



COMMISSION OF THE EUROPEAN COMMUNITIES

Brussels, 30.9.2005  
SEC(2005) 1194

**COMMISSION STAFF WORKING DOCUMENT**

***Annex to the :***

**COMMUNICATION FROM THE COMMISSION**

**”i2010: DIGITAL LIBRARIES”**

**{COM(2005) 465 final}**

## TABLE OF CONTENTS

1.	Aim and scope of this paper .....	3
2.	Digital libraries.....	3
2.1.	Digitisation.....	3
2.1.1.	Why is digitisation important? .....	3
2.1.2.	The collections of libraries and archives.....	5
2.1.3.	Challenges for digitisation .....	6
2.1.4.	Actions at European level .....	9
2.2.	Making digital collections available .....	11
2.2.1.	Availability of digitised collections .....	11
2.2.2.	Digital libraries based on ‘born digital’ material .....	12
3.	A future for the past: preservation of analogue material through digitisation .....	13
3.1.	Preservation through digitisation .....	13
3.2.	Challenges for preservation through digitisation .....	14
3.3.	Actions at European level .....	15
4.	A future for the present: preservation of digital material.....	15
4.1.	The issue of digital preservation: what is the problem?.....	15
4.2.	What has been done so far to tackle the issue? .....	16
4.2.1.	A worldwide reply.....	16
4.2.2.	Actions in the Member States .....	18
4.3.	Challenges for the preservation of digital material .....	18
4.4.	Actions at European level .....	20

## **1. AIM AND SCOPE OF THIS PAPER**

The aim of this paper is to support the Commission Communication ‘i2010: digital libraries’ and expand on it with further background information.

The paper deals with questions related to accessibility and preservation of information in the digital environment. The institutions most concerned by these questions are libraries and archives, which traditionally have been entrusted with the task of making information available and keeping it for later generations. The paper deals with all types of content - including books, newspapers, audiovisual material, maps, pictures and music - and with all public institutions involved in collecting and preserving content: national and local libraries, research libraries, audiovisual archives, historical archives, etc. Examples drawn from the different sectors are used throughout the paper. Many of the issues raised are also relevant for museums active in making their collections more accessible by digitising them (for example digital images of paintings or historical objects) and making them available online.

Digitisation and preservation of information are strongly interrelated and therefore have to be considered together. To ensure durable availability of the digitised content, proper attention for its preservation is required. Digitisation and preservation sometimes overlap, as when digitisation is itself an integral part of the preservation process. In that case, digitisation and preservation coincide. Furthermore, the same institutions – libraries and archives – are normally responsible for digitising and preserving information and making it accessible for all.

## **2. DIGITAL LIBRARIES**

A digital library presupposes the existence of digital material and the possibility to make it available through electronic networks. Most material from the past exists only in ‘analogue’ format. A first step to making the material available online is digitisation.

### **2.1. Digitisation**

#### *2.1.1. Why is digitisation important?*

##### *The social and cultural dimension*

European libraries, archives and museums contain a wealth of information, representing the richness of Europe’s history, its cultural diversity and its scientific achievements. The degree of access to this information determines how far people can experience their cultural heritage and benefit from it in their work or studies. By digitising their collections and making them available online, libraries, archives and museums can reach out to the citizens and make it easier for them to access material from the past.

The digitised material can be a key asset for educational purposes and enrich Europe’s cultural life. The online availability of works from different cultural backgrounds and in different languages will make it easier for citizens to appreciate their own culture heritage, as well as the heritage of other European countries.

Digital libraries can contribute significantly to Community policies, in particular in the areas of the information society, multilingualism and culture.

The recent initiative of Google to digitise large collections from several major libraries has raised a series of issues related to the presence of Europe's cultural heritage on the Internet.

*On 14 December 2004, Google announced that it is working together with five major libraries to turn their collections into searchable digital content. It will scan millions of books of the New York public library, and the libraries of Harvard, Stanford, Michigan, and Oxford University. For texts in the public domain, it will make the full text available as part of search results. For texts under copyright, it will work with publishers and authors to determine how much of the text will be shown.*

Most of the works to be digitised through the Google initiative will be books in English – although the collections of the libraries also contain books in other languages.

#### *The economic dimension*

Libraries, archives and museums are major sectors of activity in terms of employment and investments. The figures below illustrate the size and outreach of the library sector.

*Core figures for libraries in the EU 25 (based on 2001 data)<sup>1</sup>:*

*- 336,673 full time equivalent staff*

*- more than 138 million registered users (around 30% of the population)*

*- more than 14 billion € spending on libraries*

Their impact on the economy at large is substantial. Information is the fuel of our economy, with content industries totalling some 5% of Europe's GDP<sup>2</sup>, and with ever more organisations depending on the right information flows to take informed decisions.

*A recent study found that the British Library generates some 534 million € worth of value per annum, both in direct value to the library's users (87M€) and the indirect value to society (447M€). This is 4.4 times the annual government funding of 122 million €.<sup>3</sup>*

Digitisation of the content resources held by libraries, archives and museums and making them available electronically could considerably increase their economic impact. This is for example the case for scientific libraries. Access to information is essential for the progress of science, since all inventions depend in one way or another on the achievements of the past. Broad access to scientific material is a key factor to science and innovation and indirectly for economic growth and employment.

---

<sup>1</sup> *International library statistics: trends and Commentary based on the Libecon data*, by D. Fuegi and M. Jennings, June 2004. <http://www.libecon.org/pdf/InternationalLibraryStatistic.pdf>

<sup>2</sup> See for example *The contribution of copyrights and related rights to the European economy*, by the Turku school of Economics and Business Administration for the European Commission, October 2003. The report arrives at a figure of 5.3% of GDP for the EU 15.

