

*erpa*training

“Policies for Digital Preservation”

Seminar Report

ERPANET Training Seminar, Paris
January 29-30, 2003



ERPANET Training Seminar Paris, Seminar Report

CONTENT

| | |
|---|-----------|
| 1. Introduction..... | 2 |
| 2. The Seminar..... | 4 |
| Context and Objectives | 4 |
| Issues to be addressed | 7 |
| Impact on organisations | 11 |
| Implementing digital preservation policies: Experiences | 14 |
| The preservation policy for statistical data in France | 14 |
| Steps taken in the UK Higher Education sector | 15 |
| Experience at the European Publications Office | 16 |
| 3. Conclusions..... | 18 |
| Appendix One: Seminar Programme | 20 |
| Appendix Two: Speakers at the Training Seminar | 21 |
| Appendix Three: Participants at the Training Seminar | 23 |



Le Centre des archives contemporaines

1. Introduction

One of the consequences of the increasing use digital information is that organisations are becoming more and more aware of the necessity of developing preservation policies. This was the main driver for organising this training seminar. Organisations, public and private, need to maintain their digital information assets understandable and usable through time and technological changes. Experience with developing digital preservation policies and with applying them, however, is still very scarce. The transition to a real digital environment is still at its very beginning.

This seminar intended to provide some guidance on this new challenge by inviting speakers that could share their thoughts and experiences and discuss the issues with seminar participants.

Seminar Setting

This seminar was co-hosted by the 'Direction des Archives de France' and the 'Centre des Archives Contemporaines' and took place in the very well equipped auditorium of INSEAD in Fontainebleau.¹ A conference translation service facilitated communication with speakers presenting in French and English. The social programme included a trip to the marvellous castle of Fontainebleau and a dinner offered by the 'Direction des Archives' to participants and speakers and provided a good opportunity to know each other better in a more casual environment.

About 50 participants came from different countries in Europe, but also as far as French Guyana. They represented a broad range of institutions in the public as well as the private sector and brought to the conference a valuable mix of experiences and perspectives.

The speakers came also from different backgrounds and presented the audience with different practical experiences, insights, and views, both at an organisational and national level. It gave a good impression and understanding of the many perspectives that can and should be taken into account when defining and implementing preservation policies.

The feedback of the participants confirmed that the seminar was very much appreciated as well as that there is a pressing need for more practical approaches and for other opportunities to discuss the many issues.

Aims and objectives

The seminar was structured in a way that helped participants to walk through the issues, starting from the relevance of the organisational context, along the issues that have to be addressed, the intended or assumed impact of policies on organisations, and finally to the actual implementation of policies. As such it went from outside or the context of the organisation to the internal aspects and vice versa. The objectives of the seminar were to provide insight in the issues, to identify the contextual influences, and to discuss the possible approaches for formulating and implementing policies.

In order to encourage discussion and a more focussed exchange of information practical sessions were scheduled at the end of each day. During these sessions participants could discuss issues in smaller groups based on some questions and share thoughts with each other and the speakers.

¹ We are very grateful for the excellent facilities and hospitality of our hosts.

In the following a report will be given of the presentations and the discussions during the seminar. This report will follow the structure of the seminar that started out with

- context and objectives of digital preservation policies,
- the issues that should be addressed,
- the impact on organisations,
- and finally the issues in relation with implementation.

It will show that developing a digital preservation policy is a complex matter with many interrelated issues and of interaction between an organisation and its environment.

2. The Seminar

Context and Objectives

The seminar was opened by the director of the 'Direction des Archives de France', Madame M. de Boisdeffre, who gave an excellent introduction to the seminar and overview of the scene. She discussed not only some initiatives in the area of digital preservation, but also emphasised several important issues, such as raising awareness especially among responsible managers, the need for education and training, and for standards, better insight in cost factors in close relation with risk analysis and finally the need to identify critical success factors in order to be able to evaluate preservation programs. It showed the complex mix of aspects and issues that have to be addressed by organisations in dealing with digital preservation and also the need for multidisciplinary collaboration. The other presentations on the first day offered different perspectives on the scope of preservation policies at different levels.

There was no real discussion of what a preservation policy is about. It was the implicit assumption that any organisation that creates, manages, and receives digital information or objects has to develop such a policy, irrespective of the type of digital object. In this respect the different disciplines as represented in the audience had to translate the presentations into their own environment. It showed also the apparent commonality of issues between different communities and consequently that there is a solid basis for more collaboration. Apparently all speakers agreed silently on the fact that preservation policies not only regard technological survival of digital objects, but also include or should include organisational, financial and cultural aspects. The scope of a policy therefore implicitly seemed to encompass the whole life cycle of these objects.

Thomas Schärli presented in this respect the approach the Swiss archival community has taken in order to develop a kind of national policy. Based on an increasing awareness that something should be done to ensure the preservation of valuable digital information, several activities were undertaken. A working group tried to inventory individual initiatives, and to set the course, but it turned out that this was not an easy thing. Archivists are not really trained in developing strategies and they underestimated the scope of the problem. Schärli identified the problem of the rather isolated archival domain and he emphasised the necessity for archivists to become a more active and accepted player in the broader field of the digital world. The fact that there is not a definite answer yet to the problem of digital preservation, and that in the Swiss case senior archival managers were not involved also did not help in formulating a course towards the future and to acquire the necessary commitment at all levels. Nonetheless, a small working group tried to develop a strategy. In 2002 they succeeded to publish a report with the intent to develop a vision on archival service in 2015. The report discusses a program with 4 scenario's, called 'portal', 'top-down', 'subito' and 'dretissima'. The latter refers to a scenario in e-Government context. The basic idea is to bring together the different initiatives and instruments into one perspective, such as for instance a centre of excellence, an alliance of software users, the Swiss ARELDA-project and a standardised platform. In this changing world appraisal and selection of what information should be preserved will be a crucial issue, as well as standardisation.

One of the effects of this high-level national policy according to one of the seminar participants is that it will provide guidance to local initiatives in Switzerland. It would also mean that preservation policies can occur and are needed at different levels, contextualised with different levels of detail and objectives.

