European Commission

INSAR
INformation Summary on ARchives

EUROPEAN ARCHIVES NEWS

SUPPLEMENT IV (2000)

Proceedings of the DLM-Forum on electronic records

European citizens and electronic information: the memory of the Information Society

DLM-FORUM
Electronic Records

Brussels, 18-19 October 1999

- This publication contains the full texts in English and abstracts in German and French. These abstracts are placed directly after the full texts.
- The full texts for all contributions, in German, English and French, can be found on the DLM-Forum website: http://www.dlmforum.eu.org

- Diese Veröffentlichung enthält die vollständigen Texte in Englisch und Zusammenfassungen in Deutsch und Französisch. Die Zusammenfassungen befinden sich direkt hinter den vollständigen Texten.
- Die vollständigen Texte aller Beiträge sind in deutscher, englischer und französischer Sprache auf der DLM-Forum Website verfügbar: http://www.dlm-forum.eu.org

- Cette publication contient les textes complets en anglais et des résumés en allemand et en français. Les résumés se trouvent directement après les textes complets.
- Vous pouvez trouver les textes complets de toutes les contributions, en allemand, en anglais et en français, sur le site web du DLM-Forum: http://www.dlmforum.eu.org

EUROPEAN
ARCHIVES NEWS
INSAR

- Secretariat-General of the European Commission

- Committee of preparation:
Lino FACCO
Hans HOFMANN
Christina BECKERS
Peter BERNINGER
Jef SCHRAM
Johanna MAAS

- For further information and a free copy of INSAR, please contact:
European Commission
Tel. (32-2) 29-56810
(32-2) 29-56721
Fax (32-2) 29-61095
E-mail: archis@cec.eu.int
Address: rue de la Loi 200
Office: JECL 3/40
B-1049 Brussels
BELGIUM

Proceedings

of the DLM-Forum

on electronic records

European citizens and electronic information: the memory of the Information Society

Brussels, 18-19 October 1999

A great deal of additional information on the European Union is available on the Internet.
It can be accessed through the Europa server (http://europa.eu.int).

Cataloguing data can be found at the end of this publication.

Luxembourg: Office for Official Publications of the European Communities, 2000

ISBN 92-828-8806-1

Printed in Belgium

Johannes Hofman

Johannes Hofman (1948) has studied history, archival theory and informatics and he is working at the National Archives of the Netherlands in The Hague. During the last years he was involved in several projects concerning the management of electronic records within the National Archives and within the archival community in the Netherlands. Furthermore he fulfilled different functions in the field of archival automation within the National Archives and as such he was responsible for the development of information systems and the realisation of new strategies based on information technology.
The first project on electronic records in the Netherlands was initiated by a co-operative which has as participants apart form the National Archives, the Municipal Archives of Amsterdam, The Hague, Rotterdam and Utrecht, as well as the Royal Society of Archivists in the Netherlands. During the years 1991-1995 this project had as an objective the definition of strategies in managing electronic records. From the end of 1995 until 1997 he was director of a project, set up for the realisation of a technical infrastructure to preserve and control electronic records for archives.
In his present position he is working at the National Archives of the Netherlands and senior consultant for the government program 'Digital Longevity'. This program is initiated by both the ministry of the Interior and the ministry of Education, Culture and Science, with the objective of creating appropriate conditions for electronic record keeping

'Shooting at a moving target'. The development of a repository for the preservation of digital information

Johannes Hofman

1. Introduction

For centuries much of human knowledge and history has been documented and passed on to following generations on stone, clay and paper. The emergence of computers a few decades ago suddenly offered us a new and promising opportunity to process, to store and to disseminate information. The world wide web for instance provides us with a powerful tool to disseminate and access an enormous amount of information no matter where it is and without any time constraints. The increasing speed of developments in society and of our way of life seems to influence also the way we deal with information. Information seems to be omnipresent, and if available on digital media it tends to become more and more volatile and without any permanence. That is reinforced by the rapid technological developments in information and communication technology (ICT). One of the consequences is the difficulty of preserving digital information through time. With the increasing use of ICT however the awareness of the volatile character of digital information and the need to preserve it more permanently are being raised in particular and first within those disciplines or communities that are charged since long time with the task of preserving information through time, librarians, archivists and museum-curators. Libraries try to preserve electronic publications and websites, archives digital records, broadcasting and audio-visual organisations digital videos and films, and museums digital objects. As one of the memory organisations the National Archives of the Netherlands has until now preserved paper records, but is now confronted with digital information or records with which it is not familiar, but for which it is responsible too.

