Web where volumes will not exceed a few tens of kilobits/second on dial-up telephone connections, and a few hundred kilobits/second on broadband networks (MP3, MPEG 4, Real audio or video, Windows Media, Quicktime).

During a digitization operation, various types of copies will frequently be made in parallel. Each copy will be stored on a medium chosen according to intended use. Media very frequently differ (hard disk, magnetic tape, optical disc) according to copy type (backup, publication, consultation).

After processing, it is necessary to carry out quality controls and to check the recording and associated data.

If digitization hardware and software has to be acquired, it is recommended that a clause be included in the purchase contract forcing suppliers to make available means of recovering the data generated with their applications, to facilitate their long-term preservation and use. Clearly, the purchase of products based on stable, open standards offers the best guarantee of uninterrupted service.

In the case of subcontracting, the schedule of conditions must guarantee that the people who will maintain the data will have the hardware, software and documentation necessary for long-term use.

File formats

Given the number and variety of standards and norms, it is recommended that a standards universe be selected from the outset, and organizations should as far as possible stay with it. In addition, formats chosen should always be based on open, multi-platform international standards. These are categorized according to the type of object they deal with:

- for still images in dot mode, the most widespread formats are the following: TIFF (Tag Image Archive Format), which corresponds to data obtained by means of a scanner; GIF (Graphics Interchange Format), found in Internet applications; and JPEG, a compression format which is tending to become dominant on the Internet
- among the graphic formats which make it possible to preserve the structure of a graphic, one may cite for example, CGM (Computer Graphics Metafile), the standard for vector graphics, which provides a good guarantee of permanency
- video formats, including in particular the standards MPEG-1, intended for CD-ROMs, MPEG-2 for digital television and DVDs, MPEG-4 which is more suitable for interactive applications related to multimedia and the Web, and MPEG-7 which is more a standard for research -oriented description; AVI (Advanced Visual Interfaces), QuickTime and RealVideo are proprietary formats used primarily for diffusion on the Internet

- Audio formats, in particular MP3, in very widespread use on the Web, Wav (from Wave audio) and MPEG for its audio component
- text formats which treat the structure, characters and presentation of the text. The UNICODE standard for example deals with alphabets and ideograms, and is gradually replacing the old ASCII (American Standard Codes for Information Interchanges) system. RTF (Rich Text Format) makes it possible to preserve an enriched version of a text while remaining in theory fully convertible to all other formats. A great number of proprietary formats exist in different forms in this area, where new versions quickly drive out earlier generations. No format can really guarantee permanency. It should be noted that PDF (Portable Document Format) allows consultation on different platforms. Although this is a proprietary format (owned by Adobe), it is an open format, and is tending to spread as a *de facto* standard for the consultation of bulky text documents, which has led to its adoption by many libraries and websites. There are also standards for the description and logical processing of text documents which proposes a mark-up language to international standards such as SGML (Standard Generalized Markup Language) and its extension XML (Extended Markup Language) as well as HTML, which has become very widespread on the Internet.

Media

Two families of storage media are in use for digital data:

- magnetic tape media, the oldest and most fragile media type; these mainly consist of magnetic diskettes, cartridges, and DAT (digital audio tape) and DLT (digital linear tape) cassettes. The principal advantages of magnetic tape media are their storage capacity (300 gigabytes in 2003) and lowest cost per kilobyte. Their principal defect is their slow access time. Indeed, they do not allow information to be read directly, but must be unspooled and the file copied on to a hard disk before it can be used. These media are thus reserved either for backup functions, or are used in association with media with faster access times in robotized units allowing for differentiated access times, or dedicated to the very high storage capacity needs of audiovisual recordings. In the very high capacity area, current offerings are concentrated around Quantum's S-DLT, LTO from IBM and

Hewlett Packard, and S-AIT from Sony. These three standards offer storage capacities in excess of 100 GB.

- optical media, in general optical disks which allow direct access to information, but which have lower storage capacities. Many optical-disk-based solutions have existed, generally known as DOD (Digital Optical Disk). These are non-standardized solutions, and as each only occupied a niche market, they have not enjoyed continuity. On the other hand, the variations of the audio CD launched by Philips and Sony in 1980 have had considerable success:
  - — These are the CD formats CD-ROM (pressed), CD Worm, (Write once, read many) and rewriteable optical disks, which combine two storage technologies to obtain access speed, density and the possibility of data re-recording,
  - — and DVD formats, of which the video version has achieved very fast and widespread popularity. The DVD-ROM, which offers a capacity eight times greater than that of a CD-ROM, exists in various rewriteable formats (DVD-R, DVD-RAM, DVD-RW, etc.), and a battle of the formats is currently raging among manufacturers.

Beyond these external media, the continuing fall in the cost of hard disks has begun to make these a credible medium for preservation purposes.

