Digital Preservation – Challenges and Opportunities

Author:
Börje Justrell

CIDOC06
GOTHENBURG
SWEDEN

Introduction

This paper examines some main challenges and opportunities connected to long-term preservation of digital information. From an archivist's point of view, there are no solutions available today that can technically preserve digital resources over time with minimum loss or decay and, at the same time, keep their accessibility and usability untouched. Therefore, long-term digital preservation is a focal area of research and development in the archival community. But, the question of preserving information in binary form has a wider range then that: today's digitised society makes long-term digital preservation a concern for the whole cultural heritage sector, archives as well as libraries and museums.

It is easy to paint the future of digital preservation in black; the line of foreseers of judgement day for digital resources is of impressing length. I will briefly touch the most troublesome challenges in keeping digital information accessible and usable over time. They are all well-known, but the purpose of mention them is to set the scenario.

Computers have been used in state administration for decades, in Sweden since the 1960ths and in other Western countries for a longer period than that. So, there are evidently lessons learned to benefit from in long-term digital preservation. I will look at some of them, mainly those made by the Swedish National Archives, but my focus in this paper is, however, on opportunities. The digital world provides new conditions compared with the analogue world, which sets new landmarks for the cultural heritage institutions when they use and preserve digital resources. These landmarks need to be identified and disseminated.

Challenges

The most important challenge is time. More and more advanced computer technology is introduced on the market. The road map for bringing out new models of personal computers and laptops is today built on intervals of 15 months. Up graded versions of magnetic tapes for storage

– normally with doubled capacity for each generation – reach the market in cycles of about 1.5 year. These are examples of time frames set not by technical reasons, but to balance the economical and administrative willingness of the customers to renew their technical infrastructure. The short time span between new generations of hard- and software put a heavy pressure on those who have to preserve the digital information long-term, especially when safeguarding digital resources for the future is identified and accepted as a vital part of a sustainable knowledge society. The cultural heritage sector is today faced with growing demands among citizens, that to some extent involve a need to make current and future digital information accessible and usable over time: they claim for improved participation in societal processes, better quality and transparency in cultural and societal experiences, development of knowledge on a personal and societal level, and last but not least cultural resources for research. Matters are not simplified by the fact that society are in the beginning of a digital revolution that will affect every inch of it and with effects that we can not foresee today. Professor Luciana Duranti at the University of British Columbia defined already in 1999 the core of the situation:" The last decade and a half has produced more records than any previous similar period of human activity. The fact, that the majority of these records is less reliable, retrievable or accessible than ever before, is one of the ironies of the modern information age".

There is just one conclusion to draw: if we do not manage to handle digital resources properly over time, there will be no records of today or tomorrow left for coming generations. We will get a black hole in the history book.

Lessons Learned

During the last two decades, digital preservation has moved from finding the ideal long-term storage media to weighting the advantages and risks of different strategies, and to define practical solutions based on standards that may use several available digital preservation strategies concurrently. It is also an established truth today that digital preservation begins early in the life-cycle of digital records. The complexity of these records and their environment makes it necessary to involve archival procedures at an early stage in records creation in order to safe guard

possibilities for preservation and future accessibility and usability. Digital archiving has to be defined as a chain of actions, starting with records creators handling metadata and appraisal issues when still running their systems, and ending with archival institutions taking care of digital systems transferred to them to be accessed by present and future end-users.

And once in the archives, what happened there? The first thirty years use of digital technology has taught the archivists that digital information grows, lives and dies, and that the task of keeping this information alive requires frequent adaptations to and perpetual evolution of the digital repository systems. Therefore, rather than making temporary plans in the hopes of a future permanent solution, it would be prudent to think of long-term digital preservation as an evolving, ever-changing "ecosystem". What is needed is, in other words, a permanent strategy for handling perpetual changes …! But, we also need a clear vision why we are preserving digital resources for the future. For the Swedish state archives, the current vision for digital preservation is to guarantee long term accessibility and usability in order to support transparency in societal processes, cultural heritage, and enhanced knowledge (i.e. lifelong learning).

