Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação, v. 17, n. esp. 2 – III SBCC, p.1-18, 2012. ISSN 1518-2924. DOI:10.5007/1518-2924.2012v17nesp2p1

LONG-TERM DIGITAL PRESERVATION AND SCHOLARLY COMMUNICATION

Elena Maceviciute[i]

PRESERVAÇÃO DIGITAL À LONGO PRAZO E COMUNICAÇÃO CIENTÍFICA

1 INTRODUCTION

In 2008 the Swedish School of Library and Information Science of the University of Borås joined the European Union research project SHAMAN – Sustaining Heritage Access through Multivalent Archiving. In one of the previous articles we have explained the aims of this project as follows:

> The aim of the SHAMAN Integrated Project is to investigate the long-term preservation of large volumes of digital data in a distributed environment by developing a preservation framework that is verifiable, open and extensible. The approach will investigate all aspects of digital preservation from ingestion to dissemination in an environment where the collections, producers, consumers and curators are geographically distributed and the content of the collections is of a dynamic nature. Furthermore, it is developing corresponding preservation tools for analysing, ingesting, managing, accessing and reusing information objects and data across libraries and archives. Three prototypical applications are intended to support trials and validation of the result in memory institutions, industrial design and engineering and, finally, experimentally, also in scientific application domains. The SHAMAN data grid infrastructure was developed in close cooperation with US project partners. (Maceviciute and Wilson 2011: 2).

[i] University of Borås, Sweden; Vilnius University, Lithuania - elena.maceviciute@gmail.com

This project opened a number of interesting avenues of thinking either about preservation of digital objects and the differences of organizations and communities that see it from various points of view and construct its necessity in very different ways. This paper aims to present the understanding and perspectives of digital preservation from the points of view of scientists and scholars as well as professional keepers and curators of our documented heritage as revealed through investigation of SHAMAN requirements, evaluation of SHAMAN results or of those presented in their texts.

2 LONGEVITY OF MEDIA

Let me start with an overview of some science fiction literature pertaining to preservation of information.

There are many fiction books presenting the image of future libraries (see an essay by James Gunn, Libraries in science fiction) and books. But to my greatest surprise not many mention any means of knowledge preservation. Some provide a glimpse of it through the means of education and knowledge transfer, but preservation is taken for granted.

The first known knowledge preservation medium in literature is the Biblical fruit in the Garden of Eden – a perfect container of compressed information accessed by unauthorized users.

In one early fantastic novels of an apocalyptic future, David H. Keller's "The Cerebral Library", from 1931, human readers read a book a day; after five years they are killed and their brains are put in a jar to provide instant access to everything they have read. The librarians of that era had to cope with more serious problems than theft, vandalism, inadequate budgets, and low pay.

Ray Bradbury's Fahrenheit 451 (1953), on the other hand, was placed in a future society in which books were burned and a few rebels memorized favourite texts so that they would not be lost. Walter M. Miller in his "Canticle for Leibowitz" depicts monks rescuing books from deliberate destruction and safekeeping them for the future, 600 years after a nuclear catastrophe.

The English philosopher Olaf Stapledon in his novel "Star Maker" describes the formation of collective minds from many telepathically linked individuals, on the level of planets, galaxies, and eventually the cosmos, who share and develop further all they know.

Enc. Bibli: R. Eletr. Bib. Ci. Inf., ISSN 1518-2924, Florianópolis, v. 17, n. esp. 2 – III SBCC, p.1-18, 2012.

Isaac Azimov in his Foundation trilogy written in the 50's presented his Foundation as the finest achievement of humankind, whose knowledge would be lost if it was not preserved by the Foundation. One hundred thousand encyclopaedists are isolated on a distant planet to write the Encyclopedia Galactica. It would reduce 25,000 years of barbarism, which would normally follow the fall of the galactic empire, to only 1,000. In some passages Azimov refers to mundane microfilm as a discreet way of passing on knowledge, but the main flaw in his thinking is the idea of centralisation which, of course, could not take into account the present distributed communication networks, but disregarded even libraries and documentation centres proliferating around the world in every remotest corner at that time.