World-wide many initiatives exist at the moment to deal with issues around long-term preservation of digital information and to come up with solutions. Examples are the European NEDLIB project aiming at a repository for electronic publications, the English CEDARS-project, the US collection-based long-term preservation project of NARA in collaboration with the San Diego Supercomputer Centre, the (more theoretical) international Inter PARES reset project, the emulation project of the University of Michigan in co-operation with the (English) CEDARS project, and also the Dutch Digital Repository project.([1]) All these projects have in common that they try to solve specific issues around the long-term access of digital objects, be it electronic publications, videos, digital records, or anything else. However, there exist no real solutions yet, only best practices.

In this paper I will explain what is going on in the Netherlands in this field. Apart from that I will try to elaborate on the possible European dimension of the Dutch projects.

2. The issue of digital preservation

We experience the impact of ICT on society and our every day life. Still we do not know yet what the extent of it will be. One of the areas that is influenced is our way of recording information. ICT has changed for instance the nature of records that are created in government organisations. They are no longer fixed and physical entities, but fragile, volatile and intangible 'things'. How can we preserve them through time?

As said the continuously changing information technology, the variety of existing software and hardware platforms, combined with the lack of standardisation make long-term preservation of digital information (or more generic, digital objects) almost impossible.

The changed nature of records has made us also aware of the necessity to articulate more explicitly what records actually are, how to describe the context in which they are created and used, what they consist of, and what of them is essential and should be preserved. It is also clear that the existing rules and procedures concerning record keeping and archival management are no longer adequate in a digital world.

(1) CEDARS-project: CURL Exemplars in Digital Archives, see www.leeds.ac.uk/cedars/. NARA-project: see www.sdsc.edu/NARA NEDLIB project: www.konbib.nl/coop/nedlib/ Inter Pares project: www.interpares.org Further reading: *Digital Culture: maximising the nation's investment, a synthesis of JISC/NPO studies on the preservation of electronic materials*, ed. by Mary Feeney, London 1999. This booklet gives an good overview of the issues around digital preservation.

Moreover, these questions can only be answered when we understand why records are created, used and kept. One of the reasons, given for instance by Sue McKemmish, is to construct identity, not only of an individual, but also of society.([2]) Other reasons are the role records play in governance, in enabling accountability (political, juridical, social, cultural, historical, etc.), in constituting collective memory, providing evidence and as source of value-added information.

Answers to all these questions explain why we want to preserve records and enable us to identify the requirements for preservation of records or digital objects through time. In this respect it is crucial to distinguish between the intellectual and the technological dimensions of records or digital objects. We are actually dealing with two different categories of objects, the intellectual or logical objects such as records or electronic publications and physical or technical objects such as computer files. They both require their own management and maintenance and it is essential to be aware of this distinction while searching for solutions. What we want to preserve is the record as an intellectual or logical object, not as a physical entity. Technology is used only as a ('physical') means of expressing and recording intellectual things or objects. Nevertheless this physical carrier or medium (computer file and associated hard- and software) requires its own management.

Intellectually it is necessary to describe the objects (e.g. records) in digital form, to know what they are and that they exist, and to be able to handle them. Description offers stability that is needed in the ever evolving digital world. In this respect a new phenomenon seems to have emerged that of metadata, a somewhat mysterious term that encompasses a myriad of data about the objects we want to preserve. It is information we need in order to identify records, to establish their authenticity, to manage them, to make them available, to retrieve them etc. These metadata are not new and exist to a certain extent also in a paper world. The virtual character of digital objects however requires that they are more explicitly described.

With respect to technology it is necessary to find out and identify what preservation strategies can deal with the continuous changes and developments in information technology taking into account the preservation requirements. Although paper still plays an important role in preserving information, it will be ever more difficult to use paper in this respect as ICT will be developing. Since we cannot predict future developments, these strategies should be as soft- and hardware independent as possible. In developing a persistent digital repository or preservation function for digital objects this is essential.

I will now focus on the developments in the Netherlands especially how the National Archives is dealing with this issue.

3.The Dutch 'Digital Repository' project

Since 1991 thinking on the issues raised by electronic records has been an ongoing activity in the Netherlands. Several reports have been published. These reports and the accompanying activities have contributed to a growing awareness among senior managers of government organisations, records managers, responsible authorities, and, although to a lesser extent, in archival institutions, of the consequences of the use of ICT on record keeping and subsequently of the challenges they face.

It has in 1996 lead to the establishment of the joint program called 'Digital Longevity' in which both the Ministry of the Interior and the National Archives participate. The main objective of this program is to co-ordinate the different projects in the field of electronic records. This concerns among others projects on record keeping requirements, on the development of regulations and guidelines, and on the development of a digital repository.