On practical level, the working procedures will change in cultural institutions performing digital archiving. With a rapidly growing amount of digital information to take care of and an increasing speed of data transfer to digital archives, solutions for preserving digital information must have a higher level of automation and self-reliance than solutions for traditional paper records. The diversity in sizes and complexity of the institutional and organisational circumstances among records creators also imply that the digital repository systems and the administrative procedures surrounding them must be highly scalable and adaptable to various levels of input, storage and access. Some countries, for example Estonia, has drawn the conclusion that what is needed is one single digital repository on national level covering all cultural heritage institutions, archives, libraries and museums. The Swedish National Archives has proposed to the Ministry of Education and Culture that this solution ought to be looked at also in Sweden.

Opportunities

eGovernment

eGovernment has given records and archives management a new role in the interaction between citizen and society. The former Swedish state commission on IT declared that a corner stone in a successful eGovernment process is the citizen's confidence in the technical applications.

From an archival point of view confidence in this sense has two dimensions: transparency and trustworthiness. If citizens do not trust the authorities' computer systems, they will for sure avoid eGovernment services and instead rely on paper based records. But, if they could verify trustworthiness by viewing all activities and all information in case files, their attitude should maybe be different. A high degree of transparency in computer based solutions will probably increase the public's willingness to use eGovernment services. What the authorities need to focus on in their eGovernment work is, basically, classic records and archives management issues: how to guarantee authenticity of digital resources over time, how to physically and technically preserve electronic documents over time, and how to retrieve electronic documents and make them accessible and usable over time. In some countries, for example Denmark, United Kingdom and US, the National Archives have got the money to solve these tasks.

Cost savings

Some parts of society are still acting in the traditional analogue world; some minor parts have already entered the digital world. But most parts are acting in between, which is very expensive. Computers have rationalised the administrative work profoundly, but once the society started to define its activities as e-actions (eBusiness, eGovernment, eLearning, eCommunications and so on), computers were degraded from being the solution to be just building-bricks in a huge digital construction, joint together by information in digital form. And if the information needed isn't in digital form, it has to be converted, which takes time and costs money. Reuse of information – sometimes several computer generations old – is getting more and more important. Software applications built on reuse of information are already in place (data mining, data marts, Artificial

Intelligence etc), and storage media – today hard drives – are cheaper than ever before and continuously equipped with higher capacity. But, this is not enough; reuse of digital information means also – as mentioned earlier - to guarantee authenticity over time and to make the information accessible and usable also in the future. Today every organisation has to build its own solution for that. If the Swedish National Archives, like in other countries, will get the resources to develop general models, methods and tools for digital preservation in the public sector, state agencies and municipalities will save money, probably in total hundreds of millions of SEK.

The professional toolbox for digital preservation

It is obvious that the curators in digital archiving today have a toolbox for their professional work that covers at least the most basic needs. I will give some examples from the archival world:

In the 1990s the professional international organisation for archivists, ICA (International Council on Archives) finished a several year work on metadata standards, covering the description of records and the information about the records creators. These standards, the International Standard on General Archival Description ISAD (G) and the International Standard on Archival Authority Records for Cooperative bodies, Persons, and Families ISAAR(CPF), are used by archives all over the world. Today, there is also data exchange formats adapted to these metadata standards: Encoded Archival Description (EAD) to ISAD (G) and Encoded Archival Context (EAC) to ISAAR (CPF).

For issues regarding system structures several de facto standards are in use today, for example MoReq (Model Requirements for Management of Electronic Records) and the British Functional Requirements for Electronic Records Management Systems.[1]

International Standardisation Organisation (ISO) has settled a number of standards of importance in long-term digital preservation, like on operational level the records management

[1] http://www.nationalarchives.gov.uk/electronicrecords/reqs2002

standard ISO 15489, on conceptual level the reference model for an open archival information system ISO 14721, and on format level the electronic document file format PDF/A for long-term preservation ISO 19005-1:2005.

And there are more to come.

Enhanced access through digitisation

In most Western countries, a rapidly growing part of the cultural heritage is in digital form. Earlier, the digital archives contained mostly born digital documents, books, newspapers and pictures, but today analogue materials converted to digital form through digitisation is the major part. A few years back digitisation was often looked upon as something less cultural and more of a play ground for those interested in computer technology. Today, it is a consensus among the cultural heritage institutions that digitisation is essential to improve accessibility and usability of their holdings and collections. It is also accepted that some basic conditions must be fulfilled: the digital information has to be easy to find, easy to retrieve, and easy to understand – even cross-sectorial; and it need to be retrievable, accessible and usable not only on short term but also in the future.