The choice of a storage medium is essentially related to the nature of the data to be stored, and in particular its storage capacity requirements. Choice also depends on the uses and functionalities of consultation and access envisaged. In all cases, it is recommended that a single media type be chosen, or failing that only a small number of media types, in order to homogenize collections as much as possible.

To ensure the physical conservation of these media, it is essential to comply with certain rules concerning temperature and humidity. Standardization organizations such as the International Standards Organization, the American National Standards Institute and the International Council on Archives recommend re-recording at intervals of about 10 years.

3.3. New laws of preservation

Continuous migration to new media

Let us state once again that, in the analogue age, the permanency of a work depended on the physical resistance to ageing of the elements of which it was made: carrier, ink, magnetic coatings... but whatever this resistance, all media will ultimately die. This will take much longer for a very hard, homogeneous material on which the method of inscription is crude (granite steles, pyramids, clay tablets) than for a videocassette for example. But if it is not regenerated periodically, the medium will die because cycles of analogue regeneration gradually erase the inscription.

In the digital age, it would be very hazardous to bet on the durability of this or that medium, marked, like so many of the great majority of modern consumer goods, by short life cycles and accelerated technological obsolescence. On the other hand, binary digital signals are highly resistant to copying. The preservation of a computer file will thus depend on the systematic organization of migration cycles from one medium to another, and as a further guarantee, backup copies will be made and stored in different locations.

Digital information can thus attain a form of eternity, but with the proviso that we constantly make it our concern. This radically changes the preservation process: previously, its essential phases were collection, storage, and then communication. Preservation now becomes a key function, since any archive in a state of artificial survival requires a controlled environment in terms of hygrometry, temperature and light, and more particularly its preservation needs to be administered. There are however positive elements which facilitate these cyclical migrations: storage costs are falling very quickly as technology advances, and transfer speeds from one medium to another are constantly accelerating.

In this area, we are only at the dawn of technological possibilities. Already, integrated systems and the automated preservation of digital information include optional monitoring of the deterioration of physical carriers (analysis of dropout rates in the magnetic coating) and automatic migration to a new carrier when certain thresholds are exceeded.

The other great advance brought about by computing to guarantee information is redundancy. We already touched on this it when mentioning backup copies, but computing has generalized these systems with the advent of error correcting codes, partial or total redundancy on hard disks and storage in distant physical sectors.

In the digital world, permanency will thus be maintained by speed: faster information turnround, faster information access speeds, faster analysis and correction of errors, faster recomposition of data integrity.

In the analogue age, preservation meant pausing, stopping time, suspending the life cycle of an object; in the digital age, this is reversed, for preservation means keeping the object moving.

It is not an established fact that the total cost of preserving digital of information is ultimately higher than in the analogue world. Technological competition exerts a strong downward influence on costs for many aspects of the process, but the cost structure has radically changed, imposing on most institutions profound changes in their working methods. Preservation of analogue materials could allow itself to be negligent, simply storing materials on shelves somewhere, whereas computing does not permit this attitude.

This new set of issues surrounding archive management has rapidly given birth to new software solutions which computer manufacturers and software publishers have called *Digital Asset Management.*

Allowing formats to evolve: migration, emulation, encapsulation

We have already seen that the greatest difficulties raised by the digital society undoubtedly rested on the multiplicity of logical formats, often proprietary – operating systems, data files, executable programs – and their very rapid rates of evolution.

However acute they may be, these difficulties have not remained unanswered, and today technical solutions make it possible to forecast how long-term archiving of digital formats may be achieved. The solutions that certain heritage institutions have started to put in place are of three types:

migration, emulation and encapsulation. They can either be carried out separately, or combined.

The solution most commonly used is migration, an operation during which the bits of a file or program are modified so that they can be read by a new operating system or a new version of an application. The old data is thus reprocessed, with the aim of making them compatible with a new environment.

This solution can be implemented in the case of a physical transfer from a medium with a different logical data format, or on the level of the operating system or file formats, or even on the level of the format of the data or application. Any migration can thus be performed by modifying part of the coding of the initial content so that they can be exploited by recent systems.

Given the technological instability which we have already emphasized, systematic and regular migration of all documents should be planned for at each change of version. As an example, the BNF undertook the migration of 15 million printed pages digitized in image mode for consultation purposes on Internet Gallica. This migration of approximately 1 TB of data, without changing the data format but merely modifying it according to the change of type of media, takes approximately 10 days, taking account of all the technical difficulties which may arise in the course of processing[1].

Another solution, emulation, consists in creating a program to simulate the computing environment necessary for the exploitation of the data on a current platform after their technical environment has become obsolete. Using a piece of software known as an emulator, the behaviour of a system which has disappeared is simulated. Whereas migration transforms the way the actual data are stored, emulation intervenes at the moment of their restitution. This technique has the advantage of keeping intact and unmodified the original bits of an archive or program, but it does not always make it possible to recreate optimal conditions for their use, as for example page formats may be disturbed or audio/video synchronization may be distorted. Although, theoretically, this seems a less destructive solution, it is the still the subject of important research projects, in particular those aiming to

1. C. Lupovici, *Les principes techniques et organisationnels de la préservation des documents.*

create the emulators of the future. In addition, it poses the problem of how various layers of emulators should be preserved in cascade.