The approach adopted in this program is to take into account all activities around (digital) records and record keeping, and to consider them as a coherent set, from the creation of records in a government organisation until their permanent maintenance within archival institutions. Although this could be seen as records continuum thinking, I am afraid it is not (yet) experienced in the Netherlands in the same way as in Australia. What is similar however, is the notion that it is no longer possible to use the life-cycle approach with different and discrete stages, but that it is necessary to adopt the idea that record keeping in all its aspects (including long-term preservation) should be seen as one coherent and continuous process and should therefore be based on one concept. As such the program aims at the whole government on all levels, including the provinces and municipalities.

within government organisations and archival institutions.
In this position he is among other things involved in establishing government regulations for digital records and in formulating strategies for digital preservation.
On the international scene he is (since 1993) member of the Committee on electronic and other current records of the International Council on Archives.
Within the European Union he is member of the DLM-Monitoring Group and as such involved in activities to stimulate co-operation in the field of electronic records. This committee is established as a follow-up activity of the DLM Forum on electronic records, which has been held in December 1996 in Brussels.
Furthermore he is associated with the 5th framework for research, especially the Information Society Technologies (IST) program, of the European Union, as expert and reviewer.
In 1998 he conducted a peer-review for the Swedish National Archives on its strategies for electronic records.
Finally he is as investigator and representative of the National Archives of the Netherlands participating in an international research project initiated by the University of British Columbia (UBC), called the Inter Pares research project, which has as objective to investigate the long term preservation of the integrity of electronic records.

(2) Sue McKemmish, *Evidence of me..*, in: Archives & Manuscripts, Vol 24, No 1, May 1996, p.28-45.

SHOOTING AT A MOVING TARGET

The development of a digital repository

Hans Hofman

National Archives of the Netherlands

One of the main themes in this program is the quest for solutions concerning preservation of digital information or records. As a consequence of the adopted approach this is not seen as a problem for archival institutions only, but also as an aspect of record keeping within government organisations. Digital records require always measures for maintenance, even when they have to be preserved for shorter periods, such as 5 or 10 years.

The idea of the Dutch program is of course not to solve this world-wide problem on its own, but to try to contribute to a solution through research in this field and to acquire at least some experience. Based on this experience it is the intention to identify viable strategies, and to come up with solutions for the short term until better solutions will be found.

In this context a project on digital preservation, called 'Digital Repository' project, has been established. The objectives of the project are:

- to acquire concrete experience with managing and preserving digital records
- to acquire knowledge and insight in what preservation strategies exist and how they can be used for what types of digital records.
- to raise awareness among records managers in government organisations and archivists within the National Archives and show them what is needed in this respect
- to achieve a political strand and show that something concrete has been done. This is important to get the necessary and additional funding that is needed.

The target behind it is *the long-term preservation of digital records or information in an authentic, usable and understandable way*.

After some preliminary studies on digital preservation the current project for the development of a digital repository started in November 1998 with the first phase. The objective was (and still is) to do something practical and to design and subsequently to build something that could serve as a first digital repository, where digital records can be preserved. The point is of course how do you do such a thing? In order to get an overview of the state of the art concerning available preservation strategies, to establish the boundaries of the project, and to make a global design of a preservation system, RAND Europe was asked to conduct a study.

This report, which is available on Internet, provided us with an overview of the existing preservation strategies, a model for a generic preservation process of digital records, a strategy for preserving digital records, and a proposal how to proceed.([3]) Based on discussions around the issue of authenticity of (digital) records and what should be preserved, one of the conclusions of Jeff Rothenberg, the main researcher, is that emulation seems to be the most promising and viable preservation strategy that meets the requirement of authenticity. In his opinion emulation surpasses all the other strategies, such as migration and standardisation.([4]) Everyone who knows about Rothenbergs writings will not be surprised by this point of view. The overall conclusion or better recommendation of the report is to build a testbed for doing experiments, especially with emulation. The main arguments are that in the present situation there exist no clear solutions yet, that there is no practical experience with emulation yet and that in the very dynamic environment of ICT better insights are needed to see what is feasible at the moment.

The proposed model for a generic preservation process consists of a dual strategy: a top-down or conceptual approach and a bottom-up approach. The first one includes several steps:

1. analyse functions that records must support,
2. derive authenticity characteristics (from these functions), and
3. decide what attributes have to be preserved.

The bottom-up approach analyses the technological alternatives. Both approaches provide input for the last step, the choice of the most appropriate technological strategy in connection with the authenticity requirements.

