



Information Society  
Technologies

# erpastudies

**a r p a t**

---



erpastudies

[www.erpanet.org](http://www.erpanet.org)

ERPANET – Electronic Resource Preservation and Access Network – is an activity funded by the European Commission under its IST programme (IST-2001-3.1.2). The Swiss Federal Government provides additional funding.

Further information on ERPANET and access to its other products is available at <http://www.erpanet.org>.

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>).

ISSN 1741-8682  
© ERPANET 2004

## **Table of Contents**

<b>Executive Summary .....</b>	<b>2</b>
<b>Chapter 1: The ERPANET Project.....</b>	<b>3</b>
<b>Chapter 2: Scope of the Case Studies .....</b>	<b>4</b>
<b>Chapter 3: Method of Working .....</b>	<b>6</b>
<b>Chapter 4: Introduction to the ARPAT .....</b>	<b>7</b>
<b>Chapter 5: Details of the Interviews .....</b>	<b>8</b>
<b>Chapter 6: Analysis .....</b>	<b>9</b>
Perception and Awareness of Digital Preservation.....	9
Preservation Activity.....	10
Compliance Monitoring .....	12
Digital Preservation Costs.....	12
Future Outlook .....	13
<b>Chapter 7: Conclusions .....</b>	<b>14</b>
<b>Appendix 1: References .....</b>	<b>15</b>
<b>Appendix 2: List of acronyms used.....</b>	<b>16</b>

## **Executive Summary**

The ARPAT<sup>1</sup> is the Regional Agency for the Environmental Protection of Tuscany, the first of several regional environmental agencies (ARPA's) created by the Italian Parliament through regional legislation in the mid-1990's.

The ARPAT has only recently begun to express an interest in digital preservation and is currently in the process of reorganising and rationalising its document management system in general. This is taking place in a staggered process that first analyses the existing document management infrastructure, then develops a classification and retention schedule for all important documents and records held by the organisation, complemented by an archival study considering organisational and archival requirements.

This analysis is simultaneously providing a great trigger to stimulate the agency towards a deeper reflection of the needs and concerns for preserving a huge amount of documentation that is born and can only be preserved digitally. Of particular concern within ARPAT are both the scale and the range of data types that must be preserved. Despite this, the ARPAT has yet to implement digital preservation within its IT and archival infrastructure.

---

<sup>1</sup> Agenzia Regionale per la Protezione Ambientale della Toscana.

## **Chapter 1: The ERPANET Project**

The European Commission and Swiss Confederation funded ERPANET Project<sup>2</sup> (Electronic Resource Preservation and Access Network) works to enhance the preservation of cultural and scientific digital objects through raising awareness, providing access to experience, sharing policies and strategies, and improving practices. To achieve these goals ERPANET is building an active community of members and actors, bringing together memory organisations (museums, libraries and archives), ICT and software industry, research institutions, government organisations, entertainment and creative industries, and commercial sectors. ERPANET constructs authoritative information resources on state-of-the-art developments in digital preservation, promotes training, and provides advice and tools.

ERPANET consists of four partners and is directed by a management committee, namely Seamus Ross (HATII, University of Glasgow; principal director), Niklaus Bütikofer (Schweizerisches Bundesarchiv), Hans Hofman (Nationaal Archief/National Archives of the Netherlands), and Maria Guercio (ISTBAL, University of Urbino). At each of these nodes a content editor supports their work, and Peter McKinney serves as a co-coordinator to the project. An Advisory Committee with experts from various organisations, institutions, and companies from all over Europe give advice and support to ERPANET.

---

<sup>2</sup> ERPANET is a European Commission funded project (IST-2001-32706). See [Hwww.erpamet.org](http://www.erpamet.org) for more details and available products.

## **Chapter 2: Scope of the Case Studies**

While theoretical discussions on best practice call for urgent action to ensure the survival of digital information, it is organisations and institutions that are leading the drive to establish effective digital preservation strategies. In order to understand the processes these organisations are undertaking, ERPANET is conducting a series of case studies in the area of digital preservation. In total, sixty case studies, each of varying size, will investigate awareness, strategies, and technologies used in an array of organisations. The resulting corpus should make a substantial contribution to our knowledge of practice in digital preservation, and form the foundation for theory building and the development of methodological tools. The value of these case studies will come not only from the breadth of companies and institutions included, but also through the depth at which they will explore the issues.

ERPANET is deliberately and systematically approaching disparate companies and institutions from industry and business to facilitate discussion in areas that have traditionally been unconnected. With these case studies ERPANET will broaden the scope and understanding of digital preservation through research and discussion. The case studies will be published to improve the approaches and solutions being developed and to reduce the redundancy of effort. The interviews are identifying current practice not only in-depth within specific sectors, but also cross-sectorally: what can the publishing sector learn from the aeronautical sector? Eventually we aim to use this comparative data to produce intra-sectoral overviews.

This cross-sectoral fertilisation is a main focus of ERPANET as laid out in its Digital Preservation Charter.<sup>3</sup> It is of primary importance that disparate groups are given a mechanism through which to come together as best practices for digital preservation are established in each sector.

### *Aims*

The principal aims of the study are to:

- build a picture of methods and match against context to produce best practices;
- accumulate and make accessible information about practices;
- identify issues for further research;
- enable cross-sectoral practice comparisons;
- enable the development of assessment tools;
- create material for training seminars and workshops; and,
- develop contacts.