If the science fiction writers have neglected the issues of preservation and imagined the destruction of our knowledge only in terms of burning, breaking, tearing or flooding, they had a good reason for that. We know that stone carvings lasts for millennia, acid-free paper can be trusted for at least 1,000 years, microfilm seemingly can last for 500, photographic slides for 100, digital linear tape for 300, DVD for 100 years. The newer the media, the less it lasts with slight variations or we are not quite certain of their longevity. But we used to have long-lasting media.

Unfortunately, we do not have it anymore. We know that photographic slides last, but they cannot be displayed without the projectors. We scan the slides and convert them into digital slides and then software comes into sight. As soon as it becomes obsolete our slides will be inaccessible for viewing.

Whether born digital or digitized the digital record is associated with:

- the software employed to produce it;
- the hardware employed to produce it;
- the record format or formats (e.g., pdf files with embedded jpg files);
- the hardware for reading the format(s);
- the software for reading the format(s);
- the institutional policies on documentation (optional for our personal documents);
- the standards observed in all of these (SHAMAN 2008).

The complexity of digital records that you see on the screen, or rather the fast change in each of the displayed elements is the cause of the obsolescence of our digital records, those

stored in the bottoms of desk drawers on obsolete floppies and CD-ROMs, not because the records have vanished, but because we have no means to access them anymore.

This obvious threat of loosing the digital record in the nearest future stimulated the measures taken by national and international bodies to find the ways of keeping it intact and brought digital preservation to the agenda of technological development.

3 LONG-TERM DIGITAL PRESERVATION – CONCEPT AND ELEMENTS

What do we mean by digital preservation? There are several definitions that characterize preservation as an activity over time that should help to reduce the threat posed by technological change, as listed below:

- the management of digital information over time (Wikipedia);
- the long-term maintenance and upgrade of digital files on digital storage media (UNESCO 2003);
- the process of ensuring that a digital object is accessible over the long term on digital storage medium (Digital Preservation Coalition);
- the series of managed activities necessary to ensure continued access to digital materials for as long as necessary. Digital preservation refers to all of the actions required to maintain access to digital materials beyond the limits of media failure or technological change. (Harvey 2005: 13; Beagrie and Jones 2001: 24).

As you can see, there are two obvious aspects of digital preservation in these definitions: the time perspective and a dynamic process. The main goal is related to the timeline.

3.1 Goals

We think of the long-term digital preservation mainly from the perspective of memory institutions, such as museums, libraries and archives. This usually includes a perspective that can be named as “communication with future” (this was a slogan of SHAMAN project) and the time line of “eternity”. So, in fact digital preservation presumes as its main goal the communication with the posterity by ensuring eternal access to digital objects. Is it really the case for all?

3.2 Objects

Preservation must ensure that the digital objects are accessible. What are these digital objects then? The definitions list digital materials, digital files, and digital information.

These objects should be preserved in such a way that they have meaning in the future – that is, the original encoding of the object must be readable by future technologies and the meaning of the object in its original context must be understood. Otherwise, we will not be able to view our heritage. Or it may happen such as with many ancient scripts: they are preserved and visible, but no one knows what their meaning is. For example, without the Rosetta Stone, the meaning of Egyptian hieroglyphics would have remained unknown.

3.3 Activities and actors

With physical objects we usually assume that they have to be put somewhere safe and guarded, so that someone in the 25th century could open the door, get a volume that we have published in 2011 and admire our remarkable intelligence as we do with what has reached us from ancient Greece or Roman Empire. This is imagined as a quite passive process of safeguarding, though of course requires quite many steps of regulating the climate, inspecting publications for decay or mould, conservation and so on. But digital preservation seems to be defined in more active terms.

The question of actors is implicit in this issue: who should take these actions and have power of decision? How much action has to be taken and how often, for what and what kind of action?