Another promising solution, on which consortia of libraries and archiving centres throughout the world have been working, is encapsulation, also known as the reference model. This technique results from the OAIS (Open Archival Information Systems) model, which was developed for the preservation of space data by the Consultative Committee for the Space Data Systems and which has just been made the subject of an ISO standard[1].

The OAIS model, developed within the framework of a standardization group dedicated to exchanges of satellite data, gradually gained ground as a general model offering a conceptual framework making it possible to identify the elementary functional components of digital archives, and establishing concepts and terminology for the description of architectures and data models.

Encapsulation thus appears as a means of bringing together around the content itself information containing the instructions necessary to the decoding of its bits in the future by any system. It contains a series of layers of instructions, an external layer which will carry a text readable by all, which describes the contents of the encapsulated element and how to use it, and an internal layer which specifies the software, operating system and equipment characteristics to be reconstituted so that the object itself can be read. Encapsulation makes it possible to make both the content and the information attached to it autonomous, and seems set to become a relatively viable method for the long-term preservation of text files in particular. It remains of dubious interest for other types of documents, which suffer from technological dispersion and the excessive proliferation of new software types, compression systems and formats launched on to the market each year.

Ultimately, the solution will probably reside in the definition of a preservation format dictated by the archiving institutions, and to which they will have to commit in the long term. Homogeneity of formats will then allow automatic conversion to the following generation.

1. C. Huc, *La pérennité des documents électroniques* (the permanent safeguarding of electronic documents), http://www.archivesdefrance.culture.gouv.fr, 2001.

3.4. Documentation

We have already mentioned the documentary revolution brought about by digital technology. The corollary of the explosion in the number of documentary units generated by the information society is an unprecedented capacity to access them by their contents.

- From a documentary point of view, analogue was dead matter. It taught us nothing about itself, it reflected back no information. All information about documents had to be generated by various professionals, who produced catalographical, bibliographical, archival and documentary information which made it possible to structure collections passively and to create access keys for readers. Given the quantities to be processed, which were already increasing rapidly in the analogue age, it was often a Sisyphean task, always unfinished, leaving large grey areas even in the best managed catalogues, and ultimately giving the reader information which was altogether only modest.

With digital, not only can a document be summoned up instantaneously without a storekeeper having to go and find it on a shelf, but its contents are in addition directly comprehensible by a computer and the primary document itself, from the point of view of information retrieval, is nothing more than the addition of a string of access keys. This considerable development has made the fortune of search engines.

This aptitude of digital technology at present applies only to textual information. But already, there are some very advanced research programs in existence which make it possible to transcribe the spoken word into textual information, even in very noisy environments such as radio. There also exist a number of research programmes on the analysis of still and moving images, and it is to be hoped that one day great knowledge bases will approach human semantics and our capacity to recognize, designate, set in resonance and connote concepts.

- The problem is thus no longer that documents are silent, but that there is too much noise, an overflow of information. And each of us is aware that most requests addressed to search engines quickly become unexploitable. This phenomenon led the founders of Web to recommend a semantic

Web[1], in which data would be more structured, both data specific to the document and the data accompanying it (known as metadata). This necessary development will need greater involvement on the part of data producers in terms of the quality of data produced.

Without entering into detail, mention must be made of the intense activity on the world standardization front which reigns in this area of document structuring. Textual documents were very structured early on, using the SGML format; HTML format, the standard format on the Web, was not structured so strongly, which would have appeared too constraining. But the evolution of the Web brings us back to these problems with the current spread of the XML standard. In the field of moving images, the MPEG 7 and MPEG 21 standards are very much directed towards the standardization of various types of metadata (documentary, legal and technical information, data about uses, etc.) which accompany content data and will enrich it throughout its life.

3.5. Pilot projects for permanent preservation

From the beginning of the 1990s, by which time digital technology had penetrated all forms of expression and human creation, experimentation and research programmes generally carried by groupings of scientific research and archive institutions and libraries, began to explore how this new situation should be taken into account in archiving and preservation policies.

A pioneer in this field was the Internet Archive, an American foundation set up in 1996 as a private, not-for-profit company. This collects all Web pages freely available throughout the world, a collection of more than 10 billion pages, or five times the quantity of documents held by the Library of Congress. In October 2001, the foundation set up a ″time machine″ to give free access to its data through the Web[2].

The ground to be explored in this study – surveys of practices and needs, the state of the art in terms of tools, appraisals – of technical and procedural

1. T. Berners-Lee, *The Semantic Web*, presented at the 9th International World Wide Web Conference, Amsterdam 2000.
2. http://www.archive.org

solutions, development of models – was, and remains, vast. While lessons are already emerging from all this, there still remain many questions and uncertainties, with which local practices and experimentation and research programmes are confronted.