(3) The report is written by Jeff Rothenberg and Tora K. Bikson and can be found on our website: www.archief.nl/digiduur. Although the website is in Dutch, one can find the English title: '*Carrying authentic records through time in a usable and understandable way*'.
(4) Not everyone to say the least shares his ideas. See for instance David Bearman, *Reality and Chimeras in the Preservation of Electronic Records*, in: D-lib Magazine, Vol 5 No 4, April 1999.

In discussing the follow-up activities the idea emerged at the National Archives that not only a testbed should be built, but also a first instance of a digital repository itself. A testbed is very useful for acquiring more experience and insight, but it does not offer any facility for really preserving digital records yet. I agree there is some ambiguity in this: how to build a digital repository when you don't have sufficient insight what is feasible? Nevertheless, there exists a strong feeling that for the short-term something more concrete than a testbed is needed, how primitive it might be. This is reinforced by the fact that in 2000 a new regulation on the arrangement and accessibility of archival records will become valid. This regulation is based on the Dutch Archives Act of 1995 and provides rules and requirements for ensuring the long-term accessibility, authenticity and longevity of records, both paper and digital, and concerns not only government organisations, but also the archival institutions. It would have made the position of the National Archives very weak if there would have been no facilities for preservation.

So enough reasons for adopting a two track approach, that encompasses the development of a testbed and the building of a primitive digital repository.

In the testbed project we have started this summer a European tender. The scope of the project is not only to test the emulation strategy, but to widen it and do experiments with other strategies, which are useful in this respect, as well. Examples are the approach adopted by the National Archives and Records Administration (NARA) of the US and the strategy of migration, that is recommended by people as for instance Charles Dollar and David Bearman and with which most experience has been built up so far, mostly in national archives of the US, Sweden, Finland etc.

The purpose of the testbed is defined as to find out what preservation strategy is best for what type of digital records, taking into account the authenticity requirements. The underlying hypothesis is that there is no single preservation strategy yet that meet all requirements. This project is scheduled for three years.

The other project is the development of a small and simple digital repository within the National Archives. This project should be finished in the beginning of next year. This sounds, and I think is, very ambitious. The goal is to be able to preserve simple types of digital records, that will be acquired in the near future from government organisations. It concerns textual documents, such as word-processing document and e-mail, small and simple databases and simple spreadsheets. It is the intention to preserve text documents in PDF and/or in XML as being required in the above mentioned Regulation on digital records. Next to these formats the documents should also be preserved in their native format.

At this moment the design is being made. This should define the basic functionality of the repository and the datamodel.

It is obvious that the time-table of both projects is not synchronous. Because of the European tender it will cost more time to start with the testbed project. Hopefully that will be possible in the beginning of next year, when the repository is almost ready. The experiences with the development and building of the digital repository however could and probably will help in developing and using the testbed. One of the benefits now is already that more people at the National Archives are getting involved and that there is a beginning of common understanding among them on the issues around digital preservation.

Both projects have an innovative character and address specific issues with respect to organisational, archival and technological aspects, that give a view of what the transition from paper to digital record keeping and archiving requires.

Issues

Organisational

One of the purposes in both projects is to involve government organisations. They should provide the digital records for (long-term) preservation, and for carrying out the experiments. Apart from

Issues

- the acquisition of digital records
- the functions of a repository
- the metadata
- standards (PDF, XML)
- intellectual and physical/technical perspective

that they should play an important role in the evaluation of the results especially of the testbed project. In this respect they have to establish and assess the authenticity characteristics of the records.

In practice this is not as simple as it seems to be. It is not the attitude of the people involved, that makes it difficult, no they are most prepared to commit themselves, but the problem is to find appropriate digital records. Although most of the information and records produced in government organisations is in digital form, still most of these digital records are printed on paper and preserved as such. Even e-mail messages, the characteristic expression of digital communication, are treated in the same way. Furthermore there is a lack of overview of what exists and is created in digital form. The cause has to be found in the fact that a record keeping regime for digital records is still missing. Records managers are still mainly dealing with paper records, although they are aware of the issue of digital records. There are hardly any rules, guidelines and/or policies yet in this field. So it is hard to find out where digital records are, let be to know what digital records exist. Nonetheless the projects help in getting the necessary information and push knowledge about this forward. It also helps in coming to grips with digital records, especially with identifying and describing (metadata) them.

This problem of finding digital records leaves you with a strange feeling and it makes you wonder whether you should be worried about this or not worried at all, but feel reassured that things are not as bad as people say they are, when you deal with digital records and memory. Nevertheless it is necessary to find and to acquire a sufficient variety of digital records with which the experiments can be carried out and that could provide us with better insight into the different administrative, archival, and technological aspects.

