Digitisation of Cultural Documents

Maria PETRESCU
Octavian Goga County Library, Cluj

Keywords: digitisation, cultural documents, preservation

Abstract:
This paper deals with the digitisation of documents, the advantages and characteristics of digitisation, as well as legal aspects connected to this enterprise. Digitisation (numerisation, scaning) is an information codification procedure which allows document copying. It can also be defined as a method of transforming information documents such as books, photography, sound or visual recording into electronic documents for purposes of preservation, information circulation, and a better accessibility to the public at large.

E-mail: lolapetrescu@yahoo.com

*

> Motto: Digital documents are addressed
> to the Large Public from everywhere!

"Molière's character was surprised when finding out that he spoke in prose. The shock of present-day people is even greater when they find out that, regardless of their language and national alphabet, they all use a universal script of only two signs, and what is more, these are the numbers 1 and 0, called bits or digits".[1] "The word digital, used in computer science or electronics, has an exact meaning: codified in the form of numbers".[2]

Digitisation (numerisation, scaning) is an information codification procedure which allows document copying (manuscripts, music, photographs, etc).

---

[1] Mircea Malița, *Zece mii de culturi, o singură civilizație* (Ten thousand cultures, one single civilization), Bucharest: Editura Nemira, 1998, p. 93.
[2] *Evoluția tehnologiei* (The evolution of technology), Oradea: Editura Aquila '93, 2001, p. 56.

Digitisation is the transformation of book pages or microfiche in numerical data. Digital images thus created are then sent to a server and published online, allowing online users access to that document.

Digitisation can also be defined as a method of transforming information documents such as books, photography, sound or visual recording into bits.

"Bits (short for *binary digit*) represent the smallest information unit displayable on a computer".[1]

So that people may use and understand digital data, they are reconstructed in an analogous way with help of special processors called digital-analogous converters (DAC) found on video and sound cards of computers.

"The digital revolution which influences all aspects of our lives is grounded on the fact that any data type (text, image, sound, even 3-D objects) may be converted into a series of numbers. The ability to codify any type of information as a series of binary numbers has revolutionised, ever since the 1980s, the way in which people communicate, transmit information, buy and sell goods, create works of art and entertainment, and even manufacture products. Huge quantities of digital data are created and processed every second all over the world. There is a wide range of variety and purpose of the products created from digital data, comprising everything from educational and entertainment products to business databases and 3-D media."[2]

One purpose of digitising cultural and other documents is to make information resources more accessible and easier to use in the online environment.

Digitisation has several advantages:
- digital data are easily copied and manipulated;
- numerical information may be compacted (re-codified to smaller size) with only a little or no information lost;
- numerical data can be stored and transmitted by the same kind of method, irrespective of the data type (text, image, or other);[3]
- these data are very fast (the search engine finds texts of thousands of pages in just a few seconds) and easy to use (at any time, not limited to the opening hours of a library);

---

[1] Margolis, P. E. *Dicţionar P. C.* (P.C. Dictionary), Bucureşti: Editura Nemira, 1997, p. 40
[2] *Evoluţia tehnologiei*, p. ???
[3] http://www.anbpr.org.

- many unknown works can be made accessible to a wider public by digitisation (such as Special Collections works);
- it offers the possibility of creating clearer, brighter images, black-and-white or coloured, larger (or smaller), according to the users' needs and preferences;
- duplicates last for ever, can be preserved in time, circulate incomparably easier, are presented, archived, or used more comfortably, etc.
- offers a greater accessibility to the public, meaning that the use of a high capacity automatised storage system makes digital collections available 24 hours a day, employing a very low number of personnel. Digitisation offers thus distance access as well, allowing for a larger number of people to have access to the same kinds of documents at the same time;
- unique and valuable;
- eliminates distances, wasting time, or other kinds of losses;
- makes the search easier and more efficient.

Digitisation offers the possibility of searches in well localised catalogues on the Internet according to various criteria. Furthermore, it:

- offers the possibility of using a great number of editing functions (copying, shortening, prolonging, attaching, etc.). In the case of comparative study, several documents or fragments can be analysed at the same time. The user may study several documents, may shorten them, attach sound or music to them, can adapt them according to his wishes or needs, without harming the original;
- it makes possible to preserve and conserve the original document. In case of several collections, digitisation facilitates the access to these documents.