The initiatives are numerous and each one deserves a visit to learn all the lessons and to take maximum advantage of the knowledge acquired. We will simply draw attention to a number of pilot programmes because of their institutional and scientific positioning, in particular because they call on international expertise, and because their work enjoys widespread acclaim which enables their results to be shared.

• **Pandora (Preserving and Accessing Networked Documentary Resources of Australia)**[1] constitutes one of the first large-scale programmes to be supported by the Australian authorities. It aims to implement an archiving system for electronic publications and Australian Websites. Following a preparatory phase begun in 1996 by the National Library of Australia, a set of recommendations was published. These define in particular principles and methods of selection as well as the modes of organization and management of the archive. Publication is defined here in general terms: anything published on the Internet is regarded as a publication, internal documentation alone being expressly excluded. The sites selected for conservation purposes must relate to Australia or deal with subjects highly relevant to the country, and must have been conceived by an Australian.

Selection is carried out on the basis of contents and a priority ranking is attributed to publications which are authoritative and have a long-term value for research.

A second phase of work entitled ″PADI″ (Preserving Access to Digital Information) has broadened the field of partners, by associating with the project a broad panel of international experts and Australia's State libraries, as well as the National Film and Sound Archive, the National Museum of Australia, professional associations involved with ICTs, research centres, Australian universities, etc. Its objective is to design and put in place a

1. http://pandora.nla.gov.au/index.html

national model of shared archiving covering electronic data and the Australian Web.

To supplement these arrangements, the Public Record Office and the National Archives of Australia have broadened their task of managing electronic archives to include the Web sites of government authorities (public and Intranet sites) and have devised management principles which incorporate best practices. The National Archives draw attention to the fact that documents contained in Web sites are not always regarded as archives. These sites must therefore be managed rigorously. Also, in the Internet world, identifying and managing documentary material involves taking responsibilities, and is subjected to certain procedures.

• **The PRESERV programme[1]:** this programme is operated by the Research Libraries Group (RLG), Digital Library Federation (DLF) and the On-line Computer Library Center (OCLC). It follows various studies, surveys and experiments carried out in the universe of libraries confronted on the one hand with the rise in volumes of electronic publications, and on the other with the need to manage the digital collections for which they are responsible. It explores archiving conditions for electronic publications, digital archives and documents resulting from digitization policies, from both the methodological and practical points of view. It attempts in particular to define the attributes of the digital archive with regard to the purposes of research, to clarify the concept of collection of heterogeneous data, and questions of certification, structuring metadata, and the use of metadata within the framework of a long-term conservation policy.

• **NEDLIB (Networked European Deposit Library)[2]** is a programme which was supported by the European Commission from 1998 to 2001. It brings together eight large European libraries: the Royal Library of the Netherlands, the national libraries of France, Germany, Norway, Portugal and Switzerland, the Library of Florence and the University Library of Helsinki. The publishers Elsevier Science, Kluwer Academic and Springer Verlag are also involved. The programme aims to build a model of functional and technical specifications relating to electronic documents published on

1. http://www.oclc.org/research/preserv
2. http://www.kb.nl.coop/nedlib

physical media or on the Web. It has considered all functions, from collection via the development of a prototype, documentation and cataloguing, to preservation, including a complete review of migration and emulation techniques.

• **The CEDARS programme**[1] was developed by a consortium of British (Oxford, Leeds and Cambridge) and American (University of Michigan) university research libraries. Established in April 1998, it sets out to explore questions relating to the acquisition of electronic publications, their preservation, description and finally access to them; its objective is to make available to community libraries a guide detailing from both the methodological and the operational points of view best practices making it possible to integrate digital corpora of texts into pre-existing collections. In October 1999, the programme launched a new initiative aimed at looking further into the problems surrounding emulation techniques. This new programme, called CAMILEON (Creative Archiving At Michigan and Leeds Emulating the Old on the New), was due to run until September 2003. It explores how recourse to emulation as a preservation strategy can remedy the obsolescence of the systems used to produce digital data, and maintain possibilities of access to these contents in the configurations of their initial creation. It aims to develop and test emulation tools as well as to develop a set of recommendations concerning both migration and emulation.

• **The InterPARES programme (International Research on Permanent Authentic Records in Electronic Systems)**[2] is currently one of the most important research and development programmes devoted to the long-term preservation of digital data. It brings together teams of researchers, archive institutions, content producers and manufacturers from over twenty countries, and is controlled by the University of British Columbia, Canada. It has operated to date in two phases: InterPARES 1 began in 1999 and was completed in 2001 by the publication of theoretical models and a methodology for the area, followed by InterPARES 2 which is expected to run until the end of 2006. Its object is to identify and analyse the physical and intellectual parameters affecting the maintenance of digital data, and to identify new archiving and conservation methods and procedures resulting from this.