<sup>3</sup> Pung, C., Clarke, A., and Patten, L. (2004) Measuring the economic impact of the British Library, *New review of academic librarianship*, 10(1), 79-102.

Once digitised, Europe's cultural and scientific heritage can be a rich source of raw material for added-value services and products in sectors such as tourism, and the cultural industries. Following digitisation of its archives the reuse of audiovisual material of the RAI increased 85% over three years<sup>4</sup>.

At the same time, digitisation efforts will have a considerable spin-off for other industries. The Google initiative shows that public domain material can be an important driver of traffic on the Internet. In addition, increased efforts to make Europe's cultural collections available online will give a boost to firms developing digitisation technologies.

If properly maintained, digitised material can be used over and over again, by several users at the same time. An investment in digitisation is likely to pay off over time, not only in cultural terms, but also in economic terms.

### 2.1.2. *The collections of libraries and archives*

The collections of Europe's libraries are impressive in variety and volume.

*The total number of books and bound periodicals (volumes), held by European libraries (EU 25, all libraries, including national and local libraries, research and academic libraries) was 2,533,893,879 in 2001. This total amount is relatively stable over time.<sup>5</sup> The 'Bibliothèque Nationale de France' holds some 13 million books and periodicals, 200,000 of which are rare books. The total number of items held by the British Library is 150 million items, with 3 million new ones added to its collections every year, together taking up 625 kilometres of shelf space.*

Digitising all this material seems not realistic in the light of the costs of the digitisation process and subsequent preservation. Choices will have to be made as to the material to be digitised, notably on the basis of user demand and preservation criteria.

These figures refer only to books hosted by libraries. Many libraries also hold important image and map collections, and archives equally contain vast amounts of content.

*For example in Portugal, the material of the historical archive "Torre do Tombo" takes up 70 kilometres of shelf space. In Spain, material of the General Archive of the Administration takes up 200 kilometres. The latter holds over 20 million maps, mostly from the Ministry of Public Works, over 14 million photographs, etc. The material is unique and frequently consulted. Only a very small part has today been digitised (350,000 images).*

The same goes for audiovisual archives. Assets are huge, and mostly in analogue format.

*A survey of ten major broadcasting archives found 1 million hours of film, 1.6 million hours of video recordings and 2 million hours of audio recordings. Total European holdings of broadcast material are probably 50 times larger. Most of the material is original and analogue.<sup>6</sup>*

---

<sup>4</sup> UER report on archives, 2003

<sup>5</sup> *International library statistics: trends and Commentary based on the Libecon data*, by D. Fuegi and M. Jennings, June 2004. <http://www.libecon.org/pdf/InternationalLibraryStatistic.pdf>

<sup>6</sup> Survey by the IST Presto project which finished in October 2002. <http://presto.joanneum.ac.at/index.asp>

Many libraries and archives have started digitising their assets, but only a small part of European collections has been digitised. Currently, digitised material can be estimated at between 1 and 2% of total unique collections, on average across the EU, although individual institutions may reach far higher percentages. Digitisation activities exist in all the Member States, but efforts are fragmented and progress has been relatively slow.

### 2.1.3. Challenges for digitisation

**Financial challenges:** Digitisation is labour-intensive, and therefore costly. It takes a considerable upfront investment, which in most cases by far exceeds the means of the institutions holding the information. It requires substantial investments in equipment and staff dealing with the digitisation process but the actual production of a digital version only represents part of the cost of digitisation.

*Costs involved in the digitisation process concern in particular: selection of the material to be digitised, clearing of rights issues, actual digitisation costs, creation of metadata, making the material available to users, quality control and assurance, storing and preserving the digital material. One analysis indicates that the technical digitisation for mixed collections accounts for under one-third of the costs (32% overall), with metadata accounting for 29%. The remaining costs are in other activities such as administration and quality control.<sup>7</sup>*

Calculations of the costs of digitisation for books vary widely.

*A detailed study for the national library of New Zealand arrived at a cost of NZ \$ 0.48 (around 0,27 €) per page for the actual digitisation, in particular costs for the equipment and staff operating the equipment.<sup>8</sup> Another recent source prices image digitisation (monochrome) in a range of 0.12-0.30 €, whereas greyscale would cost between 0.26 and 1.00 €. Full text processing and treatment by optical character recognition (OCR) gave a price of 0.08 € per page.<sup>9</sup> Figures cited for the Google operation are in the order of magnitude of 150-200 million dollars (122-164 million €) for the digitisation of 15 million works. At an average of 200 pages per book, this would come down to a cost of 0.05-0.075 US dollar (0.04-0.06 €) per page or 10-15 US dollar (8-12€) per book.*

The above figures are based on a highly standardised digitisation process. Obviously, also the required quality of the digital copy plays a role, and costs can be far higher for special items such as old or fragile books, such as the famous 15<sup>th</sup> century Gutenberg bible.

*The British Library first made its Gutenberg Bibles available on the web in November 2000. The pages received one million 'hits' in the first six months, showing the popularity and huge interest in the work and the potential of new technologies to reach out to a wider audience.*

---

<sup>7</sup> *The costs of digital imaging projects*, S. Puglia, RLG Diginews, Oct 1999. Although relatively 'old', the analysis is interesting, given that since then the prices of the equipment for digitisation have gone down.

<sup>8</sup> *Digital library development review for the National library of New Zealand*, S. Ross, 2003.

<sup>9</sup> *Retrospektive Digitalisierung von Bibliotheksbeständen*, Thaller ed., University of Köln, January 2005.