At the other side especially the project for building a digital repository involves people at the National Archives into concrete activities. In the first phase this will be the acquisition of digital records, and thinking about the requirements and procedure of transfer. Furthermore they have to identify the functionality of the repository, including the metadata, and to find out about the organisational aspects of the implementation of the system. Finally when the repository system will be finished, they will have to manage and control the system. That will have quite some impact on the existing procedures and rules. Digital records require another treatment and this whole project or exercise will help in establishing new procedures etc.

It will also require new knowledge and new skills and people who have feeling with this area of digital records. This point is crucial and still a problem, because there is still a lack of adequate training courses.

The design of the repository: the main functions

One of the first things in designing the digital repository is to establish the scope of it, especially the interaction with the environment. And the question to be answered here is, what is actually a repository for the preservation of digital information? Is it only the function that enables the technological survival of digital objects or is it more something like a record keeping system?

The modelling of the digital repository is based on the Open Archival Information System reference model (OAIS), which is now under review as a draft ISO standard ([5]).

The main functions of a repository according to OAIS are:

- ingestion, which deals with the registration and validation of content and technological aspects.
- storage, which deals with the storage of the digital objects and the migration of them
- administrative or data management; in this function all information about the activities carried out are managed, and
- a delivery function that provides access.

(5) For information about this reference model: ***www.ccsds.org/RP9905.html***

The maintenance function is included in the archival storage function, but is rather superficially identified. That is strange, because that is in fact the core business of an archival information system that should serve as a repository for long-term preservation. This function should take care of the technological survival of digital records or objects by using strategies as migration, or conversion, or emulation. In the model only the migration strategy is taken into account.

It is furthermore not clear whether the data management function in the model is sufficient to take care of intellectual metadata. This metadata concerns all information about the provenance and authenticity of digital records and as such this function is necessary for preserving digital records.

With respect to electronic record keeping the reference model can serve to a certain extent as a model for a record keeping system. In the world of records there are at the moment actually two preservation systems, the record keeping system within government organisations and the archival system within archival institutions. The difference is the long-term preservation, the functionality is almost similar.

The main functions of ingestion, maintenance and access of records are conceptually the same in a paper and a digital world. The important difference is the technical maintenance of the records, which is in a digital world entirely different.

In the repository as now being developed retrieval, though in a rather simple way, will be possible through the institutional information, such as about the record creating organisation, the business function and the business process.

Archival requirements

One of the objectives of the testbed project is to establish what preservation strategy is best for what types of digital records. In order to answer this question it is necessary to know what should be preserved. It is too vague just to say digital records. The point is to establish or articulate what the requirements are when we want to preserve digital records? Terms used in this context, as in the mentioned report of Jeff Rothenberg, are authenticity, usability and understandability. Especially, as already mentioned, the authenticity requirement plays an important role. Authenticity means that the records are what they purport to be, That will contribute to the evidential value of the records and should therefore be retained. How can authenticity be established and be measured in a world of virtual and volatile records? In order to achieve that it is necessary to capture and retain the structure, content and context of the records as they are at the moment that these records play a role in a business process. Since records are bound to the business process in which they are created and communicated, these authenticity characteristics will differ for each business process. Unlike for paper records, which are physical entities and where the authenticity is more or less fixed, a description of these characteristics for each type of digital records within each business process is necessary.

Based on the essential authenticity characteristics the requirements or criteria for preservation can be derived and a preservation strategy chosen. The issue of authenticity is however not exclusive for records, but concerns also electronic publications and other digital objects such as videos, movies, multimedia etc.

One crucial activity in designing the repository system is to define what metadata should be captured and preserved. In general three categories can be distinguished: metadata about the institutional context in which the records are created and used (provenance), metadata about the records themselves both the intellectual and technical aspects in order to capture the authentic characteristics, metadata about the management and maintenance of the records once they are ingested into the repository.

Another question to be answered is, how do we ensure access over time? This encompasses not only the technical aspect of readability, but also the issue of understandability of the records. Here we enter the already mentioned world area of metadata. Many initiatives in different disciplines, such as libraries, science, archives, and audio-visual institutions, around the world are trying to articulate what metadata are necessary for what purposes. Some, such as the audio-visual community, focus on description of the content, others, such as the record keeping and archival community on the description of the context. In the latter especially the provenance of the records is important for understanding the background of the records when they were created and used properly.