Digitised copies:

- facilitate and save documents and reduce the frequency of direct consultation and as a result the originals' wear factor, protect them from bad weather, theft, etc. Thus it is possible to protect rare, fragile, bulky, precious, sensitive, or old documents, improving their preservation and saving them from deterioration or disappearance.
- favour the use of documents in the sense of valorising special, prestigious collections, manuscripts, which may thus be made accessible to the users;

- enhance the process of permanent formation, training, and education of the library's users;
- they are the most up-to-date method of circulating information;
- they represent a highly efficient way of knowing and promoting our specific, national values, as well as international ones;
- contribute to the national and international digital stock;
- they achieve the dream of scientists of the past: the creation of a universal library (a project which is being currently worked on).

The disadvantages of digitisation could be:

- several types of documents, especially images, take up very much space;
- at present, the cost of digitisation is quite high;
- the digital copy made by current technologies may not be as accurate as the original.

As apparent from those said above, the advantages of digitisation are incontestable, and digitisation itself can not only be considered as a preservation procedure, but it also facilitates the study of fragile documents, offers a greater capacity of information storage, and all these secure optimal research conditions for the Large Public from everywhere, whom we all serve.

In order to emphasise the potential of digital technologies regarding the dissemination and easy access to information, the European Commission recommends the following three main directions:

- the digitisation of traditional collections with the purpose of large-scale use in informational society;
- preservation and storage in order to secure the access of future generations to digital documents and to prevent the loss of valuable materials;
- online accessibility as a precondition for the maximisation of benefits that citizens, researchers, and companies – that is, the users – may obtain from information.

In Romania, this modern procedure of archiving and preservation of documents seems to be only in an incipient stage (a pilot project at the G. Asachi County Library in Iaşi), and, unfortunately, not supported by Romanian cultural policies.

In Europe the consensus has been reached that digitisation offers strategic opportunities and that it must be continued at a higher speed.

The European Council has declared to act as a coordinator of this vital activity, stating that "There is a real demand for digital content

among citizens as well as within the research community. Digitisation and online accessibility of our cultural heritage can fuel creative efforts and support activities in other sectors, such as learning and tourism, thereby enhancing competitiveness and growth across Europe".[1]

The importance of the digitisation of European cultural heritage and digital preservation has already been recognised by the European Commission in its *eEurope* action plan. This has lead to the acceptance of the Lund Principles in 2001 with the pertaining action plan, as well as to the foundation of a National Representatives Group on digitisation problems. "The Lund meeting (Sweden) in April 2001 was a very important event, which showed that the coordination of national initiatives is both desirable and necessary, stating that digitisation provides a key mechanism for exploiting the unique European heritage and supporting cultural diversity, education, and the foundation of e-content industries".[2]

The degree of economic provocation shows that a European-level or at least national cooperation would be welcome in order to reach to an optimal use of available resources. This national or European cooperation would make it possible to avoid multiple digitisations of identical documents (and thus save expenses), by digitising one document only once, and make it accessible online for all interested uses, regardless of where they are in the world.

Although the digitisation of the cultural and scientific European heritage is a major necessity from a cultural and economical point of view, only less than 2 % of the stocks of libraries and archives are now available in a digital form.

The majority of the European Union Member States worked as partners in various European digitisation projects (Minerva, Minerva Plus, etc.), and it can be appreciated that all of them continue the process of digitisation depending on their cultural policy, as declared on various international meetings. One such meeting was that of Alicante (Spain) in 2002, where each participating state (Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Portugal, Sweden, etc.)

---

[1] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2006:297:0001:0005:EN:PDF

[2] *Linii directoare Pulman: Noi orientări în bibliotecile publice* (Bucureşti: Editura Biblioteca Bucureştilor, 2002, p. 137.

presented a description of their own policy about the digitisation of their cultural and scientific heritage.

For example, as the French Minister of Culture and Communication, Jean-Jacques Aillagon declared that "today there are several projects launched in France in order to secure the digitisation of the cultural heritage. Large institutions elaborate strategies for ensuring the preservation of a true digital heritage. The National Audiovisual Institute has launched a plan of preserving and digitising their own heritage, while the National Library of France intends to digitise and archive French web pages.