The digitisation of audiovisual material has its own cost-dynamic, depending on the type of material and its state of conservation. Digitisation of one hour of media material can cost between 100 € (audio) and 2000 € (film).<sup>10</sup>

While the costs of digitisation are high, the incentives for the organisations carrying out the work may be limited to considerations such as increasing their visibility and saving physical storage space. Most benefits will in fact affect the economy and society at large. This may well discourage organisations holding the materials to digitise, in particular if they have to find the resources within their limited budgets.

*Although as a matter of policy, digitisation and its financing continues to be the responsibility of the individual cultural institutions, the Danish government allocated, in 2002, 13 million DKK (1,75 M€) from the profits of UMTS licenses to digitisation projects. From the same funds 53 million DKK (7,10 M€) were made available to the major cultural institutions for research (22 million) and digitisation (31 million) for the period 2003-2005.<sup>11</sup>*

In some cases, national digitisation efforts are supported by EU structural funds. This is for example the case for Lithuania.

Given the costs involved, choices are necessary as to what can be digitised and when. A careful assessment will have to be made at every stage as to the necessary investment in relation to the expected social and economic benefits.

**Organisational challenges:** *Digitise once, distribute widely* seems to be a rational strategy that can benefit all the organisations involved. In order to spend a limited budget in the best possible way, duplication of effort (for example digitising the same works or collections several times) must be avoided. The ‘digitise once, distribute widely’ strategy can only be applied if the economic models of the institutions involved allow for it.

A ‘digitise once, distribute widely’ strategy can benefit from a sustained co-ordination effort, based on the dissemination of good practices and an exchange of information on digitisation work undertaken. Cross-border collaboration can also enhance European added-value, where the final result is more than the sum of the parts. This is for example the case when digitising collections related to authors who have lived in more than one European country.

Digital libraries are not the sole business of the public sector. On the contrary, publishers, technology providers, learning and cultural industries etc. have a crucial role to play. Publishers could, for example, facilitate the development of digital libraries with copyrighted material.

Public-private partnerships to make the information available or sponsoring by private companies to digitise could contribute to ease the financial challenges involved in digitisation. The collaboration between public and private entities can take different forms, varying from sponsoring to partnerships in which the private company uses the digital material in its business. Overall this type of collaboration is not yet well developed. Within the

---

<sup>10</sup> PRESTO project cited above.

<sup>11</sup> Minerva project, *Progress report of the National Representatives Group: coordination mechanisms for digitisation policies and programmes 2003, Denmark.*  
<http://www.minervaeurope.org/publications/globalreport/globalrep2003.htm>

arrangements, it is important to make sure that the information remains accessible for all and avoid exclusive arrangements based on material held by public institutions.

*Private-public partnerships can play an important role in digitising and making accessible historical collections. A good early example is the case of the 'Archivo General de Indias' which holds the historical collections on the discovery, exploration and administration of the Americas by the Spaniards (9 Km of shelves). At the occasion of the 500 Anniversary of the Discovery of America, 8 million items, or 10% of the total collections were digitised during the period 1986-1992, with the sponsorship of El Corte Inglés and IBM Spain. This included a huge effort of description and cataloguing of the selected documents. Since then, the 'Archivo General de Indias' continued to digitise its collection, but at a much lower pace. Today 12% of their assets are available in digital format.*

Change may be called for not only in the collaboration between, but also within the organisations. Investments in digitisation have to take into account the absorption capacity of the institutions involved. In addition, in many cases it will be necessary to upgrade the skills of staff working in libraries, archives and museums in order to make the most of the new technologies and online accessibility. New types of skills are necessary to deal with the technological tools, next to the extensive expertise that already exists within the institutions. An investment in digitisation could lose much of its value if the skills of the relevant staff to handle the material are neglected. Therefore adequate training of the existing staff and targeted attention for the necessary skills while recruiting should be part of a successful digitisation and subsequent online accessibility strategy.

**Technical challenges:** While substantial progress has recently been made in the area of digitisation, especially in regard to optical character recognition software for English-language texts and printed text fonts, further improvements in the efficiency of support tools would significantly contribute to making Europe's cultural and scientific collections available for all. In particular, there is a need to adapt optical character recognition to languages other than English, backed by appropriate font recognition, integrated spell checking and correction and document segmentation techniques.

*Linguistic issues are a key factor. One of the reasons why Google is able to ensure a huge throughput of English-language books at minimal cost is that English-language optical character recognition (OCR) systems have not only been tested and run on most print fonts over the past 10 to 15 years, but have been enhanced by automatic or semi-automatic spelling correction algorithms. These are based on linguistic analysis and segmentation and on huge dictionaries of normal words as well as geographical and personal names, names of organisations, etc.*

In addition, improved high-speed automatic book (including page-turning) and document feeding equipment linked to appropriate scanners for various formats is required to handle the enormous volumes of books and archival documents to be brought on line. By reducing the human effort currently required in digitisation work, this will make the process more cost-efficient, affordable and increase the quality and reliability of the digital versions.

Other technical aspects may also play a role, since the mere availability of digital copies does not automatically translate into a European digital library. These challenges range from the interoperability between libraries and the formal description of the collections (metadata and identification systems), to tailoring search technologies to the needs of libraries and archives.



Over the medium term, progress in the areas of automatic handwriting recognition, image identification and tools for automatically summarising and/or assigning content-based metadata or keywords to documents will improve their classification, thus improving the pertinence of searching for different purposes. Finally, progress in identifying materials warranting urgent digitisation on the basis of realistic user requirements will also contribute to the efficiency of digitisation work for the short to medium term.