1. http://www.leeds.ac.uk/cedars
2. http://www.interpares.org

The programme aims to propose standards, organization principles and recommendations for actions to be implemented by governments, cultural and heritage institutions and manufacturers in support of the preservation of digital data.

Alongside these international co-operation programmes, many national scientific, cultural and heritage institutions have committed themselves to exploring and implementing ways and means of archiving the digital data for which they are responsible. Thus, for example, the Austrian National Library has joined forces with the Technical University of Vienna to design a system for the collection and preservation of digital data; the National Library of China is running a programme to set up a digital library; and in the Netherlands, the universities of Amsterdam, Tilburg and Twente are carrying out a research and development programme broadly based on the OAIS model in collaboration with IBM; and South Africa's principal cultural and scientific institutions are exploring the feasibility of setting up a digital archive of the South-African nation. In Taiwan, nine national institutions are associated in a digital public records project; in the United States, the Library of Congress is responsible for leading and running the 100-million-dollar National Digital Program. In Canada, the Canadian Heritage Information Network, supported by the federal government, makes available to cultural and educational institutions guides, advice and resources for the digitization and preservation of their collections. In Europe, various projects and concrete achievements are under development in particular in Norway, Finland, Sweden, Great Britain and France. These initiatives are largely supported by the European Community, whose Council published in May 2002 a resolution "on preserving digital content"[1].

As can be seen, societies and organizations are gradually becoming aware of the risk of amnesia which threatens them, albeit without having found a universal cure which would enable them to face up to the challenge. This is why it is more than ever necessary to continue research and development efforts, more than ever necessary to organize co-operation and to share knowledge and assets. The challenges of this new heritage force us to do so:

1. http://europa.eu.int

it is a new heritage spread on a worldwide scale, its preservation too requires worldwide initiatives.

While it is undeniable that progress has been made along the way of long-term preservation of digital documents, there still remain large grey areas relating to the conditions in which might make it possible to guarantee the permanency of contents and access to them.

Thus, the preservation of digital heritage currently remains unknown territory for most institutions. When responsibilities in this field are officially entrusted to them, they will have to adapt their organizational structures and redefine the tasks of their workforce.

Indeed, the immaterial nature of digital production will force them to conceive of new ways of preserving heritage, to redefine new procedures and methods of collecting, storing, indexing and consulting their collections which are more suited to cyber-contents generated in a context of the greatest technological instability.

It is essential therefore that the tools to preserve the memory of a document be developed at the same time as the medium is constituted.

Chapitre 4 Actions for decision-makers

The preservation of digital heritage depends on a certain number of decision-makers, of which the first are the governments of States. Indeed, market forces alone cannot guarantee the preservation and promotion of the world's digital heritage. From this point of view, the pre-eminence of the public authorities, in partnership with the private sector and civil society, is essential.

The political will of governments to preserve digital heritage must logically lead them to implement in each country a suitable legislative framework defining the roles and responsibilities of the various institutions in charge of heritage at national level. It must also lead them to release the financial resources essential to the preservation and consultation of this heritage.

Appropriate legislation

Governments and decision-makers must understand that the preservation of digital heritage is an urgent problem which cannot be solved overnight. The risk of losing essential materials in which invaluable resources have been invested is an extremely real one (according to the Library of Congress, the average lifespan of a page on Internet is two months, and almost half the total content posted on Internet disappears within one year).

Traditionally, the preservation of cultural heritage obeys legal frameworks and procedures which rest on well-defined criteria. National libraries usually gather and preserve publications by means of the legal deposit of their country's national production. In addition, a vast body of legislation on archives defines the missions of archive institutions, and the objects which they are to collect and preserve. Supplementing the task of public records offices, specialized archives and museums are responsible for collecting and preserving for heritage purposes audio, photographic, cinematographic or audiovisual materials.

The legislation involved can vary considerably from one country to another, according to how they allot these responsibilities to this or that institution. It also varies considerably with respect to the categories of documents

to be preserved. There is, however, broad agreement on two guiding principles: cultural heritage must be preserved, and citizens must be able to consult it. This is why archive institutions need suitable legislation regarding digital heritage, to enable them to define their tasks, and to select and collect the documents to be preserved.

Lawyers will point out that digital space, in particular that which exists in networks, is not naturally one of law. A study by the French Conseil d'Etat (Council of State) conducted in July 1998 on the Internet and digital networks (*Internet et les réseaux numériques)* pointed out that the law, which is applied on a territorial basis, is based on behaviours, and stable, homogeneous categories, all of which elements are absent from the Internet. According to some, it was this antagonism with respect to the law that favoured the emergence of the network in the first place, free from all constraints except those fixed by the community of researchers who originally created it.

Existing legislations must therefore be adapted or extended to the new and complex digital environment. Moreover, proposing new legislation is a long and difficult task, and there is no miracle formula.

The preservation of heritage is sometimes governed by specific laws on legal deposit, or in other cases by laws on author's rights or copyright, and in still other cases by laws instituting a national archive.