The way a policy is defined depends very much on the context of the organisation. It will be the needs of an organisation in the first place that determine what the objectives of a preservation policy should be. Those objectives will of course differ depending on the nature of the business. Government organisations are mostly very much embedded in a legal and democratic context, business companies are guided by the level of risk exposure and the regulatory environment of the business they are in, while preservation is the core business of cultural heritage institutions that will guide their policies. That is the reason why the latter domain took the lead in putting this issue on the agenda and in defining such policies. With the increasing creation and proliferation of digital information the need for such policies is spreading into other domains.

Issues that have to be addressed in this respect concern the fact that legislation has to acknowledge the existence of digital information, which is not always the case yet. Secondly the framework for defining a policy should be the continuum or life cycle concept, which takes into account all processes from design of systems, along the creation of information to access and the long term preservation. In this context the different actors can and should be identified that have any responsibility in each of these processes. It is important to establish what role they play within the perspective of creating, managing, making accessible, and preserving digital information. Based on this analysis action can be defined and taken for raising awareness among them, guidelines can be formulated for managing the information, and finally training programmes can be developed. Awareness raising has to concern not only the information professionals and record creators, but as Madame De Boisdeffre indicated, also the software suppliers. In this respect one of the questions that was raised during discussion, was to what extent legislation could or should influence industry by prescribing or setting standards. That may depend on the cultural context, but it has also a strong relationship with developments in e-government or e-commerce, which require standards and agreements among partners in order to enable it.

Development of training programmes and manuals that could help people understanding the issues and guide them in the different areas of digital information are other essential matters. These programmes and manuals should discuss concrete problems that arise from practise. On the other hand the development of reference models should prevent organisations of continuously re-inventing the wheel. These models and standards will also provide the necessary framework for positioning the indicated problems.

Essential in this respect is the fact that activities, which support proper information management including preservation, have to be an integrated part of the activities people perform. So far many organisations still grapple with the possibilities offered by information technology and how to apply them within the context of their business. The point, however, is that information management and preservation should not be seen as a separate domain that can be dealt with later. It has to be embedded and integrated in the day-to-day business activities.

Peter Emerson emphasised that not everything has to be preserved and that aiming for too much may lead to losing everything. He also indicated the differences in scope and perspectives between public and private sectors. Public sector will have a longer term view and broader scope based on societal needs than private companies that tend to have a much shorter cycle determined by the nature of their business. Nonetheless, the issue of preservation is the same for all organisations, since on average the lifecycle of information is 7 years, while the average life cycle of the creating systems is 3 years. Also, the decision to preserve or not to preserve for long term has to be made at the point of creation. It is not a sound option to do it later in the life cycle, as it will be more difficult, and, more importantly, more costly. Appraisal decisions are based on the business needs and on legal requirements. Wrong decisions may impact on the ability to do business. Therefore all decisions and the reasons behind them have to be documented properly for accountability reasons. The actual decisions on what to preserve are in practice made in all organisations, public and private, at the desktop, mostly driven by immediate business needs. The immediacy of information in business is a new phenomenon, which is guiding actions regarding information management. Determining the value is based on the functions and activities of an organisation. In this respect a functional analysis or activity modelling is necessary to identify what functions an organisation is responsible and accountable for and subsequently what activities an organisation does in carrying out the functions. This may be a cumbersome task. There has to be made a distinction between core and support functions. The first are the functions for which the organisation exists, the latter those enabling them. In assessing the value one could distinguish different levels of ranking the preservation needs and apply them to the identified functions. According to Emerson policies have to address and answer at least the questions of what, why, when, how, where and who. They will be discussed further in the following section. It is the assessment of the value of created information and records that is, however, the core activity, that should guide the possible answers.

Policies should be independent of technology, since technological evolution is still taking place very rapidly, but also because it is not apt to deal with preservation issues and there is no definite solution yet. Business requirements for preservation should be leading and translated into procedures and methods based on the current technological context. They should also be communicated with the suppliers of software and hardware, so they will be aware of them and can accommodate them more adequately. However, because organisations are still in a transitional state in incorporating IT into their business, defining policies entails for the moment a pragmatic approach and attitude. Another reason for this is that aiming for too high and visionary goals may endanger the immediate goal of preserving information that has been created or will be soon. The first thing is to find a balance between the necessity to save what is already existing and the longer term need for an appropriate infrastructure that enables proper management of digital objects, including preservation. Quick wins may also help to convince senior management of the opportunities and available possibilities. Their attitude is still rather distant, as several speakers indicated and this is a point of continuous concern. The commitment of decision makers is a precondition, however, for being successful in this area. Conducting a risk analysis, that identifies the value of information in doing business or carrying out activities and the consequences of losing valuable information assets, will be a useful instrument for raising awareness among them. This also relates to the implementation of a policy, which foresees

different stages. That could start with a risk analysis for developing a business case, followed by immediate and realistic steps to be taken in order to deal with the current situation, and end up with the more structural and fundamental changes that have to take place in order to achieve the ultimate preservation goals that are set for an organisation.

Summarising, policies have to provide not only a framework for information creation and preservation activities, but also practical guidelines for immediate use to safeguard information created now. Apart from that policies can envision possible future or ideal situations, but at the same time have to be realistic and provide approaches for the current situation, which will be in most cases far from ideal and represent a very unstructured and inadequate environment.

Issues to be addressed

This session went into the issues that have to be addressed in a preservation policy. Speakers from the archival and library community presented their experiences and insights. It became clear that those organisations that had already formulated policies, such as the Archives Contemporaines de France, now have to review them, because the technology has developed in such a rapid pace that these policies no longer are adequate. On the other hand those who have not yet defined them, have to start immediately as already indicated. Taking the items of Emerson as mentioned before as a starting point at least questions as to who, why, what, when, where and how have to be answered. The who question should identify the different actors and the role they should play. That includes the responsibilities and the implicit accountability and embraces the whole organisation, from the top to the information creators. It is interesting to see that the Italian government institute for information technology, AIPA, has determined in a regulation that government organisations should appoint a 'preservation manager'.