**Legal challenges:** Digitisation presupposes making a copy, which can be problematic in view of intellectual property rights (IPR). Directive 2001/29/EC of the European Parliament and Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society<sup>12</sup> foresees in its article 5.2 the possibility for exceptions or limitations to the reproduction right for specific acts of reproduction made by publicly accessible libraries, educational establishments or museums, or archives, which are not for direct or indirect economic or commercial advantage. The exception is however not mandatory and has led to different implementations in the Member States.

In many cases, the costs of establishing the IPR-status of a work will be higher than the digitisation of the work itself. This is true for literary works, and even more so for audiovisual material, where there may be tens of rightholders. The challenge of successfully dealing with IPR-issues is a key factor for the speed of digitisation. Solutions have to be found that respect the legitimate interest of creators, while enabling full use of the potential of the new technologies.

#### 2.1.4. *Actions at European level*

The need to digitise Europe's cultural and scientific heritage was already recognised by the Commission in the eEurope action plan. This gave rise, in 2001, to the Lund Principles and the corresponding Lund action plan.<sup>13</sup>

*The Action Plan was drawn up following a conference on the relevant issues held in Lund, Sweden, with experts from all Member States on 4 April 2001. It addresses a number of technical aspects relevant for digitisation, amongst other things ways to improve national digitisation policies and cross-border co-operation, discovery services for digitised resources, and the exchange of good practices. Over the last few years, the Action Plan has served as a framework of reference for digitisation in several Member States, including the new Member States. The Action Plan does not contain any concrete targets committing the Member States to achieve quantitative objectives, which makes it difficult to measure real progress in the Member-States over time.*

The Member States, supported by the Commission, have given an active follow-up to this work. A National Representatives Group was set up to act as a steering group for the co-ordination of digitisation policies. The activities are supported through the MINERVA project funded under the Information Society Technologies programme.

---

<sup>12</sup> OJ L 167, 22.6.2001, p. 10.

<sup>13</sup> A first exchange of views at ministerial level on the Lund principles took place in the informal cultural Council in Bruges in December 2001. You will find the text of the Lund Action Plan from the following page: <http://www.cordis.lu/ist/ka3/digicult/europe-overview.htm>

*MINERVA is a network of Member States' Ministries which aims at the harmonisation of activities carried out in digitisation of cultural and scientific content. The project also works with international organisations, associations, networks, international and national projects involved in this sector. It was extended to the new Member States through MINERVA PLUS project.*

EU research funding has resulted in a portfolio of projects aiming at making Europe's cultural heritage more accessible through the use of new technologies. In the period 2002-2005 some 50 million € of co-funding went to projects relevant for digitisation and accessing the resources in archives and libraries.<sup>14</sup>

An example of a project that has helped to make progress in the area of digitisation techniques is the MEMORIAL project that ran between 2002 and 2004.

*MEMORIAL, coordinated by Gfai, Berlin, and bringing together museums and technical partners, has developed tools for digitising very large volumes of prisoner transfer records from the Nazi concentration camps. This was achieved by improving document segmentation techniques, enhancing the quality of the text by using multispectral analysis and adapting optical character recognition software - including appropriate language support - to deal with typewritten documents, usually in the form of carbon copies. Success in the area covered paves the way for future applications in other areas. Co-funding for the project was 1.5 M€.*

Projects supporting co-operation between Europe's national libraries have developed from simple exchange of cataloguing records to 'The European Library project' (TEL).

*TEL has now been launched as an operational service providing a single gateway to the collections of Europe's national libraries. The site covers 150 collections, representing about 11 million records and digitised objects and items.*

In the audiovisual sector, an initiative is underway to improve the interoperability between film archives.

*The Commission has mandated CEN to adopt a European standard on cataloguing and indexing practices of cinematographic works and on the interoperability of film databases. The standard will address in particular the harmonisation of terminology, and a set of common rules on cataloguing and indexation.*

Digitisation of cultural heritage is part of the cooperation projects co-financed under the "Culture 2000" programme in an increasing number of cases. For instance, this programme is co-financing SHPAENA; this cooperation project between five European News Agencies aims at digitising historical photographic archives.

At the level of the Community Institutions, the EU Bookshop aims at becoming the virtual library for all the publications of the European Union since 1952. The EU Bookshop opened at the beginning of 2005.

---

<sup>14</sup> In most of these projects digitisation per se was not the only issue. Several projects were concerned with digital representations of archaeological sites, objects, etc, rather than dealing with text or flat images.

## 2.2. Making digital collections available

Libraries and archives are using the new information technologies to provide better levels of service.<sup>15</sup> Citizens for example access information on local government and events through local library websites. As a result, they expect to be able to use the Internet to access also the digital resources owned by libraries and archives. This would contribute to making the cultural heritage of one country accessible for citizens from other countries and could help to reach out to peripheral regions.

However, the traditional model of library services based on lending of the physical items they own is not easily translatable to the digital environment. There is no guarantee that the items will not be copied or further distributed, which might prejudice the interests of the rightholder. There is at present no digital equivalent of taking a book home from the library. Paradoxically, at a time of easy, rapid transfer of digital information from source to office or home, it is citizens who have to go to libraries to consult digital resources on-site.

### 2.2.1. Availability of digitised collections

Digitisation has the advantage of allowing wide and cheap distribution. There are, however, costs involved in making digitised collections available online. Costs for setting up and maintaining the technical infrastructure, as well as costs linked to the maintenance of the digital material itself (see section 4). Costs may vary, depending on the type of service offered to the users – for example a view only service or a full download service.