Technological

On the technological level there exist a few strategies as already mentioned. The most well-known are migration and standardisation. With both quite a lot of experience has been built up and the advantages and disadvantages are known. Since a couple of years another strategy emulation is mentioned

as a promising approach. The point here is however that there exist hardly any practical experience yet, only in the world of computer games. It is still mainly theory and there exist several variants of emulation (ranging from software-emulation to hardware-emulation). That is one of the reasons to conduct some experiments with it in the testbed. One of the things to find out is whether emulation will be the only solution. Recently the University of Michigan together with the English CEDARS project acquired funding of the National Science Foundation (NSF) in the USA for a three year project on emulation. The research focuses on the applicability of emulation for certain types of digital objects of which the look-and-feel is essential.([6])

There are other approaches as is shown by the already mentioned NARA project in collaboration with the San Diego Supercomputer Centre. In this project experiments were carried out using XML for preserving different types of digital records. The results are promising and require certainly further investigation.

Both approaches will be part of the Dutch testbed as well, in order to check in what cases each approach will be useful. As already mentioned the draft Regulation on the arrangement and accessibility of archival records (1999) prescribes XML, next to PDF, as a possible format for transfer to the National Archives.

International collaboration

Both Dutch projects have also an international dimension. It is said before, the issue of digital preservation is too complex and too innovative to be solved in one (small) country and certainly not by an archival institution on its own. One way of solving that is to co-operate with other (national) institutions, such as ministries and research institutions, and with other disciplines. Another way is international co-operation. I mentioned already the connection with the NARA-project, and with the Michigan-CEDARS-project. Apart from these there will be a link with the European NEDLIB-project, in which the National Libraries of the UK, the Netherlands, Portugal, France, Germany and Finland are working together on solutions for long-term preservation of digital publications. The Dutch National Archives is also participating in this project. This project aims at the development of a preservation system also based on the OAIS reference model and will come up with a demonstrator next year.

Another major project is the Inter Pares research project. This project started in the beginning of 1999 and consists of six research groups from Canada, the USA, Northern Europe, Italy, and Australia and recently also East-Asia. It is directed by Luciana Duranti of the University of British Columbia (UBC). Its objective is to determine what policies, guidelines and requirements are necessary in order to ensure long-term preservation of authentic digital records. The Dutch project offers not only a testsite for case-studies and for results of the research, but could also contribute with results from the experiments carried out in the 'digital repository' project.

So many initiatives exist, but they are hardly co-ordinated, not only between different disciplines such as libraries and archives, but also not within one community. This is to a certain extent achieved because people participate in more than one project.

It would however be useful to have some infrastructure for digital research projects that could support co-operation within Europe in this area.

To achieve that the European Partnership of Electronic Records Research (EPERR) has taken the initiative to set up a project. At the moment this Partnership consists of the national archives of England, Ireland, Sweden, Germany, the Netherlands and the universities of Glasgow, Porto, and London. A proposal for the Information Society Technologies program of the 5th Framework is being prepared. The objectives are to establish a better infrastructure in the European Union for collaboration in this field and for a research project on *preservation pathways for long-term access of digital objects.* It includes multidisciplinary research.

(6) See about this project *www.leeds.ac.uk/cedars/*

4. Summary and conclusions

I called my presentation 'shooting at a moving target'. The target is to guarantee long-term access to digital information. The background of this should be obvious, it is ICT that determines what can be preserved and how, but it is also ICT that is developing very rapidly. ICT is somewhat ambivalent, it not only confronts us with problems with respect to preserving authentic records, but it could

also help us in finding solutions for them. The real point here is, that the solution for long-term preservation has to be found in ICT itself.

So the target is moving indeed. The issue is to move in the same direction as the target and aim at a feasible distance. What does that mean?

The other conclusion has to be that there is no final solution yet, if ever. We have to be satisfied for the time being with best practices that are as much as possible founded on current knowledge and ongoing research. The interrelationship between practice and theory is in any case necessary. The approach in the Netherlands is based on this notion and tries to achieve better insight, expertise and knowledge in order to prevent loss of digital records, which will inevitably happen if nothing is done. A second objective is to prevent an ever-growing knowledge gap in this field. Practical experience helps thinking and in this respect we can say we are doing something practical now.

All this should lead to best practices and guidelines, that are viable and useful for the moment. Eventually it should lead to more persistent solutions.

If e-commerce and electronic government initiatives will ever be successful then measures have to be taken to ensure reliability in digital communication and information exchange, but also to ensure the preservation of reliable and authentic digital records. It are these records that provide society with a firm basis for doing business in cyberspace, no they are in fact a prerequisite for it. In order to play that role however they have to be authentic, understandable, and accessible.