We can consider the set of actions as constituting the ‘life cycle’ of the preservation process according to the following:

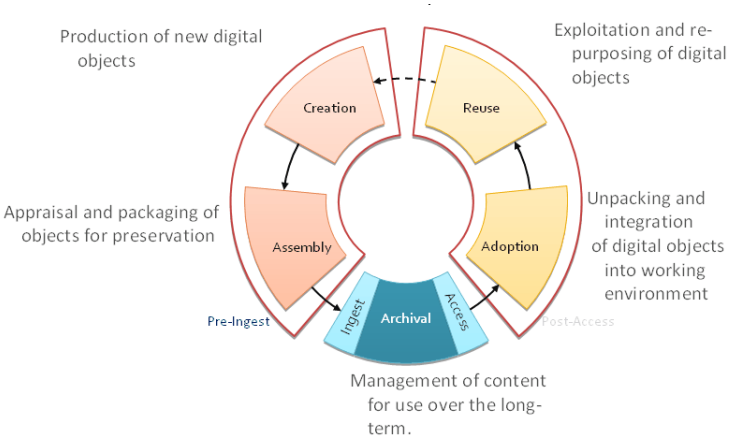

Fig. 1: Preservation life-cycle (source: SHAMAN 2009: 49)

The actors are involved in a whole life-cycle of the long-term digital preservation processes that begins with the production of the digital object (which may be any document, image, moving film etc.). This information has to be assembled in a package and archived (or ingested, in the technical jargon). The main actions for long-term preservation are taken within the archival system where the storage is organized, processes of migration are performed, more metadata added, checksums calculated, emulation mechanisms are implemented, access and discovery tools put in place, the rights of the users assigned and checked, and many other actions taken. This makes preservation different from archiving. Archiving basically involves secure storage of densely packed data, so when the data has to be accessed, it has to be withdrawn from the storage, unpackaged and prepared for the use preferably in the same environment in which it was used before, while in the preservation system, the ease of access and readiness of the data for re-use are the most important indicators of successful preservation. Thus, at some point in the future, the digital object may be retrieved from its store for re-use: in that case, the package must be unpacked and the digital object made available through the technology of the future. Re-use may involve exploitation of the same object for the original purpose or its adaptation for a new use. For example, a bureaucratic record may have been produced for legal purposes, but may be re-used for the purposes of historical research and education (Wilson 2011).

However, there are many actions that are entirely outside of the digital preservation technological life-cycle performed by very different actors.

Such actions start firstly with the creation of the digital object. At this particular moment, the appraisal of what will be preserved has to be determined. For proper preservation, it has to be appraised in terms of how long the preservation period should be,

what information about the digital object has to be sent into the future together with the object itself, so that it remains meaningful. This is never an easy decision to make, for how do we know what will be needed in the future, let alone what will be valued in the future?As we can see in our days, comic books are of great value among the collectors, but also among cultural historians. Leaflets and the ads, commercial and personal correspondence of the bygone times that were thrown out by the contemporaries are of highest research value and sought collectibles at present. So, should we preserve our e-mail messages, including spam, as they may become highly relevant for future researchers trying to make sense of our lives?

3.4 Technologies

The actions are clearly related to technologies. Technology is the only thing that is not mentioned in the definitions. It is so obvious that nothing digital can exist without digital technologies that there is no need to name it specifically. And who can be sure of what kind of technologies we will have in 20 years or even sooner? As now we know, it is technology itself or rather its development that causes the need for preservation of digital objects and, especially, the possibility of making sense of such objects. It is the element in the equation of digital preservation that attracts the biggest attention and financial resources, since it is quite easy to demonstrate that the problems caused by technology can also be solved by the understanding of the information carried by it. So, it is clearly a problematic element.