Furthermore, it is important that intellectual property rights are fully respected. Under current EU-law and international agreements, material resulting from digitisation can only be made available online if it is in the public domain - the reservoir of generally available and usable content<sup>16</sup> - or with the explicit consent of the rightholders. If there is not such an explicit consent, material that is under copyright can only be made available by public libraries in an on-site consultation service. Therefore a European digital library will in principle be focused on public domain material. An online library offering works beyond public domain material is not possible without a substantial change in the copyright legislation, or agreements, on a case by case basis, with the rightholders.

As a consequence, digitisation of Europe's cultural heritage is likely to lead to a wide online availability of older material that is no longer protected by copyrights, against paper based availability (or very limited digital availability) of most material produced in the 20<sup>th</sup> century. This may affect the way people consult and use information.

The digital environment has added a completely new dimension to the value of public domain material, since this material can be distributed through the Internet without any restrictions. Recently, the public domain has been under some pressure. The harmonisation of the term of copyright protection until 70 years after the death of the author<sup>17</sup> has, for example, brought material which was out of copyright back under copyright protection.

---

<sup>15</sup> See on the specific issue of the archives the Commission proposal of 18.2.2005 for a Council recommendation on priority actions to increase cooperation in the field of archives of Europe, COM(2005) 53 final.

<sup>16</sup> This concerns for example material for which the copyrights have expired.

<sup>17</sup> Council directive 93/98/EEC of 29 October 1993 harmonising the term of protection of copyright and certain related rights. OJ L 290, 24.11.1993, p. 9.

*This was relevant in the event of a planned public reading of Joyce's Ulysses in 2004 at the event of the 100<sup>th</sup> birthday of its publication. The public reading was stopped by Joyce's heirs, invoking their copyrights. The work had fallen into the public domain under Irish copyright law, but copyright was revived again as a result of the terms directive. Emergency legislation was enacted to make the public reading possible after all.*

Libraries and archives hold considerable amounts of material which is out of copyright and can therefore be made available online. However, even if works are in the public domain, the situation is not always straightforward.

*There may be rights attached to the different editions of a work that is itself no longer protected by copyrights, for example rights to introductions, covers and typography. The 'Bibliothèque Nationale de France' has digitised collections of 19<sup>th</sup> century authors published in the 1970s. Most of the titles are in several volumes. Because of the copyrights of the preface writer, only volumes 2, 3, ... are available online. Where possible, the 'Bibliothèque Nationale' digitises older versions to replace the modern ones.*

A particular problem is the issue of orphan works. These are copyright-protected works whose owners are difficult or impossible to identify or to locate. The problem is particularly relevant in the context of audiovisual archive resources. Since the ownership of these works is uncertain, it is impossible to get the explicit consent of the owners for the re-use or the distribution of the material. This obviously affects the possibility to integrate the material in new creative efforts.

*The problem is also a barrier for video on demand services, in particular for audiovisual works from Europe preceding the digital age, where rights issues can lead to prohibitive administrative efforts without a guarantee of success. There can be tens of rightholders – the persons involved in creating the music, the scenario, the adaptation etc, or their heirs – who all have to give their consent.*

The Scandinavian countries have in place an extended collective licensing scheme for the non-commercial use of orphan material, which allows its use, but also foresees remuneration for the authors in case they are identified.<sup>18</sup> Also in other parts of the world the issue of orphan material is receiving increased attention, as the recent consultation of the US copyright office on this topic shows.<sup>19</sup>

### 2.2.2. Digital libraries based on 'born digital' material

Digital libraries can be the result of the digitisation of material from the past, but they can also be based on material that was originally produced in digital format. More and more of this born digital information will be captured and stored in an organised way for present and future use. Repositories where digital material is stored can quickly build up to sizable digital

---

<sup>18</sup> Extended collective licenses enable users to obtain lawful user rights without the often time consuming and costly procedures of locating multiple right holders that are sometimes difficult or even impossible to find. Article 50 of the Danish copyright code applies the use of extended collective licences to a number of areas of non-commercial exploitation of protected works, such as reproduction for educational activities, reproduction by libraries, museums and broadcasters.

<sup>19</sup> <http://www.copyright.gov/orphan/index.html>

archives and libraries. ‘Classical’ libraries are among the institutions taking care of these repositories. In several countries their role in collecting and preserving digital material is backed by changes in the legal deposit laws, allowing them to obtain the complete digital cultural and scientific output.

Collecting the material in this way to form digital libraries does not imply the right for libraries and archives to distribute it. However, with the consent of the rightholders digital information can be made widely available for everyone to access. Voluntary sharing of content can have an important place in shaping the common information space.

*The Creative Commons initiative, which started in the USA, is gaining ground in different European countries. It provides a set of user-friendly online licenses giving creators of content the opportunity to protect some of their rights, while giving away others. Licences are available in human readable, lawyer readable and machine readable versions. In May 2004, the BBC announced that it would make part of its archives available online to the public for use through a Creative Commons license. This Creative Archive scheme seems to have been slowed down, amongst others because of rights clearing issues for the underlying material.*

The trend of voluntary sharing is noticeable in the scientific sector, where so-called ‘open-access’ models are emerging. These models are based on the principles of free, worldwide access to the information and the deposit of articles in an online repository, and are often physically linked to an academic or scientific library. They will lead to huge digital scientific libraries accessible for all.

Commercial publishers are critical of the new models, alleging that open access publishing will undermine the current peer review system which guarantees quality and selectivity, and will make it impossible to find quality material in a mass of information. Those advocating open access claim that it will considerably bring down the costs to the user, multiply access and take away entry barriers for younger researchers. Another argument used is that publicly funded scientific information should be accessible for all to verify and build upon.