In our search for solutions we have to be aware of the following issues:

1. Practical experience with digital preservation, especially with modern forms of digital records, is still scarce and real solutions are not yet available; nevertheless some progress has been made during the last years. The existing experience is mainly acquired with the preservation of databases. Every new project in this field that could increase our knowledge should therefore be welcomed.
2. The issue can not be solved alone and it need to be multidisciplinary, not only archives or public administration, but also research institutions, industry (both as supplier and as facing the same problem) should be involved.
3. An infrastructure for collaboration (for instance in Europe) is missing, there is a need and there are opportunities, but somebody or some organisation has to take the initiative. The coming EPERR proposal could be a starting point. An important activity in this respect would be to inventory all the existing initiatives, and to exchange information about the (intermediary) results and experiences. The Dutch project is one of many, and should be embedded in a wider network of collaboration in this field. The result of the survey should lead to a kind of register of preservation activities and initiatives in Europe.
4. Last but not least it is necessary to develop demonstrators to teach interested people, to show archivists or librarians involved what preservation of digital objects of all sorts means and what is required, to acquire some practical experience, and to raise awareness.

„Schuß auf ein bewegliches Ziel". Die Entwicklung eines digitalen Archivs zur Aufbewahrung digitaler Informationen

Johannes Hofman

Im Zuge der Entwicklung der Informationstechnik (IT) haben sich die Speicherkapazitäten für Informationen in den letzten zehn Jahren enorm erhöht. Neben ihren Vorzügen wirft diese Entwicklung auch eine Reihe von Problemen auf, die in der Welt der Papierdokumente bisher unbekannt waren. Eines davon ist die Langzeitkonservierung digitaler Informationen, und vor allem digitaler Aufzeichnungen. Mit der IT hat sich der Charakter von Aufzeichnungen gewandelt. Es handelt sich nicht mehr um feststehende physische, sondern um flüchtige, nicht greifbare Gebilde, so daß sie leicht verlorengehen können. Die raschen Entwicklungen in der IT haben auch zur Folge,

daß die Konservierung der Aufzeichnungen über die Zeit hinweg sehr schwierig ist. Anforderungen an Datenverwaltung und Archive, wie sie in der Papierwelt vorhanden sind, müssen auf eine digitale Welt umgestellt werden. Eine der wichtigsten Anforderungen ist die Bewahrung der Authentizität der Aufzeichnungen. Sie müssen tatsächlich das sein, was sie zu sein vorgeben. Dazu ist es notwendig, Struktur, Inhalt und Kontext der Aufzeichnungen so zu archivieren, wie sie sich zu dem Zeitpunkt darstellen, an dem die Aufzeichnungen im Geschäftsprozeß eine Rolle spielen, und sie als solche zu erhalten.

Die ständig im Wandel befindliche Informationstechnik, die Vielzahl unterschiedlicher Software- und Hardwareplattformen und die fehlende Standardisierung machen dies jedoch nahezu unmöglich. Derzeit besteht der am häufigsten genutzte Ansatz darin, die Aufzeichnungen erforderlichenfalls auf eine neue Softwaregeneration oder ein anderes Speicherformat zu konvertieren bzw. zu migrieren, wenn die alte Software bzw. das alte Speicherformat überholt ist. Meist geht dies mit einem bestimmten Grad an Standardisierung einher, der die Vielfalt der Software und Speicherformate einschränken soll.

Es fragt sich, ob dieser Migrationsansatz am besten geeignet ist, um die Authentizität digitaler Aufzeichnungen zu sichern. In vielen Fällen gehen bei jeder Migration oder Konvertierung Informationen verloren. Deshalb werden andere mögliche Konzepte untersucht. Eines davon ist der Emulationsansatz, der sich auf die Nutzung der ursprünglichen Software durch Emulation der alten Hardwareplattform konzentriert. Auf diese Weise dürfte die Authentizität am besten gewahrt bleiben. Obgleich interessant, handelt es sich bisher noch um ein theoretisches Konzept. Es bedarf noch eingehender Forschung, um herauszufinden, ob es wirklich tragfähig ist.

Klar ist auch, daß sich Lösungen nur durch fachgebietsübergreifende Anstrengungen finden lassen, an denen nicht nur Archiveinrichtungen, sondern auch die Softwarebranche und Forschungsinstitutionen beteiligt werden sollten.

Ein Teilgebiet des Aktionsbereichs „Benutzerfreundliche Informationsgesellschaft" (IST) des 5. Rahmenprogramms ist speziell dem Thema Konservierung digitaler Informationen gewidmet. Stimuliert wird auch die Zusammenarbeit zwischen Archiven, Bibliotheken und Museen, die zum Teil vor den gleichen Problemen stehen.