Each nation possesses its own cultural heritage, its history, its institutions, and its own specificity, and deals with preserving them the best it can by digitisation. All these aspects participate in constructing a memory which must be preserved for future generations".[1]

I. Stages of a digitisation project

A. Planning a digitisation project

Each digitisation project has its logic, or its purpose. Sometimes such a project is necessary for providing more cultural information on the Internet. In other cases digitisation is made in view of a better use and protection of the already existing sources of information.

In yet other cases the project is achieved by the cooperation of various institutions (mostly libraries), elaborating a web page on a useful subject, directed to a certain target-audience. The motivation of the project is paramount for deciding the criteria for the sources to be digitised.

- The **selection of documents** to be digitised represents a very important decision within a digitisation project. Usually it is an entire collection of books or rare, valuable objects which is digitised.
- The **selection of the content** (texts and images) must be made by persons who are familiar with the original collections, their documentary and physical characteristics (aspect, state, etc.), and therefore may choose those which are the most useful for the potential users and can be adapted to information searching methods. The first step is the localisation of the works, followed by the verification of the originals (page by page, to see if they

---

[1] http://www.journaldunet.com/itws/it_aillagon.shtml

are complete), and if they are deteriorated, they must be rehabilitated. Then it must be decided whether the work will be directly digitised, or it will be transposed first to an intermediary film support as a preservation copy, and digitised only afterwards.

- **Establishing the selection criteria**. These depend on the objectives of the project, and technical, financial, or copyright matters alike, as well as on the experiences of other digitisation projects.

The main selection criteria for documents, as defined by the Research Committee of the European Minerva Project, are as follows:

- uniqueness of the information;
- interest of the cultural collection for a wide target-audience;
- intellectual property rights clarified;[1]
- establishments of risks. All possible variants must be considered before beginning a successful digitisation project, while most risks should be eliminated.

B. Definition of the legislative framework

All digitisation projects raise certain legislative issues, which must be anticipated and solved before the beginning of digitisation.

Digitisation means the creation of a new copy, which raises the issue of copyright.

Online accessibility

The traditional model of library services based on reading room access to library units or home loan is not easily transposable to a digital medium.

According to the current legislation of the European Union and international agreements, digitised materials may only be accessible online if they pertain to the public domain, or with the explicit agreement of copyright holders.

Thus the European digital library will primarily focus on public domain documents.

For literature, this means only works written prior to the beginning of the 1900s or somewhat later, which do not have copyrights, depending on the year of their authors' death. Even if the works themselves are not copyrighted, the situation is still not always clear.

---

[1] http://www.culture.gouv.fr/mrt/numerisation/fr/f_minerva.htm

There might be copyright issues involved connected to certain editions of the works, such as copyrighted prefaces, covers, etc.

Directive 2001/29/EC on the harmonisation of certain aspects of copyright and related rights in the information society provides for exceptions and limitations for certain reproduction types for the benefit of public institutions such as libraries, museums, or archives. These exceptions are not applicable in all cases, and are subject to different variants in the Member States.[1]

In Romania, this matter is regulated by Law 8/1996 on copyright and related rights.

An online library which would offer works beyond the limit of public materials could not be implemented without serious modifications to copyright laws or agreements on a case-to-case basis discussion with the rightholders.

In this context the clarification and transparency of copyrights is paramount.

C. Human resources

Following the solution of copyright issues, the participating staff must be instructed with regard to several new concepts and activities differing from the usual day-by-day activities in a library.

D. Digitisation (Methods of application)

Prior to digitisation proper, documents must be processed and prepared for digitisation.

This phase is indispensable since documents must be easily identifiable.

The operation of digitisation can be made by the institution itself with its instructed staff, or by a specialised institution.

Digitisation can be made in two ways (depending on the materials to be digitised):
- with a scanner, or
- with a digital camera.

Organization of the digitisation workshop

---

[1] Cf. http://eur-lex.europa.eu/smartapi/cgi/sga_doc?smartapi!celexapi!prod!CELEXnumdoc&lg=en&numdoc=32001L0029&model=guichett, Chapter II.

The organisation of the digitisation workshop is similar to a binding workshop, only much more modernized, based on specialised staff and necessary equipment:

- a digitisation station made up of a server, an IBM-PC computer with FUJITSU or DUOSCAN (Agfa) scanners, one for formats up to A4, and one for A0 formats, or any other types of computers with standard scanners and software;
- a digital camera;
- image processing software (Photoshop);
- webpage design software;
- paper cutting machine;
- solutions and utensils for operations of rehabilitation of pages or images;
- magnetic discs for "online" memorisation.