France, Finland and Sweden fall into the first category. In France, a new law on legal deposit was adopted in 1992: it applies to printed and non-printed documents, and in particular to radio and television programmes, but also to multimedia productions and databases published on computer media; a project to extend this law to French websites is currently in preparation. In Finland, the 1980 law on legal deposit covers printed materials, audio and video recordings and a separate law, enacted in 1984, governs the legal deposit of animated images, films and videotapes. In Sweden, the 1978 legal deposit law also creates national sound and animated image archives. A more recent instrument, dating from 1993, extends the obligation of legal deposit to ″portable″ electronic documents and other types of non-printed documents.

In Australia, the United Kingdom and the United States, the provisions governing legal deposit form part of copyright legislation. In Australia, texts

relating to the legal deposit of ″library material″ form part of the 1968 Copyright Act, which confers on the National Library the status of national repository of such documents. But ″library material″ in that law refers only to publications printed on paper. The National Library and film archives have presented a joint proposal to broaden the scope of the law, while recommending that its provisions relating to legal deposit be maintained in the copyright law, rather than being incorporated into the law relating to the National Library or set up as distinct legislative instruments. In Australia, legal deposit provisions have always gone hand in hand with copyright questions, and therefore it was considered simpler to maintain this bond and to legislate on the first subject when legislation relating to the second was re-examined. The statutes governing legal deposit in the United Kingdom form part of the 1911 Copyright Act. The British Library has recommended that in the case of non-printed publications, new legislation should be completely separate from copyright law, in order to make a clear distinction between these two areas. In the United States, legal deposit obligations appear in the Copyright Act of 1976 which confers on the U.S. Copyright Office, under the Library of Congress, the power to enact rules requiring the deposit of the ″best edition″ of works in any medium. In Canada and Germany, it is laws on their national libraries that render legal deposit obligatory. In Spain, the 1971 law on the Hispanic Bibliographical Institute covers legal deposit.

In general, national libraries tackle the problem of the digital environment from the point of view of legal deposit legislation. The off-line deposit of digital products, for example CD-ROMs, is already a legal obligation in several countries. On-line electronic publications are regarded as the extension of a long tradition of publications in printed media, which national libraries have collected always and preserved. Legislation on legal deposit is thus a key element of national preservation policy. As in the case of traditional media, citizens must be able to access the digital materials deposited without endangering the legitimate exploitation of heritage in the digital environment. It is therefore necessary to propose legislation which is the broadest possible, and ensure that it applies both to publications produced only in electronic form and to those produced ″in parallel″. If there is the slightest doubt about the relevance of an object, it is better not to establish a distinction between the on-line and autonomous forms of electronic publications, and to include both whenever there is a risk of rapid evolution

towards electronic publishing on line. It will then be left to the legal deposit authority to determine which elements to retain for the national collection.

This legislation must be developed not only for materials and publications, but also for research data, for example, perhaps by making deposit a condition of research subsidies.

Application of legislation

Once the principles have been posited by legislation, the law must be applied. Countries which have begun to do this have adopted a progressive approach, but it is difficult to provide precise instructions given the importance in this respect of practices already followed for the legal deposit of printed publications, the framework proposed for new legislation applicable to electronic heritage, the types of digital products to be included within the scope of application of the legislation, and the numbers of electronic documents that may be deposited each year.

Co-operation among actors of the digital universe

The establishment of real partnership within one country is essential to the preservation of digital heritage: creators of digital documents and the professionals working in communication and information technologies must be associated with the preservation process because their co-operation could eliminate part of the burden which weighs on heritage institutions. Creators will have to be encouraged to use open standards and to document their files appropriately. It will also be necessary to convince ICT professionals of the advantages of unprotected source code software, and of the need to publish detailed and complete documentation on their products so that they can be used later in a preservation context.

Indeed, the requirements of preservation systems for digital heritage lead creators to become aware early of the fact that the choices they make at the time of creating a document determine the possibilities of subsequent archiving it, and can contribute to their future maintenance: the use of open standards and formats, suitable descriptions and documentation, and the use of permanent names for on-line resources facilitate long-term preservation and contribute to reducing their cost.

To achieve this, communications efforts by political decision makers towards their national public opinion are essential. It is important to present a clear argumentation of what is at stake in this mission. It is entirely possible to generate real enthusiasm by emphasizing the importance of national digital heritage and the possibility of preserving its essence in the future. This work of explanation and mobilization must be undertaken at the same time as the feasibility studies, so that institutions are fully motivated around this priority activity, and private partners are ready to play their part.

Within the framework of this co-operation between the actors of the digital world, problems of authors' rights and copyright should find a positive outcome through agreements between heritage institutions and publishers, with the aim of reconciling the interests of the two parties, authorizing copies for the sole purposes of preservation, while limiting access.