Secondly the issue of what information should be managed and preserved as is discussed in the previous section. The issue of what should be preserved is closely related and in fact dependent on why. The business context will be the first criterion, but in the public sector the societal needs will represent another relevant issue. This part will furthermore discuss the different type of information objects. It should not be limited to documents, but also include spreadsheets, multimedia, databases, websites, etc. Different types may for instance entail different guidelines for appraisal. Relevant issues may also be the need for ensuring possible coherence between these different sources in different systems and their interdependency. Furthermore issues such as future use, re-use and re-purposing should be assessed in this context. The possibilities of modern technologies for sophisticated use and manipulation of information resources offer new perspectives. It goes without saying that these new possibilities should not infringe on the integrity and authenticity of the information resources that are preserved. Determining what has to be preserved includes also how long. Preservation of some digital objects may be necessary for some decades, while for others permanent maintenance will be required. Again this is established in close relationship with business and societal requirements. Preservation relates also to intellectual property rights that may restrict the use of the digital sources that are subject to those rights. Organisations have to negotiate with the owners, not only about the

use, but also about the preservation activities that have to be performed on the objects in order to ensure their survival.

At what moment in time action on preservation is necessary, is another important matter. As already alluded to previously there was an agreement among speakers and participants that preservation measures should be taken from the very start of records or information creation. If taken later on in the life cycle the risk of losing valuable information will be much greater, but more importantly the costs for implementing them will be much higher. The volatile nature of digital information and its high vulnerability require immediate care after creation. It will mean that measures should be formulated and integrated already in the stage of system design. It is obvious that preservation itself is an ongoing activity that requires permanent attention and monitoring.

With respect to the question how digital objects should be preserved issues such as metadata and methods such as representation networks for reproduction (CEDARS-project) are discussed. These networks manage tools necessary to reproduce the objects. Metadata are necessary not only to retrieve and to understand the digital objects, but also to manage and preserve them in an authentic and usable way. The notion of identifying the 'significant properties' of digital objects as suggested in the CEDARS project and presented by Ellis Weinberger, is an additional approach. Describing these properties will help to preserve the objects in the form and structure as they were intended when they were created. An important part of preservation is also the requirement that all preservation activities have to be documented properly to support and account for the authenticity and integrity of the objects. Weinberger also mentioned the possible use of free and open source software that enables changes and improvements and may be helpful to re-apply on future platforms.

Where digital information has to be preserved depends among other reasons on the length of their preservation. Digital objects that need to be preserved for a while may stay within the creating organisations, but digital information identified for permanent preservation will be at some moment in time mostly be transferred to a specialised institution, such as an archives, library or data archives. Apart from that organisations may want to collaborate in the area of digital preservation and establish a network or a distributed archives. Another possibility is to outsource it. Multiple copies of digital objects stored at different locations could help to ensure their survival, if adequate preservation management is in place. Again the legal context and the continuing funding of repositories are supporting and necessary conditions for preservation.

An important aspect, it has been said before, is to inform and train staff in the application of good practices and procedures. In order to be able to cope with the new challenges and formulate policies training of people in their different roles is and will be essential. Involving the actors is not only important to take care that guidelines are executed. Getting feedback from the staff on the feasibility and efficiency of the guidelines is also necessary to ensure and maintain the success of the policy. Although all participants acknowledged the importance of training, they also identified the problem that there is a lack of good teachers, as was observed by for instance Joël Poivre. Training the trainers is therefore an issue that should be taken up more systematically.

The other recurring issue in the presentations was the need for standards and guidelines. The emergence of open standards is encouraging in this respect.

Scope and definition of policies will be influenced by decisions about collaboration with other parties, the necessity of multidisciplinary approaches, and the need to influence legislative bodies. These national bodies should formulate criteria, set legal frameworks, prescribe standards, etc. Funding is another big issue that has to be addressed. Preservation of digital information requires huge investments in new technological and organisational infrastructures and that in turn needs the awareness and subsequently the commitment of senior management and political bodies. It all refers back to the context in which organisations are working and the fact that organisations are lagging behind in keeping up with the rapid changes caused by the speed of technological developments.

A common feeling among participants was that there is still little experience in the field of digital preservation that can be used to build upon in defining a policy. This requires apart from theoretical approaches also more exploration of the issues in experimental projects. The results can then be used to refine existing policies. Another thing is the difficulty to anticipate on future developments and possible future demands. So, with little experience and a future that is difficult to predict defining policies will not be an easy task and will have to take into account these particular circumstances. This is aggravated by the fact that software suppliers are at first instance of course promoting their own interests. It will also be necessary to update and adapt defined policies continuously to keep in line with the most recent developments, both within organisations and in information technology.

A major issue that was discussed amongst participants is that of costing models. Again there is not much experience and information about the costs aspect of preservation strategies and that is one of the main criteria upon which decision makers will take their decisions. The necessity of developing costing tools, that will identify the cost factors and will enable to simulate different scenario's with the entailing costs was widely underlined. Factors that should be taken into account are not only infrastructural and technical costs, but also the value of information. In this respect Two possible influences on the **value of information** were identified:

- (1) the issue whether the information is created in a core or a support function of the organisation (as Peter Emmerson in his presentation elaborated on), and
- (2) the organisational context and possible external requirements (e.g. mandate and regulatory environment).

This touches very much upon the important issue of appraisal and selection. Libraries and archives have much experience in this area, be it from different perspectives, but outside these disciplines people are struggling with it. The development of appropriate models and tools could therefore use the experience of the archival and library communities. This leaves open the issue that appraisal may not be an issue any longer, since the growing and increasingly cheaper capabilities of IT to store information make it possible to keep everything. This opinion is sometimes heard, but does not really address the issue of accessibility and retrievability, nor the issue of preserving the huge amount of information through time. In the end it is, however, again a matter of cost and economic criteria. Will the often intricate effort to appraise information be outweighed by cheap storage?

Enrica Massella Ducci Teri presented the approach to digital preservation of the Italian Public Administration Authority for Information Technology, AIPA². This public authority was established in 1993 in order to, amongst other tasks, guide and support public administration to put information systems in place and to manage them, as well as supporting Italian government in legal questions and standardisation processes, so in all an important basis for policy making.