By now we have evolved methods and tools for digital preservation and have more or less successful projects, such as LOCKSS - Lots of Copies Keep Stuff Safe, migration of files to new formats and platforms (a variety of solutions are available for this), emulation of systems and software for accessing the old files. We are trying to employ a Virtual machine for accessing the bit stream directly and creating a preservation framework that will not depend on the technological platform and medium. I am sure that our software engineers will succeed sooner rather than later in this. There are specific discovery tools and tools that can automatically extract metadata from the objects for preservation. The grid and cloud computing are tested for storage and preservation of huge amounts of digital data for e-science. Even semantic and semiotic issues of records and the decision making for digital preservation processes are aided by a range of software tools and ontologies.

Accordingly, there is definitely no lack of ingenuity and knowledge to create several preservation systems for various needs.

4 LONG-TERM DIGITAL PRESERVATION IN SCHOLARLY AND SCIENTIFIC COMMUNICATION

Besides technology, all the other elements, though less obvious, are equally problematic. Let me demonstrate the problems with regard to the area that is under the scrutiny of the honourable participants of this conference – scholarly and scientific communication.

There are significant differences among various disciplines and even institutions with regard to the particularities of the communication processes, communication habits, traditions, norms and rules that we follow in publishing, sending e-mails and talking, attending conferences and writing projects. We are pushed into being uniform by the requirements of funding agencies and means of communication available, but we are still different in the very essence. And there is a great difference between science and humanities, which is very important in terms of digital preservation. It stems from the essence of knowledge generated in these two areas.

Humanities generate understanding of the humanity's nature – it is as complex as our human spirit and the social embodiment of our social relationships. We can study it from various perspectives, but humanities produce individual explanations that are neither correct nor erroneous as long as they contribute to the cognition of ourselves as humankind. Therefore, each contribution is only as good as it is subjective and original. Humanities' research is never cited to the same extent as physics or biotechnology articles, but it is used by the whole scientific and scholarly community.

Aric Hagberg is one of the main creators of the NetworkX software package. His team has developed this map. "Maps of science resulting from large-scale clickstream data provide a detailed, contemporary view of scientific activity and correct the underrepresentation of the social sciences and humanities that is commonly found in citation data." (Bollen et al. 2009).

The work of scholars of humanities and some social science explorers does not have to be and should not be repeatable as it addresses unique aspects in unique ways – the more unique, the better. As unique input into the understanding of human nature, each of these works is equally valuable, regardless of the time of its creation. Thus, the philosophy of Plato and Kant are as true and necessary as the thoughts of Habermas or Darlei Dall'Agnol. Even if they build upon each others' thoughts, they are unique. Does this mean that we should preserve all these works for eternity?

Enc. Bibli: R. Eletr. Bib. Ci. Inf., ISSN 1518-2924, Florianópolis, v. 17, n. esp. 2 – III SBCC, p.1-18, 2012.

On the other hand, we all can witness the inscription on the Scholar Google which says "Standing on the shoulders of giants". It actually implies that with each generation we can see further and further. However, to the full extent this is true only for science and technology, which depend on the accumulation and validation of research results, information and knowledge about the phenomena of this wonderful world. It cannot be faked (at least not in the long run), it must be testable (even if we start having huge problems with our amazing experiments), it has to be taken further from what we already know. This is a general explanation of the world that is constantly accumulating and enhancing, building on the previous knowledge to generate new knowledge. It may be the case that the old knowledge is discarded as erroneous after it has served for a while, but it is often incorporated into a fuller explanation as not erroneous, but only partially explaining a separate element or aspect of a bigger system of universe. One may say that this type of knowledge is very seldom lost altogether since one can always find the older knowledge built into the present system. So, it might seem that we do not have to worry about preservation in science and technology at all because this type of knowledge continues into the future through evolving explanations. At least this will not imply preservation for eternity, only as long as it will serve its purpose. But could this reasoning be not quite correct?

4.1 Findings from SHAMAN – requirements and needs of e-scholarship and e-science

While working on the SHAMAN project we had an opportunity to investigate the needs for digital preservation in different fields. We were collecting data on the preservation projects and practices of preservation in e-science and e-scholarship.