On European level, the Council in 2000 approved the eEurope 2002 Action Plan which stated that digitisation is strategic for the preservation of the European cultural heritage and for the possibilities to make it more accessible. During the Swedish presidency in 2001, the basic document for the coordination of digitisation, the Lund Principles, was agreed.[2] Since then, a group of expert, the National Representatives Group (NRG), representing all Member States in the European Union, meet twice a year to coordinate national activities in digitisation. In November 2005 the Council approved a new action plan for the coordination on digitisation, called the Dynamic Action Plan, where the Council underpin the strategic importance of digitising the cultural and scientific heritage for enabling digital access by all citizens to national, regional and local cultural heritage of Europe, but also for providing rich and diverse digital

[2] http://www.cordis.lu/ist/ka3/digicult/lund_principles.htm

resources that support education and research, tourism and the creative industries. One of the action areas in the Dynamic Action Plan is digital preservation, coordinated by a working group led by Sweden and Estonia.

In Sweden, the Governments strategy group for IT policies has in a report also pointed out the possibilities of using the digital cultural heritage resources in economic development and growth.

A new huge initiative by the European Commission is the European Digital Library with digitisation, internet access and digital preservation as the three main corner stones. This initiative will be followed up with funding for research and development through the coming 7th EU frame work programme. Already in the ongoing 6th frame work programme digital preservation has been successfully targeted in at least one call.

Research

It is an urgent matter for cultural heritage institutions to coordinate their activities in digital preservation. But to get hold of the situation, major actors in this sector need to join their forces also with IT-research and industry partners to develop new strategic models, methods and tools for collecting, assessing, preserving and providing access to digital resources. And such initiatives are already in the air! In 2003 the Swedish National Archives started a joint research and development project together with Lulea University of Technology and the municipality of Boden. The objectives for this project are, on short-term, to develop methods and tools for long-term preservation of digital information in the public sector. The Swedish social welfare administration has been used as the main pilot for both development activities and scientific research, and the results will be published in the end of 2006 when the project period is completed. On long-term, the goal for the project is to create a national competence centre in Sweden on long-term digital preservation and access. This is now fulfilled. In January 2006 Lulea University of Technology established, as a part of the university, a centre for long-term preservation. Partners in this centre are, besides the university, the Swedish National Archives and the municipality of Boden. During this spring it has been decided that also the Swedish National

Library and the Swedish National Archive for Sound and Moving Images will take partnership in the centre. Three main areas of activities have so far been identified in the field of long-term preservation: scientific research, development of models, methods and tools, and knowledge enhancement. On a more practical level two main questions will probably dominate the agenda: (1) the availability and accessibility of multi-sourced and multi-formatted digital resources and (2) the preservation of complex, dynamic and very high volume digital objects, including those with high levels of interactivity.

Some Conclusions

1. It is possible to avoid falling into the black hole of history mentioned in the introduction. But, it will be costly. And there are no short cuts – just a lot of sweat, tears and money to spend.

2. Long-term digital preservation is today an alarm clock which sends its signals far outside the cultural heritage sector. Solutions for preserving digital resources are needed to safeguard eGovernment, otherwise the whole concept are on risk to be embarrassingly down-sized or amputated to survive. In some countries the government has already realised that and decided to rise the funding needed. Other countries will surely reach the same level of understanding. It is just a matter of time!

3. The transformation from an analogue society to a digital one is very expensive, especially making digital information accessible and usable across time, technology and organisational boarders. The cultural heritage sector, with decades of experiences in preserving digital information and with a fresh start-up toolbox on hand, can help shortening both the time frame and the costs. But, so far, just a few institutions have realised that. Old perspectives on information management apparently die hard.

4. A big pusher for expanding activities in digital preservation has entered the arena - digitisation. The expectations on increased accessibility and usability through digitisation are high both among citizens and political decision makers. The initiatives on European level for enhanced and

coordinated digitisation will put a political pressure on the national governments to take actions. And it is not easy for them to escape, especially when millions of EURO is available in European programmes for research and development.

5. Coordinated and targeted research and development activities in long-term digital preservation have still a short history. Institutions with experiences in solving permanence and longevity issues for different types of digital resources are just in the beginning of finding each other as partners: archives, libraries, museums, audio-visual institutions, and industrial bodies. Creating a competence centre like the one on long-term preservation at Lulea University of Technology is probably the shortest and most fruitful way to get things to happen on national level. The cultural heritage sector needs a national green house for improving new ideas on digital preservation and for cultivating the seed of solutions for the future.