She explained that with the increasing use of digital processes the gap between citizen and public administration is becoming smaller. After all, the products of government bodies are information products, such as reports, checks, licenses, and statistics. In an effort to simplify processes and to provide better services to the citizen, the creation of digital documents is promoted. The goal is to preserve only digital documents, including digitised replicas of paper documents.

Building on previous laws on the legal validity of digital documents, a law was passed in 2001 (AIPA Regulation 42/2001) to ensure the integrity and authenticity of digital documents over the long term. This law concerns not only public administration, but also private companies developing policies with respect to digital documents that should be kept for legal reasons. In the scope of this law a common glossary of document and process typology was created. The law sanctions the migration of digital objects to new systems and storage media, and their conversion to new formats if necessary. It also provides a security policy based on digital signatures and time stamps. The technical renewal of digital signatures is expected to be necessary at least every ten years.

In order to ensure the integrity and authenticity of digital documents the law stipulates that every government organisation should have a 'preservation manager'. The person in this position has responsibility for the systems where the documents are stored and for keeping various software versions, and verifies periodically the readability of documents, as well as oversees other tasks involved in digital preservation. He or she is therefore very influential.

The preservation manager gets ongoing technical training, while the end-user is offered the necessary basic technical training. Training courses have also been developed for management in order to raise awareness and to develop an understanding of the issues involved in digital preservation.

This law had quite an impact and working groups comprising specific governmental bodies, institutions, and private companies have been put into place to formulate digital preservation policies on restricted operational fields under their jurisdiction; such operational fields include financial documents, government documents, diagnostic imaging, and research projects.

The goal in a future revision of the AIPA law and policy is to incorporate the recommendations of the InterPARES 2 project³ concerning authenticity, reliability and accuracy of digital objects. Also, the current law is restricted to long-term preservation processes. In a future extension, it should include the entire life cycle of digital objects with the focus remaining on long-term preservation.

During the first practical breakout sessions participants discussed questions about the benefits of having a preservation policy, the risk of not having one, the possible contextual influences on policies, and the focus. The main issues that came out of the discussions were getting the attention of senior

² Autorità per l'informatica nella Pubblica Amministrazione; <http://www.aipa.it>

³ International Research on Permanent Authentic Records in Electronic Systems; <http://www.interpares.org/>

management and the fact that much is still very volatile, because we are still in a stage of transition from paper based to digital environments. There was also the feeling that it still was too much technology driven, instead of being lead by intellectual and organisational principles. Preservation is furthermore just one of the issues, organisations have to define new ways of doing business and to go through a complete change of infrastructure. Also, just because of the rapid technological developments, policies are needed and should prevail since they envisage the long term. Nonetheless there was also the recognition that readily available solutions are needed to make some progress. A related point that emerged was whether a top-down or bottom-up approach was preferable. Participants agreed that it should be a mix of both: think globally, act practically or at a manageable level. Risk analysis will help to convince decision makers, especially when it makes clear what costs are involved in not having a policy. What will it mean for instance if digital data assets are lost or if information is not reliable?

Finally collaboration among organisations, but also among disciplines is seen as necessary. It will be most cost-effective to share information and solutions, then inventing everywhere the wheel again for instance. Difficulty may occur in the area of private companies, who are mostly not very willing to share knowledge and be open about their approaches, because of competitive reasons. The opportunities offered by the Web, however, are an important driver for more collaboration, since people have to agree on how to communicate. Standardisation may help and will provide a necessary basis for enabling e-business and e-government.

Impact on organisations

The third session was dedicated to the impact of preservation policies on organisations. What has to change or be adapted in organisations in order to enable proper management of digital information? Having a policy is one thing, but implementing another. Although the policy and strategies should be determined by the organisational context, organisational change is very much dependent on human behaviour and whether people are prepared to accept change. The cultural context within an organisation often turns out to be an impediment to a successful implementation of a policy. John McDonald, a private consultant with longstanding experience in information management within Canadian government, went into this important issue and gave a comprehensive overview of the aspects involved. Having completed the initial design and development of a preservation policy, the real challenge lies in its consistent implementation.

In his presentation McDonald gave an overview of the Management of Government Information policy of Canadian Government. In this policy it is acknowledged that a good infrastructure is needed and that this has to be supported by a strong governance structure which reflects clear accountability for information through its life cycle. The infrastructure consists of e.g. standards, practices, systems and people and is very much based upon an understanding of the business or the activities that should be carried out by government. Apart from the business perspective there is the information perspective, which looks at the role of information, and its creation, use and preservation within organisation and business activities.

The business, technology and information landscape in which a policy has to be implemented is complex, however, and in this respect McDonald makes a distinction between three kinds of environment that can be encountered within an organisation:

- *The structured environment* – An organisation with a structured environment has defined functions; highly structured business processes; assigned accountability; and a rigorous approach to systems design and development. The infrastructure in place includes policies, standards, and practices to govern and develop systems and data management, but few account for assignment of accountability for retention and long term preservation. Digital objects are seen as data and hardly as evidence of what they represent. Integrity of data is managed, but not the authenticity or retention.
- *The unstructured environment* – In an unstructured environment not just preservation, but the whole life-cycle of information management is a problem. The main reason is the emergence of PC's with personal support utilities. Corporate work procedures and standards are not applied. In this environment work processes are poorly defined and accountability for electronic records is weak. There might be a records management policy in place, but it does not yet include electronic records. Records management controls are absent, as is shown with the e-mail issue. Solutions, such as records management applications are still in their infancy.
- *The web environment* – Where in a web environment systems and policies for a digital environment are in place, they focus on content management, publishing and communication (and as such the web is becoming increasingly important). Systems do not address record keeping and sensitivity about the long-term value of information is low.

Each of these environments poses its own challenges when implementing a preservation policy. While the 'structured environment' might appear easy at first glance, one should not underestimate the difficulty of incorporating a preservation programme in an existing and perhaps rigid policy landscape. In a structured environment functional requirements with regards to retention and preservation should be incorporated in systems design using existing system development methodologies. In an unstructured environment better management of e-mail and other document should be provided as well as development of record keeping systems.