*London’s Wellcome Trust, one of Europe’s largest charities, is planning to launch a system that will archive all papers produced by its grantees in a digital library. Wellcome will require researchers to deposit a copy of the accepted manuscript within 6 months of publication.*

In July 2004, the Commission launched a study into the scientific publishing market. Results will be available in the summer of 2005.

### **3. A FUTURE FOR THE PAST: PRESERVATION OF ANALOGUE MATERIAL THROUGH DIGITISATION**

#### **3.1. Preservation through digitisation**

In one specific area there is a strong overlap between the issues of digitisation and preservation. This is the case when digitisation is used as a means to preserve analogue material that is degrading. In this case digitisation is not primarily used to make the material more accessible, but to guarantee its survival. This is most relevant for audiovisual material, where analogue formats deteriorate with time, and precious material is lost.

*Nearly 70% of the holdings of Europe's audiovisual archives is at risk, because it is decaying, fragile or on obsolete media. About 25% of the total archive is in such a state that the original will be damaged or even destroyed during cleaning and digitisation. Every year Europe's audiovisual archives lose 10,000s of hours of the oldest parts of their collections.<sup>20</sup>*

Digitisation can also be a way to save books and documents in libraries and archives deteriorating as a result of amongst others acidification. This concerns in particular paper collections from the middle of the 19<sup>th</sup> century until the 1950s.

*To illustrate the situation: because of the extent of the problem of damaged works in Germany (5 million archival units and 18 million books are seriously damaged) not every item can be preserved in its original form.*

*For the Netherlands it is observed that if one copy of every Dutch document from the 1840-1950 period is preserved, it will amount to some 400,000 books, 30,000 volumes of periodicals, 1500 metres of newspapers, and 2 million manuscripts and letters, divided over all libraries in the Netherlands. Preservation takes place through registration, filming, reliable storage and, on a limited scale, digitisation and deacidification. Accomplishing this task will require at least 20 years.<sup>21</sup>*

The example shows that digitisation is only one of the ways of preserving the material. It has the disadvantage of being relatively expensive, compared to microfilming and even compared to preserving the original. In addition, microfilms are very stable, contrary to the digitised information that requires constant maintenance to be kept 'alive' (see below). On the other hand, digitised material can be made accessible to a much larger public through the Internet, thus combining preservation and accessibility functions. Therefore, sometimes hybrid strategies are followed, combining microfilms for preservation purposes and digitisation for online availability.

For rare works which are frequently consulted digitisation has another advantage. If the digitisation quality is high enough, consultation of the digital copy can replace the physical manipulation of the original, which will add to its longevity.

### **3.2. Challenges for preservation through digitisation**

The challenges of preservation through digitisation are the same as for digitisation in general: costs, organisational issues, technical progress and legal barriers, as presented in section 2.1.3. The following examples add some elements regarding the financial challenges particularly relevant for the preservation of analogue material.

*According to the results of the PRESTO project, the total cost of preserving the world's audiovisual archives by simple format transfer would be around 100 billion €.*

*In 2003, the 'Institut National de l'Audiovisuel (INA)', indicated that an immediate complementary financing of 40 millions € was necessary to prevent 40 % of the*

---

<sup>20</sup> Survey by the IST Presto project which finished in October 2002. <http://presto.joanneum.ac.at/index.asp>

<sup>21</sup> Preservation map of Europe, ECPA <http://www.knaw.nl/ecpa/map/index.html>

*hours of audiovisual material at risk from being irremediably lost between now and 2015.*<sup>22</sup>

### **3.3. Actions at European level**

The specific issue of the preservation of analogue material through digitisation is addressed at Community level by different means and is an important point of concern in the co-ordination actions on digitisation and preservation. The area of the audiovisual heritage has received specific attention. On 16 March 2004 the Commission adopted a draft Parliament and Council recommendation on the cinematographic heritage and the competitiveness of the related industrial activities.<sup>23</sup> The draft recommendation addresses the need to ensure the availability of Europe's cinematographic heritage over time and calls on the Member States, amongst other things, to allow copying for preservation purposes. The European Parliament and Council have reached an agreement on the draft recommendation, which will be formally adopted shortly.

Also in this area, research projects co-funded by the Community contribute to the goal of keeping the content accessible for the future. Within the European research programmes, the PRESTOSPACE project is particularly relevant, since it directly deals with the challenge of getting the costs of preservation through digitisation down.

*PRESTOSPACE (9 M€ co-funding) is developing a toolkit which audiovisual archives can use to digitise audiovisual material. The tools will lead to more affordable and better quality digitisation in view of converting and preserving analogue audiovisual material. The project is led by the 'Institut national de l'audiovisuel'. Partners are 8 audiovisual archives, 3 applied research institutes, 6 universities and 15 industrial partners (all SMEs).*

## **4. A FUTURE FOR THE PRESENT: PRESERVATION OF DIGITAL MATERIAL**

### **4.1. The issue of digital preservation: what is the problem?**

With the advent of the Information Society, the supply of information is growing exponentially.

*According to a US study, print, film, magnetic, and optical storage media produced about 5 exabytes of new information in 2002 (10<sup>18</sup> bytes, comparable to all words ever spoken by human beings). Stored information grew about 30% a year between 1999 and 2002 Ninety-two percent of the new information was stored on magnetic media, mostly hard disks.*<sup>24</sup>

Most of this information is produced in digital format. This raises the question as to whether it possible or worthwhile to preserve it all. But who decides and who is responsible for preserving what and in what way? Even the question of how to preserve online digital content

---

<sup>22</sup> Rapport du Sénateur Broissia du 23 novembre 2003 sur le projet de loi de finance 2004 pour la communication audiovisuelle, quoting Mr. Hogg of INA.