Digitisation techniques

Books are digitised in the digitisation station with help of a scanner, preferably a FUJITSU FI 4120 C. This scans as many as 20 pages/minute, with A4 as the largest supported format.

Digitisation is usually made with 300 dpi (or as decided at the beginning of the project), black-and-white or coloured, based on the nature of the digitised document.

The recognition of optical characters is made possible by a specific module of the Acrobat software, creating .pdf files which can be saved on electronic support or uploaded to the Internet.

The digitisation of geographical maps larger than A3 format is made in the same way, but with an A0 scanner.

The digitisation of photographs happens with help of a digital camera.

As a result of the operation of digitisation, the documents are represented on a screen as identical with the originals.

The image obtained by digitisation is thus the preserved image, the electronic image, which, once copied on an electronic support, becomes an **electronic document**.

E. **The duration** of accomplishing a digitisation project differs on a case-to-case basis.

F. **Processing of electronic documents**

It is made according to international norms (ISBD-ER) of standardised description in order to be integrated into the library's electronic catalogue. "Then their topographic number is established, which physically integrates them in the collections and indicates their format and location of the shelves".[1]

G. Economic aspects

"The estimation of costs varies according to the quality of the original, the types of images, and the document's necessary treatment prior to digitisation".

The average price for the digitisation of one page is between 0.5-1.0 Euros.

The cost of the Google project digitising 15 million books is estimated to 150-200 million dollars."[2]

H. Guidelines for quality practices

"The digitisation process is apparently easy, because any computer with a low-cost scanner and standard software can be used for creating a readable digital copy of any document. Still, when having in mind issues such as the quality of the digitised image, the format used for its storage, description, usage, and preservation, the process becomes much more complex."[3]

The simplicity of the basic process may create the false impression that a digitisation project is easy and cheap to implement, but unfortunately this is not the case, as proven by the guides of libraries which have already undertaken such digitisation tasks.

The supply of digital documents is done relatively directly, as the development of the Internet and websites have provided the necessary infrastructure, software, and technical standards.

II. Preservation and conservation of electronic documents

"Electronic documents presuppose the existence of specific means of preservation and conservation. The storage area is organized in specially designated places. The area must offer optimal climate conditions.

---

[1] http://www.culture.gouv.fr/mrt/numerisation/fr/f_minerva.htm
[2] http://www.zdnet.fr/actualites/internet/0,39020774,39228791,00.htm
[3] *Linii directoarePulman*, p. 136.

Preservation and conservation are activities which aim at the long-time care of documents, therefore deserve our greatest attention."[1]

The international norms applicable for the preservation of electronic documents are stipulated by ISO/CD 1611, and are as follows:

- temperature between 5°C - 20°C;
- relative humidity: 30 % - 50 %;
- maximum temperature gradient: 4°C per hour;
- relative humidity gradient: maximum 10 % per hour.

"Electronic documents are stored in special furniture (closed up in special racks) designed to protect them from scratching, mechanical shocks, biological factors (rodents), and fingerprints.

In order to secure their optimal presevation conditions, their storage place (within the general storage area) must be well delimited. This place must favourably secure these types of documents, and also protect electronic support in relation to other types of support.

The storage room must be a special one, with a window protected by metal grates and PVC roller blinds for sun protection, as well as a metal door.

Environment factors must be permanently monitored, and, when they are requested, they are carefully manipulated so that they would not be mixed up with other supports, or collide with hard surfaces, or have fingerprints.

They must be manipulated delicately, so that they would not be mixed up with other kinds of documents during transportation to and from reading rooms.[2]

Conclusions

"Leading to the creation of new types of documents, digitisation overturns the centuries-old practices of libraries, and invites librarian professions to important reorientations, especially in the field of technological abilities. Although for the time being digitisation is a costly enterprise, it is hard to believe that any more competitive way could be found to reduce the pressure that falls on much requested collections."[3]

Translated by Emese G. Czintos

---

[1] http://www.bcub.ro/continut/unibib/prezervare_restaurare.php
[2] Ibid.
[3] Ion Stoica, *Criza în structurile infodocumentare* (Crisis in info-documentary structures), Constanța: Editura EX PONTO, 2001, p.56.