I must admit that I have been very astonished to find quite sophisticated digital preservation practices in many areas of humanities. It may be strange having in mind what I have said earlier about the nature of humanities, but, obviously, I was under the spell of stereotyping humanity scholars as afraid of computer technologies.

The projects were not huge due mainly to the lack of financing, but they had all the necessary elements inbuilt and were thought through to minute details. Sure enough they were carried out together with mathematicians, programmers, software engineers, and others, but involved humanities researchers in many aspects: as ideologists, requirement architects, data curators, system operators, not to mention those who were using the preserved material.

We have found proper preservation (not simply digitization or archiving systems) in humanities to my greatest amazement. They involved the sound recordings of various kinds

(digitization and preservation system), image and video recordings (digitization and preservation), language corpora databases (preservation and data-mining), scholarly e-editions (preservation and research), academic dictionaries of various languages (access, research, and preservation), historical archival documents (world-wide access and preservation), databanks of personalities (access, research and preservation) and many more. In addition, there were classifications and thesauri developed for different humanities' fields, such as ethnography, archaeology, history, or literature for accessing, appropriation and re-use of preserved digital materials. The need for preservation and re-use of the data was not disputed by anyone and the main concern was about who has the priority in using the materials.

In fact, these small but comprehensive projects were quite similar in complexity and goals to the ones that memory institutions were implementing.

During the requirements exploration phase it seemed that natural scientists, mathematicians and technology scientists were not greatly interested in preservation as long as the results of their investigations in the form of publications are accessible for peers for a certain period. In relation to the expenditure required to set a proper long-term preservation system, the outcomes of it did not seem to be absolutely necessary for them.

However, four years is a long period and we had an opportunity to witness quite different attitudes at the end of the project when conducting the evaluation of the project's outcomes with scientists. In this we worked with engineering and physics researchers who expressed now a strong need for preservation of various types of data for future re-use. The following objects had to be preserved according to the engineering researchers:

1. Original data, received from various sources, relating to various kinds of structures (e.g., dams and bridges), road accidents, water resources, and other sensor data.

2. Software programs used in the analysis of data, which needed to be maintained so that they could be re-run; for example, to enable new theories to be tested by modifying and re-running the software on the latest computing environment. At present, the possibilities were constrained by the need to maintain legacy systems, upon which such programs could be run and, consequently, a policy for the successful migration of programs from one computing environment to another was needed.

3. Finally, the results of analysis (and any intermediate data generated by the analysis) needed to be preserved for comparison with future work. (SHAMAN 2011).

The need for long-term preservation and the usefulness of the preserved data was indisputable. Some researchers had their own individual means of retaining the data from their previous activities and quite complicated storage and access means to it. It was a surprise

to us to some extent since just a few months before we conducted evaluation in industrial design and engineering company. There the long-term preservation was regarded as quite unnecessary activity considering the fact that the product information and documentation was retained only for 14 years after a certain product (like a TV set) had been stopped in production. A little more useful preservation systems were shown in the areas where the legal requirements for keeping certain types of information and documentation for longer periods were in place (medical products). But even in these cases, the main feature for the systems was that they should be "invisible" to the engineering staff, who did not want to use their time for the preservation process. The re-use of previous solutions was not seen as a sufficient justification to put significant resources into preservation systems.

In the area of fundamental physics (we did our evaluation in European research institutes), which is very different from the applied technology interests, we experienced similar attitudes. For example, one of the participants in our focus groups stated: "We don't have a digital preservation policy and it's what we need, so for us, anything we can learn is useful".

The scientists have raised a number of issues. First was the issue of international collaborations on huge amount of data from large-scale and extremely costly experiments (e.g., with the Large Hadron Collider at CERN). There was no need to preserve the original data on the location, since this was done through existing collaborative agreements. It still requires digital preservation, but by someone else. However, capturing the workflows of the experiments conducted by local researchers and preserving these, their associated data and software, was essential. Standards did exist but there was a lack of policy to enforce these standards across the different groups of researchers on the location.