<sup>23</sup> COM(2004) 171 final, OJ C 123, 30.4.2004, p. 1.

<sup>24</sup> "How much information? 2003" UC Berkeley's School of Information Management and Systems.

is far from being answered. A 2002 estimate indicated that the normal lifetime of a webpage varies from 44 days to 2 years.<sup>25</sup>

Without active intervention a lot of the material will be lost or become unreadable. One of the reasons is the succession of generations of hardware.

*To mark the 900th anniversary of the Domesday Book in 1985, a new multimedia edition was compiled and published in 1986 at a cost of 3,7 million €. In 2002, there were great fears that the disc would become unreadable as computers capable of reading the format had become rare. It could be saved because a system was developed capable of accessing the discs using emulation techniques. Interestingly, while there are difficulties accessing digital data from 1986, the original Domesday Book, now over 900 years old, can still be consulted.*

The rapid succession and obsolescence of software is another factor that may cause information to be lost. Unless data are migrated to current programs or care is taken to preserve the original source code, retrieval of information may become very costly, if not impossible. This is particularly true of ‘closed’ data formats, for which the source code is not publicly known. As generations of file formats, software, and platforms succeed each other rapidly, the moment that we cannot read the information any longer is always just around the corner.

A third reason for the loss of digital information – along with the obsolescence of computer programmes and that of software – is the limited lifetime of digital storage devices.

*Contrary to what many people believe, the lifetime of CD-ROMs and other digital storage devices is limited. One estimate<sup>26</sup> indicates that commercially-produced compact discs have a life of 10 to 25 years under normal preservation conditions. Recordable blanks have an even shorter life (5-10 years after recording). This will in due course lead to unpleasant surprises for individuals as well as organisations.*

Maintenance of the content and migration to new supports is necessary to ensure that the material is not lost. For this reason digital preservation is as relevant for digitised material as it is for ‘born digital’ material. The risk of losing digital material has to be taken into account in any digitisation programme. Digitisation without a proper preservation strategy may become a wasted investment.

## **4.2. What has been done so far to tackle the issue?**

### *4.2.1. A worldwide reply*

As information is produced more and more in digital format, libraries and archives, in their institutional tasks as keepers of our intellectual heritage are expected to come up with a magical solution to keep digital information alive for future generations. The truth is that there

---

<sup>25</sup> *Preservation risk management for web resources*, A.R. Kenney et al., D-Lib Magazine, Volume 8:1, January 2002, quoted by Y. de Lussenet of the European Commission on Preservation and Access in her paper *preservation of digital heritage* prepared for UNESCO (March 2002).

<sup>26</sup> Mrs. Lupovici of the French National Library at the ABDU congress on the preservation of digital documents (September 2001). The estimates are based on laboratory experiments.  
[http://www-sv.cict.fr/adbu/actes\\_et\\_je/je2001/CathLUPO\\_140901.html](http://www-sv.cict.fr/adbu/actes_et_je/je2001/CathLUPO_140901.html)



is little experience with digital preservation, that the legal framework is evolving, that resources are scarce and that the outcome of preservation efforts is uncertain.

The problem of preserving digital information is gradually receiving increased attention worldwide. The US acted in 2000, when the Library of Congress was empowered to develop a national program to develop standards and a nationwide collaborative collection and long-term preservation strategy for digital materials.

*In December 2000, Congress appropriated \$99.8 million (81.8 M€) for this effort, calling on the Library to spend an initial \$25 million (20.5 M€) to develop and execute a congressionally approved strategic plan for a National Digital Information Infrastructure and Preservation Program. Congress specified that \$5 million (4.1 M€) of the appropriation may be spent during the initial phase for planning as well as the acquisition and preservation of digital information that may otherwise vanish.*

In 2003, the General Assembly of UNESCO adopted a charter on the preservation of the digital heritage.<sup>27</sup>

*It stresses the need to ensure that the world's digital heritage remains accessible to the public and that the digital heritage materials, especially those in the public domain, should be free of unreasonable restrictions. It underlines the need for action to avoid loss of digital material and the need to alert policy makers and the general public to the potential of the digital media and the practicalities of preservation.*

In reaction to the rapid disappearance of web-content, several private and public initiatives decided to tackle the preservation of today's web for future generations.

*An example is the US based non-profit Internet Archive, started in 1996, aiming at capturing and preserving the worldwide web content that is freely available.<sup>28</sup>*

*The IIPC is a collaborative effort (6 national libraries from EU-countries, the US library of Congress, the National Library of Australia, Library and Archives Canada, national libraries of Iceland and Norway, and the Internet Archive). It aims at acquiring, preserving and making accessible knowledge and information from the Internet for future generations everywhere, promoting global exchange and international relations.<sup>29</sup>*

*In Europe, the Swedish Kulturlw3 Heritage project of the Swedish Royal library is of interest. The project captures all websites in the Swedish domain or providing content on Sweden with the help of robots.<sup>30</sup>*

All in all, in spite of these initiatives, and of the efforts undertaken by libraries and archives in the Member States to tackle the issues of born digital material, many questions remain open and investments made in the past in the production of content can be irremediably lost.

---

<sup>27</sup> Adopted at the 32<sup>nd</sup> session of the General Conference of UNESCO, 17 October 2003.