Common to all these environments is that different disciplines are involved in information management. The success of implementing a preservation policy lies among other things in bringing those disciplines together. For example a web master and an archivist may be suspicious about each other since they may think they have conflicting interests. It will be necessary to find a responsible person to integrate the diverse tasks and communities involved. One issue is communication between the various communities. This is to some extent a problem of terminology. The word "record", for instance, has different meanings in different communities. But communication problems go beyond terminology and relate to concepts as well. For example, database specialists are in place to manage the integrity of data, but they may not understand concepts such as authenticity, evidence, and preservation.

An big issue is also to find the right balance between what should be done and what the organisation can adopt or is able to absorb. That depends on whether the organisation is 'ready' to

apply certain proposed policies. To identify this readiness there has to be a mapping between the models and solutions that are proposed and seen as necessary to apply and the way the organisation is developing in areas as the business, the enabling systems and technologies and the information.

Here McDonald introduced the concept of **capability and maturity models (CMM)**. This concept comes from the software industry and is later adapted by financial management community. More recently these models are also being used in a government environment, e.g. in Canada and by the World Bank. They offer now a methodology to assess the capability and maturity of an organisation's information management infrastructure, including the long-term preservation of information. Five levels of maturity are distinguished, starting from a rather chaotic (level 1) to ultimately a controlled corporate environment (level 5). The methodology also supports the process of moving from one level to the next. The first steps are to assess (a) at what level an organisation currently is, (b) whether it is on the right track, and (c) to recommend what the next step for improvement should be. The levels that are identified are:

- *Level 1* – An infrastructure for managing information is not in place. Information is created, used and retained based on protocols established by individuals or work groups.
- *Level 2* – An infrastructure is in place for controlling the retention, protection, and disposition of information. However, the relationship between the management of information and business needs is weak.
- *Level 3* – An infrastructure is in place to ensure that information is created to support business activities, that information can be accessed and retrieved effectively and that it is retained and disposed of according to corporately approved standards and in compliance with laws and policies. The relationship between information management and business needs is strong.
- *Level 4* – An infrastructure is in place to ensure that the right information in authentic and reliable form is provided to the right person at the right time in the right format and at a reasonable cost.
- *Level 5* – An infrastructure is in place to exploit information to meet the needs of a knowledge-based organisation and its clients and partners.

Application of this approach, however, has to take into account that there is a relationship between the maturity of information management and the maturity of the organisation, business and so on. Understanding this interdependency is essential for developing and achieving the appropriate information and preservation policy.

The presentation of McDonald as well these of others made clear that one size does not fit all and that organisations have to tailor their policies to their own needs, to the context in which they are working and in accordance with the level of maturity at which they are.

Other models exist that can help and support the design and implementation of preservation policies and strategies. Models offer well-structured criteria and instruments that contain the experience of previous initiatives. A good example is the Open Archival Information System (OAIS) that is widely acknowledged as the reference model for preservation of digital information.⁴ It is used as a starting point for further guidance on preservation, as is shown by the report on "Attributes and

⁴ Reference Model for an Open Archival Information System; <http://www.classic.ccsds.org/documents/pdf/CCSDS-650.0-B-1.pdf>. ISO standard ISO 14721:2002.

Responsibilities of Trusted Repositories”⁵. As with all high level models however it needs to be customised to the distinct needs of the individual communities.

Another point that was raised was the fact that people at different levels in organisations seem still not really aware of the issues at stake. More effort has to be put in in order to raise this awareness and to show what benefits could be gained if one has an adequate policy in place or what risks could be the consequence if nothing will be done or if insufficient measures are taken.

Implementing digital preservation policies: Experiences

Although experience with implemented policies is scarce, some early examples were presented that gave an idea of what can be done. They also provided some lessons learned so far. Data archives preserving statistical and social science data were among the first to deal with the issue of digital preservation, but are now facing new developments that entail new challenges as was told by different speakers.

The preservation policy for statistical data in France

The mission of the French institute of statistics and economical studies, INSEE⁶, is to collect statistical data, which is used by the French ministry of economics, finances and industry, MINEFI⁷, for taking political and economical decisions. In this respect it was confronted very early with the challenge of managing digital data. Recognising the need for a systematic approach for archiving statistical data in digital form, the institute set up a digital preservation policy in the late 1970s.

Though several updates have been necessary since the inception of the policy, the original vision remains relevant. Its broad objectives are to guarantee the accessibility of the digital objects, to control access rights, and to establish a catalogue of the archived data. The policy defines the documents to be preserved and the period of retention for each type of statistical data. The usual retention time is between five and twenty years. Digital objects that are deemed to be of historical interest are preserved permanently. Establishing and assigning these retention schedules has proved to be quite a daunting task, which underwent several revisions in the three decades the policy has been in place.

Another focal issue in the policy is the assignment of responsibilities. This includes the appointment of a committee providing direction and developing guidelines. This committee consists of representatives from diverse associated institutions. Finally, a special department for archiving has been set up recently at INSEE.

Particular emphasis is given to documenting the archived objects. Often the documentation is dissociated from the actual objects and a more systematic approach is necessary, such as the development of a database for managing the metadata. Sufficient documentation also has to be obtained from the producer, following the guidelines of the committee. Additionally, metadata is collated regarding to the technical characteristics of the archived objects. A unique identifier is

⁵ RLG/OCLC: *Trusted Digital Repositories: Attributes and Responsibilities* (May 2002)

<http://www.rlg.org/longterm/repositories.pdf>

⁶ Institut national de la statistiques et des études économiques; <http://www.insee.fr/>

⁷ Ministère de l'économie, des finances et de l'industrie; <http://www.finances.gouv.fr/>

assigned and practices for storing the objects have been developed through the archival programme of the institute.

For ensuring the long-term preservation of the digital objects, a tight co-operation with the 'Centre des Archives Contemporaines' (CAC)⁸, in relation to the 'Constance' project has been initiated. Part of this is the transfer of backup copies of statistical data with historical value to the CAC. This involves around 5.000 of the overall 27.000 statistical objects in digital form as kept at INSEE.

Overall the digital preservation policy at INSEE has proved to be very efficient and robust. This is also due to the homogenous nature of the statistical data kept and the fact that older formats dating back to the time of the policy's inception are still being used now. The growing diversity of data formats, and new statistical data formats becoming more prevalent (e.g. the company SAS⁹ provides a widespread format), however, call for further action in this field. Following the demands of associated institutions and for the convenience of the users it is planned to install automatic conversion tools that migrate the data to new formats when requested. An important issue here is ensuring the reliability of the data after the transformation, which requires adequate procedures.