Regarding the local research, the situation is similar to that of researchers in engineering: individual researchers have a great deal of autonomy over how they conducted their experiments. The experiments tended to consist of running programs they had developed themselves and what they did with the original data, the software and the intermediate data and results. Several obsolete machines were retained, although powered down, in the event that proprietary data format needed to be re-used. Re-use of old data or old programs arose when, for example, a new PhD student found a need to apply new theory to existing data analyses. This might involve revising the original program to explore new parameters in the existing data.

The long-term use was not disputed, but different elements involved in the experiments had different long-term value.

We were very glad to receive the confirmation that the efforts of the SHAMAN were not in vain and got the answers to project-related questions. On the other hand, many of us engaged in this work started asking much bigger questions, such as the following:

- Is it possible to preserve everything?
- Is it necessary to preserve everything?
- How long can we power our technology?
- Who should be responsible for the preservation of our record?
- What to preserve and how to select objects worthy of preservation?

5 THE BIG QUESTIONS OF DIGITAL PRESERVATION

5.1 Is it possible to preserve everything?

Let us have a look at our amazing digital universe that we have created so far. The world's information is doubling every two years. In 2008, companies' servers have processed 9,75 zetabytes of information (Short et al. 2011). By 2020, the world will generate 50 times the amount of information and 75 times the number of "information containers" while IT staff to manage it will grow less than 1.5 times. New "information taming" technologies such as deduplication, compression, and analyses tools are driving down the cost of creating, capturing, managing, and storing information to one-sixth the cost in 2011 in comparison to 2005 (Ganz and Reisel 2011).

Data creation outpaces the creation of storage. Obviously, there is much data that is just ephemeral - which disappears at the moment of production (like digital TV signals). But what one should do with the rest?

We are actually very eager to have the data for science and scholarly purposes, so we can make wonderful discoveries based upon such data. One can point out bibliometric research done by Olivier Beauchesne to illustrate the collaboration among scientists and scholars in the world. Using the Facebook friendship map by Paul Butler, Olivier Beauchesne at Science-Metrix examined scientific collaboration around the world from 2005 to 2009:

"From an extensive database of academic citations I extracted and aggregated scientific collaboration between cities all over the world. … if a UC Los Angeles researcher

published a paper with a colleague at the University of Tokyo, this would create an instance of collaboration between Los Angeles and Tokyo." (Beauchesne 2011).

Imagine what kind of analysis we could do with long-term data preserved over long periods. So, possible or not, we would like to preserve as much as possible our significant data. But

5.2 Is it really necessary to preserve everything?

We know quite well the treasures that have reached us in various odd ways. We have an idea of what we have lost by what was mentioned in other works, such as some ancient Greek dramas cited in Aristotle's Poetics. But we have no idea about what we have lost, neither how much nor how good. And we seem to be doing quite well without the lost heritage as Eco and Carriére (2009) argue in their discussion on books. On the other hand, as Oscar Wilde has noted in the 19th century: "In old days books were written by men of letters and read by the public. Nowadays books are written by the public and read by nobody."

The situation seems to persist.

If there are no documents left but we still find something outstanding, such as the amazing engineering of ancient Incas and Mayas, we stubbornly work on the riddle till we get viable answers. So, maybe we should not bother with preservation at all? In regard to the selection for preservation we also have to think on who are those deciding what and how much should be preserved. Should we leave the decisions in the hands of experts as is common in archival and library practice for physical heritage or should we involve the members of the public as well? The experts are proved to be wrong or misguided at least in some cases. One just has to remember the loss of priceless American and English newspaper collections in the Library of Congress and British Library criticised by Nicholson Baker (2001) in his famous book Double Fold.

Even if we solve the storage and selection issue, the next question remains.