Weltweit gibt es verschiedene Initiativen, um nach Lösungen für diese Herausforderung zu suchen. Mitunter liegen bereits Erfahrungen bei der Konservierung digitaler Informationen vor, wie z. B. bei der National Archives and Records Administration der USA in Washington, beim Schwedischen Nationalarchiv usw. In anderen Ländern werden Initiativen und Projekte ins Leben gerufen, um Erfahrungen zu sammeln und sich tiefere Einblicke zu verschaffen. Eines dieser Projekte läuft derzeit in den Niederlanden unter dem Titel „digitales Archiv". In erster Linie soll daraus eine Infrastruktur für die Durchführung von Experimenten auf der Basis von Forschungsfragen entstehen. Dabei sollten sich diese Experimente nicht nur auf einen, sondern auf unterschiedliche Ansätze richten, um so viele Informationen und Erfahrungen wie möglich zu erlangen.

Développer un fonds d'archives numérique pour conserver des données numériques ou comment viser une cible mouvante ?

Johannes Hofman

Les capacités de stockage de l'information se sont accrues dans des proportions immenses grâce à l'émergence des technologies de l'information durant la dernière décennie. Outre les avantages de cette évolution, cette multiplication a également soulevé un certain nombre de problèmes, jusque là inconnus dans le monde du papier. L'un d'entre eux est la conservation à long terme de l'information numérique, et tout particulièrement, des documents numériques. Les technologies de l'information ont fait évoluer la nature des documents, qui ne sont plus les objets fixes et matériels d'autrefois, mais sont devenus volatils et immatériels. Cette évolution les rend très vulnérables au risque de perte. L'évolution rapide des technologies de l'information a également pour conséquence qu'il est très difficile de les conserver dans le temps. Aussi faut-il adapter les critères de conservation et d'archivage, appliqués dans le monde du papier, au monde numérique. Préserver l'authenticité des documents constitue l'un des principaux critères. Cela implique que les documents soient effectivement ce qu'ils ont vocation à être. A cette fin, il est nécessaire de garder leur structure, leur contenu et leur contexte au moment où ils remplissent leur fonction dans le processus de travail et de maintenir tout cela en l'état.

Ainsi qu'il est dit précédemment, l'évolution incessante des technologies de l'information, la quantité de plates-formes logicielles et matérielles différentes ainsi que le manque de standardisation rendent la tâche quasiment impossible. Actuellement, la solution la plus courante consiste à convertir et à transférer les documents dans une nouvelle génération de logiciels ou sur un autre format de stockage lorsque l'obsolescence de l'ancien logiciel ou format de stockage l'impose. Très souvent, cette solution est conjuguée à un certain niveau de standardisation afin de réduire la diversité des logiciels et des formats de stockage.

La question est de savoir si ce type de transfert est la meilleure solution pour garantir l'authenticité des documents numériques. Chaque transfert ou conversion entraîne, très souvent, une perte d'informations. Aussi d'autres approches possibles sont-elles étudiées. L'une d'elles est l'émulation, qui consiste à utiliser le logiciel original en émulant l'ancienne plate-forme matérielle. Cette solution donne de meilleures garanties quant à l'authenticité des données. Quoiqu'étant une solution intéressante, elle demeure tout de même un concept théorique. De nombreux travaux devront être réalisés pour voir si elle est ou non viable.

Il est également évident que ces solutions ne sont possibles qu'au prix d'efforts interdisciplinaires, associant non seulement les archives, mais aussi les fabricants de logiciels et les instituts de recherche. L'une des lignes d'action des technologies de la société de l'information dans le cadre du cinquième programme-cadre pour des actions de recherche et de recherche est consacrée spécialement à la préservation des données numériques. Cette ligne d'action encourage également la coopération entre les archives, les bibliothèques et les musées, qui partagent en partie les mêmes préoccupations.

Il existe plusieurs initiatives au niveau mondial, qui visent à trouver des solutions à ce problème. Parfois, on dispose déjà d'une expérience dans la préservation des données numériques, comme c'est le cas aux Archives nationales des États-Unis (NARA) à Washington D.C., aux Archives nationales de Suède, etc. Des initiatives et des projets sont également mis en place dans d'autres pays pour acquérir plus d'expérience et une connaissance plus approfondie. L'un de ces projets, le "fonds d'archives numérique", est actuellement en cours de réalisation aux Pays-Bas. Il devrait tout d'abord déboucher sur une infrastructure permettant de réaliser des expériences fondées sur les questions que se posent les chercheurs. Ces expériences ne doivent pas être centrées sur une seule et même approche, mais sur plusieurs afin de glaner autant d'informations et d'expériences que possible.