More work is also needed to ensure the adequacy of documentation. The awareness of the creators has to be increased, and their co-operation for producing sufficient documentation has to be acquired, though this has already improved recently. Strict guidelines are defined in order to guarantee the adequacy of the documentation. At the same time, however, the institute learned that a more flexible format of documentation is needed to represent the heterogeneous background of the statistical data.

At present not only the technological developments and the new software formats require to be incorporated in the policy, but also a change in the organisational structure is necessary to deal adequately with digital preservation at INSEE.

Steps taken in the UK Higher Education sector

Awareness in the UK Higher Education sector with respect to digital preservation is relatively high. In fact, many institutions have taken first steps to meeting digital preservation challenge and some have developed a policy. Neil Beagrie gave in his presentation some insight into these developments in the UK and introduced practically oriented guidance and case studies about the development and implementation of a digital preservation policy. He also shared some of his experience from developing preservation policies for three different organisations.

The latest of those was the "JISC Continuing Access and Digital Preservation Strategy", which was issued in late 2002 in the scope of the JISC Preservation Focus¹⁰. It emphasises the importance of proactively including the whole information lifecycle in preservation activities, also underlining the importance of partnerships with institutions beyond the members of JISC. One of the objectives of this programme is building a Digital Curation Center that could help in making preservation decisions, and could also offer practical support, for example, by establishing a technology watch. Additionally, the JISC policy will support national planning in the UK with regard to digital preservation.

⁸ <http://www.archivesnationales.culture.gouv.fr/cac/fr/>

⁹ <http://www.sas.com/>

¹⁰ <http://www.jisc.ac.uk/dner/preservation>; JISC – the 'Joint Information Systems Committee' in the UK

Another UK initiative was the launch of the Digital Preservation Coalition (DPC)¹¹ at the House of Commons in 2001. That event and the activities of this organisation have raised awareness regarding digital preservation considerably. It brings organisations from various sectors within the UK together and as such fosters the exchange of experience between those communities. With the range of stakeholders who participate the impact is more significant. Beagrie pointed out the potential for national coalitions like the DPC in other countries across Europe. At a European level, an umbrella organisation like ERPANET could then establish a network between those coalitions.

Experience at the European Publications Office

Another example is the European Publications Office (EPO)¹², which is the official editor for European institutions and organisations. Although the EPO reports to the European Commission it does not fully belong to it. Other bodies that make use of its services are the Council of Ministers, the European Investment Bank, the European Court of Justice, and the European Parliament amongst others. The office co-ordinates the printing of three official journals daily, legal documents such as treaties, internal documents, and others. Currently all those documents are translated into 11 languages, which will increase to 20 languages in 2004. Documents are provided in paper format as well as digitally, on CD's and on websites.

To fulfil its mandate the EPO collects documents from the authoring entities. It takes care of the translation of the texts and edits them, co-ordinates the printing at partnering printing companies, and disseminates the results among the European institutions as well as the general public. To fulfil all those tasks in time demands a tight workflow management, as well as clear contracts and service level agreements with the partners.

The archiving of all these digital publications is essentially outsourced. A contract with the service provider is for a limited period of 3 to 5 years. After that term the service is reassessed on its continuing appropriateness and feasibility. If the service is considered to be still appropriate for the current demands, the contract can be renewed; otherwise a new EU-wide tender is put forward. As part of this change to a new partner, the conversion of the digital objects to new software formats is considered. This procedure has already been followed three times since the activities in digital preservation of the EPO began, and they are quite happy with this policy underlining that the stability over the contracted term offers a good cost-benefit ratio.

The basis for being able to change between different service providers is the commitment to application neutral data formats. Also the separation of the archival format and the presentation format results in versatile and portable archives. Documents are stored in specifically defined SGML¹³ formats. For presentation the PDF format¹⁴ and the TIFF format¹⁵ are being used, which are de facto standards. The very generic SGML format also allows the export of the documents to XML¹⁶ taking

¹¹ <http://www.dpconline.org>

¹² <http://publications.eu.int>

¹³ Standard Generalized Markup Language; <http://www.w3.org/MarkUp/SGML/>

¹⁴ Portable Document Format

¹⁵ Tagged Image Format

¹⁶ Extensible Markup Language; <http://www.w3.org/XML/>

advantage of new tools becoming available for XML/XSLT.¹⁷ Partial data extraction is even possible for cataloguing formats like UNIMARC or USMARC as used by libraries.

Selection criteria and retention schedules proved difficult to define. This has been addressed for some of the documents for which the EPO is responsible:

- legal documents and official journals are archived permanently,
- paper series and monographs are retained as long as the client requests,
- and statistical yearbooks and the like can be disposed of after a number of years.

For other documents (for example internal paper files) retention schedules remain to be established.

Also, an approach to the preservation of web content on the Intranet as well as the publicly available sites is being developed. Web content consists of over 1 million pages and their volatility and dynamic nature poses a particular challenge.

One of the current topics of concern is the definition of a unique identifier for collection items of EPO, which would better support the migration to a new generation of technology. Future goals include to further increase the efficiency of the workflow, for example by enforcing central production tools. In addition, the EPO will try to make its services more user-friendly by investing in web services and e-commerce. Part of this will be the launch of the online service "EU-Bookshop" later in 2003.

In the second practical session participants discussed questions about who are involved or will be influenced, what factors can make a policy effective, how should a policy be embedded in the whole of an organisation's framework of policies, what are the preconditions for successfully implementing preservation policies, how can we start and so on? Participants agreed that if it will not be possible to get immediate support of senior management still small steps can be taken to improve the situation. A lot of effort is required to get management on board, while also funding often is a problem. Possible triggers that could serve as starting point are e-government projects or re-organisations. The initiate projects that only have a preservation focus may fail not only because it is not an objective on its own and has to be connected to organisational business needs. Some people also identified the already many existing projects and the 'initiative fatigue' emerging from that. Connecting and embedding the preservation requirements in other broader initiatives within organisations therefore will be more successful.