Authors' rights and copyright

Authors' rights and the copyright are indeed an important issue. Current legislation imposes such strict limitations on copying that in certain extreme cases libraries cannot even preserve the electronic reviews to which they subscribe without encroaching on owners' and creators' rights. Certainly, publishers recognize that authors' rights and copyright may represent an obstacle to long-term preservation, but they are also very wary of any arrangement which might disturb their commercial interests by facilitating access via the networks to materials deposited with them.

The World Intellectual Property Organization (WIPO) keeps watch over the adaptation of existing international conventions to take account of the problems raised by the development of digital networks. In particular, it was under the aegis of WIPO that two treaties in were concluded December 1996, the WIPO Copyright Treaty and the WIPO Performances and Phonograms Treaty. Moreover, a new treaty is currently under negotiation on performers' rights. These new conventions aim in particular to broaden the field of protection of artistic works in order to include among them new media and modes of transmission, especially digital ones. However, until now they have not solved the difficult questions concerning exceptions to authors' rights on networks, and of which law is applicable.

In addition, a very important agreement has been concluded within the framework of the World Trade Organization (WTO) on those aspects of intellectual property which concern trade.

Given the desire of authors for ever greater protection of their works against piracy, in particular by technical means, many intellectuals fear that economic logic may restrict the circulation of ideas and the access of all to culture. This concern is shared by certain international organizations such as UNESCO. It is also taken seriously by governments which have a very strong tradition of the exception to the exclusive right of authors which allows freedom of movement of works for the purpose of education and research, in particular in the Anglo-Saxon countries (the so-called ″fair use″ exception) and in the countries of northern Europe, as the study of the French Conseil d'Etat has observed.

However, documents officially and publicly expressing the intentions of governments as to the adaptation of their national literary and artistic property regime to the specifics of the digital networks remain rare.

Managing copyright issues is becoming extremely complex: the law, and the agreements between preservation institutions and publishers, cannot themselves cover all aspects of the question. For example, when a digital product requires use of software which is the exclusive property of a third party, the creator of the content is generally not the copyright holder. It should be noted that software publishers have scarcely associated themselves until now with preservation efforts, and software is not generally covered by legal deposit legislation.

Both creator and custodian have a responsibility as regards the safe keeping of documents. Insofar as creators are not always aware of all the risks, heritage institutions must enlist their co-operation to make possible the permanent preservation of the document, from the moment when the document was created.

Organization

To build a solid infrastructure able to ensure that a system of distributed digital archives can function presupposes the existence of reliable

organizations able to keep materials "alive" over the long term. Libraries and national archives, as well as a number of specialized research institutes and data archives, currently assume this role. But there also exist many other institutions which could be called upon to intervene in the preservation of certain types of documents (photographs, audio recordings, works of art, television programmes, etc.) in digital form, or in the preservation of documents for a particular community (local or regional institutions, research institutes in particular disciplines, etc.)

If organizations dealing with digital archives must be reliable, this is first and foremost in the interests of the cultural and democratic benefit of national communities, and beyond that, of humanity itself. By taking initiatives to test preservation models, national institutions can also help other organizations to understand the requirements of an operational preservation system and to set up systems in their own areas.

Restoring the lost unity of the Web archive

The Web today is world-wide, it is one. Now, most preservation projects select from this total space thematic sub-sections, or reconstitute territories on the scale of their country and their national institutions of memory. This prospect seems normal and legitimate, but it brings with it the risk of a loss of substance through lack of a sense of totality: i.e. all those links that lead across borders, and which will thus be broken.

One may dream of a central Internet archive at the worldwide level. Just such a temptation underlies the Internet Archive project[1]. But it seems inconceivable that nations will give up any sovereignty where their heritage is concerned. On reflection, moreover, this model is far removed from the Internet, the basic concept of which is the federal organization of networks without any pyramidal or centralized concentration.

The future of the archive lies more in the concerted organization of international archiving standards, which will make it possible to build hyper-archives by progressive federating and networking of heritage elements, and the world-wide interconnection of institutions of memory.

1. http://www.archive.org

Clearly, such a hyper-archive, a historical version of the Internet bringing to the Web as a whole a third dimension, should not be merged with the Web itself, so as to preserve the rights of site publishers, but rather should follow its trail. It will not be possible to achieve this objective without international co-operation and strong motivation on the part of all concerned.

Financing

To be able to ensure adequate preservation of digital heritage and general access to it, constant efforts are needed on the part of governments, authors, publishers and the other industries involved, and the institutions in charge of the preservation of documentary heritage. Also, institutions in charge of the long-term preservation of digital material must have guaranteed financial permanency.

Heritage institutions on which the right is conferred to collect digital documents must be assured of sufficient funding to acquire the equipment and premises necessary for the storage of these data, and to recruit sufficiently qualified personnel to deal with them and to make them accessible. However, in the majority of cases, if not all, the institutions themselves are not financially able to maintain the digital documentation they collect or to provide consultation services without additional funding specifically assigned for that purpose.