5.3 Will we be able to power our technology?

We will never be able to access the stored information without technology as we can do with a paper book. And technology needs energy, which turns the long-term access to

digital information into a different kind of problem altogether as we are rapidly exhausting energy resources on Earth, and it is not quite clear how much of it we will be able to produce from renewable sources.

5.4 What should we preserve and who should be responsible for the preservation?

At present, funders, scholars and scientists, librarians and technology specialists, politicians and governors definitely agree that we have to engage in digital preservation. And I agree with them completely. We should do so for our own sake, because we are not able to use anymore our old files from 10 years ago; we may lose the materials that we need to have within 20 years from now. We are at the very beginning of the road, but by now we have come up with the idea that digital preservation is a matter of policy more than of technology. All the questions that I have posed previously should be answered by us in the high level policies of states and organizations by organizations and individuals, state and private bodies involved in digital preservation. We need to make sense of what we will be preserving and find the ways to represent this sense to our future generations and to our future selves as well. We also have to decide to whom will be given the responsibility of the long-term digital preservation, though the answer seems to be quite obvious here – Google and the like who are heavily involved in digitization, have legal settlements with authors and publishers and enough capital to implement the ventures of amazing size. Or maybe not?

The Harvard chief librarian Robert Darnton who is also a known figure in historical research and publishing is absolutely convinced that it is only in academic libraries that we can trust to take care of our intellectual heritage in the long run. And it has not so much to do with the ability and experience of these institutions, but more with their longevity, which is a significant factor for preservation anyway.

Organizational longevity refers to the durability or continuance of organizations. The average life span of a multinational corporation is approximately of 40 to 50 years, with many companies failing or being absorbed in takeovers and mergers within the first 15 to 20 years. This figure is based on most surveys of corporate births and deaths. There are some exceptions, of course. The Stora company, for example, is a major paper, pulp, and chemical manufacturer; it has had the character of a publicly owned company from its very early beginnings, more than 700 years ago, as a copper mine in central Sweden. The Sumitomo Group has its origins in a copper casting shop founded by Riemon Soga in the year of 1590. Examples like these are enough to suggest that the natural average lifespan of a corporation

should be as long as two or three centuries (de Geus 2002). But this is far less than the average life of big research libraries that, though having tendency to burn or change under hard pressures, keep going on from the moment of creation.

So Robert Darnton continues: “Companies decline rapidly in the fast-changing environment of electronic technology. Google may disappear or be eclipsed by an even greater technology which could make its data base as outdated and inaccessible as many of our old floppy disks and CD-ROMs. Electronic enterprises come and go. Research libraries last for centuries. Better fortify them than to declare them obsolete, because obsolescence is built into the electronic media.

Google will make mistakes... Bits become degraded over time. Documents may get lost in cyberspace... Hardware and software become extinct at a distressing rate. Unless the vexatious problem of digital preservation is solved, all texts “born digital” belong to an endangered species. The obsession with developing new media has inhibited efforts to preserve the old. We have lost 80% of all silent films and 50% of all films made before World War II. Nothing preserves texts better than ink imbedded in paper… The best preservation system ever invented was the old-fashioned, pre-modern book.” (Darnton 2009: 36-37)

Umberto Eco echoes the thoughts of Harvard librarian: “We have seen that modern media ages fast. Is it worth risking overloading ourselves with objects that most probably will become unreadable and impenetrable? We have enough scientific proof that paper books are superior to any other object that our cultural industries have supplied on the market lately. If I had to choose something easily transported and something that has already proved its ability to withstand the threat of time, I would choose a book.” (Carriére and Eco 2009: 69)

I realise that I have chosen humanities scholars and for some of you their conclusions might sound conservative and stubborn to remain faithful to ancient times.

But in regard to such position I also have support of the founder of the Internet Archive, Brewster Kahle, who decided to back up the digital versions of books by creating a physical archive of books (Internet... 2011) after becoming concerned about the number of books being pulped after digitization. Therefore, paper is considered to be an adequate back up for digital preservation systems.