<sup>28</sup> <http://www.archive.org/web/web.php>

<sup>29</sup> <http://netpreserve.org/about/mission.php>

<sup>30</sup> <http://www.kb.se/kw3/ENG/Description.htm>

#### 4.2.2. *Actions in the Member States*

In response to the challenge of preserving digital material, Member States have taken legislative action, introducing obligatory deposit schemes for digital material or extending existing legislation to digital material. The legislation of a large majority of Member States now also covers digital off-line material (such as CD-ROMs). Some Member States, such as Finland, Sweden, and the UK have extended these schemes to include dynamic online publications (Internet), whereas others (such as France, Lithuania and Slovenia) are considering similar legislation and/or experimenting with dynamic web-archiving.

The idea behind the schemes is that it is better to deal with the issue in a pre-emptive and organised way instead of having to implement costly recuperation schemes later or to lose digital material that may be relevant.

*According to a report of the Science and Technology Committee of the House of Commons,<sup>31</sup> delays in implementing the mandatory legal deposit to digital material and the inability of libraries to pay for material that publishers are not willing to relinquish on a voluntary basis (which libraries would receive for free if it were in printed form) is already leading to a gap of up to 60% in the deposit of electronically-delivered publications, including STM journals, which represents a significant breach in the intellectual record.*

Legislation needs to be accompanied by the adaptation in the preservation infrastructure, necessary to deal with large quantities of digital objects. National preservation plans, where they exist, tend to concentrate in the first place on safeguarding analogue material at risk, rather than to venture in the area of the preservation of born digital material.

### 4.3. **Challenges for the preservation of digital material**

**Financial challenges:** At present, the real costs of long-term preservation are not clear.

*Costs will depend on evolving factors such as storage costs but even more so of the number of migrations needed over time. Costs involved in the migration will include the efforts necessary to check the integrity of the digital object after migration.*

It is, however, clear that due to the limited resources available for the preservation of digital content, institutions responsible for preserving information are forced to make choices (which in itself generates a certain labour cost). Selection for preservation is not necessarily a one off choice. Depending on factors such as the archival value and the use of a digital object as well as the resources available, a choice can be made at later migrations as to whether it is worthwhile to continue investing in it.

Since in many cases the same institutions are responsible both for digitising and for preserving content, it can happen that these two 'priority areas' compete for the same scarce resources.

**Organisational challenges:** In the area of preservation, some collaboration across borders exists, aiming at an exchange of good practice and a more co-ordinated approach.

---

<sup>31</sup> HC Science and Technology Committee report 'Scientific Publications - Free for all?' - HC 399-1, July 2004, p. 93.

Nevertheless, the risks of widely divergent approaches and duplication of efforts remain. Here too, good collaboration between public and private players is essential. Some interesting initiatives exist, such as the agreement between the National Library of the Netherlands (KB) and Elsevier:

*Under the agreement, the KB receives digital copies of approximately 1,500 journals covering all areas of science, technology and medicine, currently published by Elsevier. The KB will ensure preservation, for example by migrating the content and associated software as technologies change. The KB will provide access to the journals to all who come to the library and are permitted access to the library's collections.*

The issues on the preservation of scientific information in a digital environment, however, go far beyond the future availability of scientific journals. Appropriate mechanisms are to be defined to avoid unpleasant surprises over the medium term, such as loss of data and experiments which cannot be repeated.

Preserving information in the digital age requires new ways of working. For efficient preservation organisational issues within the institutions will have to be addressed, such as upgrading the skills of the staff working in libraries and archives.

**Technical challenges:** Preserving digital content so that it can be accessed, trusted and re-used in the future is emerging as a key challenge. It embraces technology aspects of hardware and software, storage devices, peripherals, as well as increasingly complex questions of how to organise, describe and store these resources. Limited work has been done on how to keep text, data and images, for example by migration and emulation, but there is inadequate knowledge yet of dealing with the specific characteristics of emerging digital formats, such as digital audio, digital video, digital models and simulations.

Increasingly born-digital objects are dynamic, meaning that they change as a result of user interactions or adding new data. There is a need to develop methods, tools and technologies that can preserve dynamic content. Solutions will need to be built upon automation – of the preservation processes, and of the analysis of the content. They will need to be scalable. Their use needs to be tested on very large volumes of data and, also, at the other end lead to affordable tools usable by individuals without archiving or technical skills.

**Legal challenges:** IPR-issues are also relevant in relation to the preservation of information, since preservation depends on copying and migration. In this context, the introduction of technological protection measures and digital rights management solutions raises a whole set of new issues. Copying and migration are exactly what the technological protection measures are trying to prevent. For effective preservation unprotected copies need to be made available by those who produce the information.

Another challenge is linked to the new legislative measures extending legal deposit obligations to dynamic online material. If all Member States follow suit there is the risk of 25 (or more) sets of rules, imposing different obligations on content producers with cross-border activities.

#### 4.4. Actions at European level

Some initiatives at European level have addressed the issue of digital preservation. A Council resolution of May 2002 underlined the importance of the issue and called on the Member States to tackle it.<sup>32</sup> The issue of the preservation of digital information is also touched upon in the Commission proposal of 18 February 2005 for a Council recommendation on priority actions to increase cooperation in the field of archives of Europe.<sup>33</sup>

Furthermore, a limited number of projects under the research programmes have started to address the issue of digital preservation. One example is the ERPANET project.

*The ERPANET project that ran between 2001 and 2004 (co-funding 1,24 M€) addressed the lack of awareness, fragmentation of knowledge and skills amongst the stakeholder communities about how to handle existing preservation problems. It brought together museums, libraries and archives, ICT and software industry, research institutions, government organisations, entertainment and creative industries to address the challenges posed for preservation by the widespread use of digital technologies.*

---

<sup>32</sup> Council Resolution of 25 June 2002, Preserving tomorrow's memory: preserving digital content for future generations, OJ C 162, 6.7.2002, p. 4.

<sup>33</sup> COM(2005) 53 final