Indeed, it should be noted that the complexities of storing digital documents does not cancel out the storage costs, but rather changes the nature of these costs, and that heritage institutions will have to manage two types of documents: those on traditional media, and immaterial documents. The technology necessary for the preservation of digital documents requires investments in research and development which are certainly significant in absolute terms, but which remain negligible when compared with the resources invested in the creation of the materials themselves, or with the cost which the loss of these materials would represent for society if suitable preservation systems were not developed.

Research and training

Public financing must be also envisaged in resolute support of research into technologies and promising models, in order to develop as quickly as possible fully operational preservation systems for digital heritage.

Lastly, financing must be found for the creation of vast training schemes. These are necessary to disseminate widely the skills and experience acquired until now by the managers and other personnel of heritage institutions. Indeed, the preservation of digital heritage requires new forms of organization, new working methods and new ways of thinking. These programmes must not only deal with technical aspects, but also offer training which will enable staff to adapt to a changing environment and the changes of direction that it imposes.

The preservation of digital heritage may be considered to be a factor for creative development, as the preservation and exploitation of traditional cultural heritage has been in the past, and will continue to be increasingly in the future.

International co-operation

Considering the existing digital divide, it is necessary to reinforce international co-operation and solidarity to allow all countries, in particular developing countries and countries in transition, to ensure the preservation of their digital heritage and continuous access to it, by the pooling of experience, the dissemination of results and best practices and the concluding of twinning agreements.

Thus, the coordination of institutions must be carried out in parallel with active collaboration within the international groups of actors already involved in the preservation of digital data. This collaboration will give a clear overview of work already completed and likely to enable any State to benefit from significant experience.

For example, in April 2001, at the time of the Lund Conference in Sweden, the European Union created the National Representatives Group (NRG)

on the digitization of scientific and cultural heritage. The NRG's role is to make it possible to coordinate the digitization policies of the various Member States. The national representatives are administrative personnel or politicians appointed by their governments. The NRG established the ″Lund principles″, a declaration of intent by States to make their heritage digitization policies move in a common direction. Among these principles are the constitution of a common coordination structure for digitization in the various cultural sectors (museums, libraries, archives, archeology, etc.), a research policy for the promotion of heritage, the evaluation of digitization practices and the adoption of common standards. The NRG meetings provide opportunities to discuss the practical application of the Lund principles and to define priorities.

The cultural heritage of all regions and countries should be preserved and made accessible in order to achieve, over time, fair and balanced representation of all peoples, nations and cultures. It is also appropriate to encourage minority groups within nations to see their digital heritage as an element of the overall digital heritage of the world.

For this reason, as in the strictly economic domain, it is essential to reduce the inequalities between developed and developing countries, while enabling the latter to be able to persevere with their actions in support of heritage. Thus, the president of Senegal, Abdoulaye Wade, proposed the concept of ″digital solidarity as a strategy to bridge the digital divide″. Indeed, the preservation of the world's digital heritage is not only a question of cultural heritage. In the longer term, it will affect the very nature of the knowledge societies which are today coming into being.

Be that as it may, the field is so new, and experience for the moment is so limited, that extraordinary efforts will be needed to build the necessary infrastructure. Sufficient resources and support at the decision-making level are essential to enable future generations to continue to have access to the abundant digital resources of the world, especially those created during the last ten years.

Chapitre 5 Recommendations

The preservation of the digital heritage of humanity is a new task incumbent upon States and actors of the information society.

This is a continuous activity which requires commitment and participation not only on the part of heritage institutions but also on the part of public authorities, other decision makers, producers and users of information, software manufacturers, and international professional associations and organizations.

These solutions imply the following

1. They must be based on the experience gained in the preservation of other forms of world heritage, material or immaterial, such as manuscripts, printed or audiovisual documents, so as to create an institutional, legal and technical framework likely to optimize the preservation and accessibility of digital heritage

2. Clear objectives for preservation must be laid down, both qualitative and quantitative, to make it possible to determine which documents or families of documents must be preserved, and whether this preservation must be exhaustive and systematic or merely periodical, by means of sampling

3. Permanent access to digital heritage documents must be guaranteed, in particular to those which belong to the public domain, to provide a balanced and equitable image of all peoples, nations, cultures and languages over time

4. Teaching and training programmes should be encouraged, as should agreements on sharing resource and the dissemination of research results and best practices in order to democratize access to digital preservation techniques

5. Research into technologies and promising models should be resolutely supported, in order to develop fully operational systems for the preservation of digital heritage as soon as possible

6. International co-operation and solidarity should be strengthened in order to enable all countries, in particular developing countries and countries

in transition, to create, spread and preserve their digital heritage, and allow permanent access to it

7. The pre-eminence of the public authorities, in partnership with the private sector and civil society, should be reaffirmed over market forces which cannot guarantee alone the preservation and promotion of the world's digital heritage.

Composé par JOUVE, 11, bd de Sébastopol, 75001 PARIS
N° 328195Z. Dépôt légal : Novembre 2003