6 THE ENDING

After demonstrating all these questions and presenting positions of well known scholars, I may have become somewhat wiser but I still do not know answers. Despite this fact, I am quite sure that there are many people looking for them and there are others who are trying to find technical and organizational solutions. After all, this is what actually used to make sense – to start looking for the way to approach the aim.

I would like to round up as I started. There is after all an author who talks of long-term preservation or existence of books in eternity. This is Jorge Luis Borges who writes in "The library of Babel":

> Perhaps my old age and fearfulness deceive me, but I suspect that the human species – the unique species – is about to be extinguished, but the Library will endure: illuminated, solitary, infinite, perfectly motionless, equipped with precious volumes, useless, incorruptible, secret. (Jorge Luis Borges. The Library of Babylon).

Preservation for no use – a sad perspective and the only one that we should avoid.

REFERENCES

Baker N. (2001). Double Fold: Libraries and the Assault on Paper. London: Random House Publishing.

Beagrie N. and Jones R. (2001) Preservation Management of Digital Materials: A Handbook, London: The British Library. Retrieved 13 June, 2012 from www.dpconline.org/graphics/handbook/

Beauchesne, O.H. (2011). "Map of scientific collaboration between researchers", 23/01/2011, accessed on 11/04/2011. Retrieved 4 July, 2012 from http://olihb.com/2011/01/23/map-of-scientific-collaboration-between-researchers/http://olihb.com/2011/01/23/map-of-scientific-collaboration-between-researchers/

Bollen J., Balakireva L., Bettencourt L., Chute R., Hagberg A, Rodriguez M.A. and Sompel H. van de. (2009). Clickstream data yields high-resolution maps of science. PloSOne 4(3) : 1-11.
http://www.plosone.org/article/info:doi/10.1371/journal.pone.0004803http://www.plosone.org/article/info:doi/10.1371/journal.pone.0004803

Carriére J.-C. et Eco U. (2009) N'ésperez pas vous de débarraser de livres / entretiens menée par J.-P de Tonnac. Paris : Grasset.

Darnton R. (2009) The case for books: past, present and future. New York: Public Affairs.

de Geus, Arie (2002) The living company: Habits for survival in a turbulent business environment. Harvard: Harvard Business Press Books.
http://www.businessweek.com/chapter/degeus.htmhttp://www.businessweek.com/chapter/degeus.htm

Ganz J. and Reinsel D. (2011) Extracting value from chaos. IDC.

Gunn, J. (s.d.). Libraries in science fiction. Retrieved 12 May, 2012 from http://www.sfcenter.ku.edu/library.htmhttp://www.sfcenter.ku.edu/library.htm

Harvey D.R. (2005). Preserving digital materials. Munchen: K.G. Saur Verlag GmbH.

Maceviciute E. and Wilson T. (2011). Evaluating the SHAMAN framework by memory and industrial engineering institutions. In Proceedings of Mötesplats Borås Profession- Forskning, 19-20 oktober 2011. Borås: University of Borås.

Shaman (2009). D2.3 specification of the shaman reference architecture: Shaman-wp2-d2.3.

Shaman (2011). D. 14.4 Report on evaluation and demonstration for e-science domain of focus.

Short J.E., Bohn R.E, and Baru C. (2011). HMI? How much information: Enterprise server information. La Jola (CA): UCSanDiego.
http://hmi.ucsd.edu/pdf/HMI_2010_EnterpriseReport_Jan_2011.pdfhttp://hmi.ucsd.edu/pdf/HMI_2010_EnterpriseReport_Jan_2011.pdf

UNESCO (2003). Charter on the preservation of digital heritage. Retrieved on 15 June, 2012 from http://portal.unesco.org/en/ev.php-URL_ID=17721&URL_DO=DO_TOPIC&URL_SECTION=201.html

Wilson T. (2011). Preservation – the final frontier? Retrieved 10 May, 2012, from http://informationr.net/tdw/publ/papers/ASIST2011.htmlhttp://informationr.net/tdw/publ/papers/ASIST2011.html