¹⁷ Extensible Stylesheet Language Transformations; <http://www.w3.org/Style/XSL/>

3. Conclusions

Policies for digital preservation represent an issue that still needs a lot of attention. Little practical experience yet exists and most of the ideas are still rather theoretical. Although there are organisations that have a relatively longstanding experience with digital preservation, such as data centres, they deal with rather simple digital objects and data formats. The current challenges that are posed by modern technology have such a deep impact on society and the way organisations function and communicate, that they require other and much more comprehensive and sophisticated approaches. The presentations at this seminar showed the broad range of issues involved in this area as well as that there is a firm basis for identifying strategies and formulating policies, but that more practical experience is needed. It all starts, however, with the awareness of the issues at stake at the level of decision makers in organisations. Important in this respect is to have a clear view of the benefits that could be gained, or the other way around, what costs are involved when adequate and coherent policies, procedures and guidelines are not in place. On the other managers are also open to listen to advice that saves money, so that approach can be used as well. In the area of costing the overall opinion was that better models and tools are needed so organisations can get better insight in the financial consequences.

Other crucial issues that were identified by the participants in defining and implementing adequate preservation policies are:

- Preservation policies should be comprehensive, have a broad scope and include not only technical matters, but also organisational, human resource, legal, rights management, access, and intellectual property issues. It also has to be embedded in broader policies on for instance e-government, e-business, or information management. Preservation is just one out of many aspects that have to be dealt with in organisations that are adapting their business to the possibilities that IT offers them. That goes not only for the private sector, but also for government agencies that are under constant pressure to change and adapt their business processes and bring them in line with the higher requirements of citizens. Connecting to business drivers is recommended as a good approach to implement preservation strategies. It is important to note that change will be achieved rather through a long process, than through projects.
- Policies should be in place at different levels, not only in organisations, but also nationally and internationally. An example was given in the Swiss approach. It has to entail also close and/or better collaboration between the different stakeholders that can be identified, such as policymaking bodies, government organisations, memory institutions, possible preservation services companies, software industry and so on. It is not only necessary to share knowledge and experiences, both nationally and internationally, but it may also be possible and even desirable from a cost-effective point of view for instance that certain services in the area of preservation will be carried out in close collaboration. Moreover, if government organisations and private companies will have a shared and much clearer view of what they want, this may

- help them to influence the IT-industry in delivering more appropriate products and tools that enable more efficient preservation.
- The different actors and disciplines involved have to co-operate, such as staff-members who have to carry out tasks and should articulate their needs, IT-people that should provide technical solutions, records managers that should provide guidance on appraisal and retention of digital material and on documenting the digital sources, auditors and controllers that should point out the critical issues for an successful implementation of a policy. The interests of these different groups are or may be different, therefore successful collaboration will depend on things they have in common. It is worthwhile to identify these issues and use them as a basis.
 - Nothing works without sufficient knowledge and adequate skills. Therefore training of all people involved is urgently needed and an essential prerequisite for successful implementation of preservation policies and practices. This has to start with teaching the teachers, an issue that was identified by participants as primary problem. Also important is a good communication and PR strategy that makes clear what issues are at stake, what the benefits, and what the consequences of failure will be.
 - When a policy is in place it is necessary to monitor its application and effectiveness. In order to assess the results criteria and success indicators need to be formulated. Ongoing evaluation of practice will help to refine policies and if necessary to adapt them as appropriate.
 - Finally, participants agreed that more research is needed on these issues as well as that more experience has to be gained from practice in organisations. The initial stage in which the implementation of policies still is, makes it difficult to know what will work or not and what good approaches are.

Appendix One: Seminar Programme

*erp*ataining
Policies for Digital Preservation
Wednesday 29th -Thursday 30th January 2003
Fontainebleau, near Paris; France

Wednesday, 29th January

08:30 *Registration*

09:00 **Welcome**

Martine de Boisdeffre (Archives de France)

09:15 **Introduction to seminar**

Context and objectives

09:30 Christine Pétilat (Le Centre des Archives Contemporaines)

Thomas Schärli (Statistisches Amt des Kantons Basel-Stadt)

10:45 *Break*

11:00 Peter Emmerson (Emmerson Consulting)

Issues to be addressed

11:30 Joël Poivre (Archives de France)

Ellis Weinberger (Cambridge University Library)

12:30 *Lunch*

13:30 Enrica Massella Ducci Teri (Italian public administration)

14:15 *Break*

14:30 **Practical Session**

17:00 *Closing*

Thursday, 30th January

09:00 *Wrap up of previous day*

Impact on organisation

09:15 John McDonald (John McDonald Consulting)

Benoît Riandey (Quetelet Center)

11:00 *Break*

Implementing digital preservation policies

11:15 Neil Beagrie (JISC, DPC)

Hartmut Burghard (Multimédia - EU-Bookshop)

12:30 *Lunch*

13:30 Marie-Claude de la Godelinains (INSEE)

14:00 *Break*

14:15 **Practical Session**

17:00 *Closing*

Appendix Two: Speakers at the Training Seminar

Neil Beagrie is Programme Director for Digital Preservation at the JISC and he co-ordinated the development of a Digital Preservation Coalition in the UK. Previously he was Assistant Director of the Arts and Humanities Data Service, where he developed digital collections policy and standards, and published extensively on digital preservation issues.

Hartmut Burghard is Principal Administrator for the European Commission at the Publications Office in Luxembourg. He works in the field of automated multilingual publications. As part of the mandate of the Publications Office he ventures in issues involved in record keeping and digital preservation.

Peter Emmerson is director and principal consultant of Emmerson Consulting Limited. His consulting practice specialises, amongst other things, in the development of function-based records management systems and has both private and public sector clients. Previously, he was head of Records Services at Barclays Bank, the UK's largest and most comprehensive corporate archives and records management programme, which included strategy and policy development, records retention programmes, and systems development and design.

Marie-Claude de la Godeliniais is head of the archiving section at the national institute of statistics and economical studies (INSEE). Her responsibilities include the dissemination and preservation of statistical data, as well as studies and documents of the institute. Prior to this post she has taken various positions at the INSEE.

John McDonald is an independent consultant specializing in information management. During a career of over 25 years with the National Archives of Canada he held a number of positions that were responsible for the management and preservation of electronic records and for facilitating the management of records across the Canadian federal government.

Christine Pétilat is director of the Centre des Archives Contemporaines since 7 years. She entered the French Archives Nationales in 1974, where she has taken posts as archivist at the social ministries, the archival office of the prime minister, and coordination at various ministries.

Joël Poivre takes a responsible post at the department of technical innovation and standardisation at the Archives de France. He is a qualified archivist, and he has worked extensively on the preservation of digital documents at the Centre des Archives Contemporaines in the scope of the Constance project.

Benoît Riandey is presently Director of the Centre Quetelet, the French Center for Survey Data dissemination, and chairman of the archival committee of the INED, the French Center of Demographic Research. He has a background as a statistician. From 1995 to 2000, he was the Executive Director of the International Association of Survey Statisticians.

Thomas Schärli currently works in the statistical office of the canton Basel-Stadt in Switzerland on the development of a business process and knowledge management programme. He has worked on the development of a national strategy towards the long-term preservation of digital documents in Switzerland, and he is active in a number of other initiatives focussing on digital processes and long-term preservation.

Enrica Massella Ducci Teri currently works at AIPA, the Italian Information Authority for Public Administrations, in the area related to data quality, standard and methodology definition. She has been involved in the working group that has reformulated the Italian law concerning technical rules for long-term digital preservation and exhibition.

Ellis Weinberger is Research Associate at the Cambridge University Library. He has a background in Information and Library Science, and he developed digital preservation policies in the scope of the CEDARS Project, and developed and implemented a migration strategy for the BBC Domesday digital object on behalf of the CAMiLEON Project.

Appendix Three: Participants at the Training Seminar

| <i>name</i> | <i>position, organisation</i> | <i>country</i> |
|--------------------------------|---|-----------------|
| Pierre-Yves Aigrault | AREP | France |
| Andreas Aschenbrenner | Content Editor, ERPANET | The Netherlands |
| Neil Beagrie | Programme Director for Digital Preservation JISC | UK |
| Miguel Beuvier | Chef Restauration-Conservation, ECPAD | France |
| Didier Bondue | Directeur, Saint-Gobain Archives | France |
| Margaret Brooks | Keeper, Sound Archive, Imperial War Museum | UK |
| Hartmut Burghard | European Commission, Publication Office | Luxembourg |
| Cyrille Chareau | Adjoint au chef de service, Service des Archives et de l'Information doc. | France |
| Menehoud du Chatelle | Directrice du Patrimoine Culturel, Hèrmes International | France |
| Claude Chiesa | Archives départ. de Seine-et-Marne | France |
| Ghislain Compreignac | associated manager, Novastrat | France |
| François Danhiez | Chargé de mission informatique, Mission des Archives nationales | France |
| Aranea Dijkmans | Senior Digitisation, Municipal Archives Amsterdam | The Netherlands |
| Enrica Massella Ducci Teri | AIPA, Italian Administration | Italy |
| Peter Emmerson | Emmerson Consulting Limited | UK |
| Jean-Jacques Favreau | archivist, Region Poitou- Charentes | France |
| Emmanuelle Flament-Guelfucci | Chef de service, Service des Archives et de l'Information doc. | France |
| Frédérique Fleisch | Responsable du pôle Archives, ANAES | France |
| Eva Fonss-Jorgensen | Head of National Library Dept., State and University Library | Denmark |
| Marie-Claude de la Godelinains | Institut National de la Statistique et des Études Économiques | France |
| Monica Greenan | Content Editor, ERPANET | Scotland |
| Maria Guercio | erpanet director, Università degli studi di Urbino | Italy |
| Aude Guillon | Archiviste, Mission des Archives Nationales | France |
| Cecília Henriques | archivist, Instituto dos Arquivos Nacionais | Portugal |
| Hans Hofman | Co-Director ERPANET, Nationaal Archief | The Netherlands |
| Anita Hollier | Archivist, CERN | Switzerland |
| Christopher Jack | Manager, Records and Information Service, Audata Limited | UK |
| Delphine Jensen | European Investment Bank | Luxembourg |
| Ulla Kejser | Head of Preservation, Royal Library | Denmark |
| Hélène Lhoumeau | Conservateur, chef de la mission des Archives; Ministère | France |
| John McDonald | John McDonald Consulting | Canada |
| Peter McKinney | Coordinator, ERPANET | Scotland |
| Evelyne Van den Neste | mission des Archives nationales, AN-CAC | France |
| Heike Neuroth | Research & Development, Goettingen State and University Lib | Germany |
| Myriam Pauillac | Head of Bureau des Archives, Conseil Régional de la Guyane | South America |

| | | |
|----------------------|--|-------------|
| Philippe Penicaut | Managing Partner Novastrat Conseil Marketing Strategique | France |
| Christine Pétillat | Director, Centre des Archives Contemporaine | France |
| Joël Poivre | Archives de France | France |
| Mireille Rajinthan | Archives départ. de Seine-et-Marne | France |
| Benoît Riandey | Quetelet Center | France |
| Seamus Ross | Director, ERPANET & HATII, University of Glasgow | Scotland |
| Anne Rossignol | Documentation International Coordinateur, Hèrmes International | France |
| Isabelle Rouge-Ducos | conservatrice du patrimoine | France |
| Roseline Salmon | Curator, C.A.C. Archives nationales | France |
| Thomas Schärli | Statistisches Amt des Kantons Basel-Stadt | Switzerland |
| Daniel Schmutz | Head of Housing & Logistics, Swisscom IT Services AG | Switzerland |
| Jean-Pierre Teil | Head of Constance program, French National Archives | France |
| Mario Tonelotto | Parlement européen | Luxembourg |
| Edouard Vasseur | Curator, Archives nationales – CAC | France |
| Nathalie Vidal | Hèrmes International | France |
| Ellis Weinberger | Cambridge University Library | UK |
| Bakelli Yahia | Researcher, CERIST | Algeria |
| Jean-Daniel Zeller | Archiviste principal, Hôpitaux universitaires de Genève | Switzerland |