Potential sectors have been selected to represent a wide scope of information production and digital preservation activity. Each sector may present a unique perspective on digital preservation. Organisational and sectoral requirements, awareness of digital preservation, resources available, and the nature of the digital

---

<sup>3</sup> The Charter is ERPANET's statement on the principles of digital preservation. It has been drafted in order to achieve a concerted and co-ordinated effort in the area of digital preservation by all organisations and individuals that have an interest and share these concerns. [Hhttp://www.erpanet.org/charter.php](http://www.erpanet.org/charter.php).

object created place unique and specific demands on organisations. Each of the case studies is being balanced to ensure a range of institutional types, sizes, and locations.

The main areas of investigation included:

- perception and awareness of risk associated with information loss;
- understanding how digital preservation affects the organisation;
- identifying what actions have been taken to prevent data loss;
- the process of monitoring actions; and,
- mechanisms for determining future requirements.

Within each section, the questions were designed to bring organisational perceptions and practices into focus. Questions were aimed at understanding impressions held on digital preservation and the impact that it has had on the respective organisation, exploring the awareness in the sector of the issues and the importance that it was accorded, and how it affected organisational thinking. The participants were asked to describe, what in their views, were the main problems associated with digital preservation and what value information actually had in the sector. Through this the reasons for preserving information as well as the risks associated with not preserving it became clear.

The core of the questionnaire focused on the actions taken at corporate level and sectoral levels in order to uncover policies, strategies, and standards currently employed to tackle digital preservation concerns, including selection, preservation techniques, storage, access, and costs. Questions allowed participants to explore the future commitment from their organisation and sector to digital preservation activities, and where possible to relate their existing or planned activities to those being conducted in other organisations with which they might be familiar.

Three people within each organisation are targeted for each study. In reality this proved to be problematic. Even when organisations are identified and interviews timetabled, targets often withdrew just before we began the interview process. Some withdrew after seeing the data collection instrument, due in part to the time/effort involved, and others (we suspect) dropped out because they realised that the expertise was not available within their organisation to answer the questions. The perception of risks that might arise through contributing to these studies worried some organisations, particularly those from sectors where competitive advantage is imperative, or liability and litigation issues especially worrying. Non-disclosure agreements that stipulated that we would neither name an organisation nor disclose any information that would enable readers to identify them were used to reduce risks associated with contributing to this study. In some cases the risk was still deemed too great and organisations withdrew.

### **Chapter 3: Method of Working**

Initial desk-based sectoral analysis provides ERPANET researchers with essential background knowledge. They then conduct the primary research by interview. In developing the interview instrument, the project directors and editors reviewed other projects that had used interviews to accumulate evidence on issues related to digital preservation. Among these the methodologies used in the Pittsburgh Project and InterPARES I for target selection and data collection were given special attention. The Pittsburgh approach was considered too narrow a focus and provided insufficient breadth to enable full sectoral comparisons. On the other hand, the InterPARES I data collection methodology proved much too detailed and lengthy, which we felt might become an obstacle at the point of interpretation of the data. Moreover, it focused closely on recordkeeping systems within organisations.

The ERPANET interview instrument takes account of the strengths and weaknesses from both, developing a more focused questionnaire designed to be targeted at a range of strategic points in the organisations under examination. The instrument<sup>4</sup> was created to explore three main areas of enquiry within an organisation: awareness of digital preservation and the issues surrounding it; digital preservation strategies (both in planning and in practice); and future requirements within the organisation for this field. Within these three themes, distinct layers of questions elicit a detailed discovery of the state of the entire digital preservation process within participants' institutions. Drawing on the experience that the partners of ERPANET have in this method of research, another important detail has been introduced. Within organisations, three categories of employee were identified for interview: an Information Systems or Technology Manager, Business Manager, and Archivist / Records Manager. In practice, this usually involved two members of staff with knowledge of the organisation's digital preservation activities, and a high level manager who provided an overview of business and organisational issues. This methodology has allowed us to discover the extent of knowledge and practice in organisations, to understand the roles of responsibility and problem ownership, and to appreciate where the drive towards digital preservation is initiated within organisations.

The task of selecting the sectors for the case studies and of identifying the respective companies to be studied is incumbent upon the management board. They compiled a first list of sectors at the very beginning of the project. But sector and company selection is an ongoing process, and the list is regularly updated and complemented. The Directors are assisted in this task by an advisory committee.<sup>5</sup>

---

<sup>4</sup> See [Hhttp://www.erpanet.org/studies/index.php](http://www.erpanet.org/studies/index.php). We have posted the questionnaire to encourage comment and in the hope that other groups conducting similar research can use the ideas contained within it to foster comparability between different studies.

<sup>5</sup> See [Hwww.erpanet.org](http://www.erpanet.org) for the composition of this committee.



## **Chapter 4: Introduction to the ARPAT**

<http://www.arp.at.toscana.it/>.

In the mid 1990s, the Italian Parliament decided the role of the ARPA's as regional centres for environmental vigilance and control at local level. Tuscany was the first region to introduce such a regional agency and the ARPAT was launched in 1996. The ARPAT is an autonomous agency that offers control, information, research and consultancy services to a range of public and private bodies through a network of laboratories and offices located around the Tuscany provinces. Sustainable development and compatibility of new projects with the existing environment are key features of their activities.

The main activities that the ARPAT performs are:

- a. technical controls, measurements, laboratory analysis, technical elaborations and evaluations, technical documentation and other technical activities linked to environmental protection functions
- b. verification, consultancy, and technical and scientific assistance in environmental matters for public bodies with administrative functions
- c. organisation and management of the regional environmental information system.

Furthermore, the ARPAT is involved in collaborative research and educational activity with three universities in Tuscany, and with several other environmental, social, and economic organisations.

The ARPAT organises both regional and provincial activities. Regional activities are orchestrated by a centrally, with separate provincial departments for each province. Information resources are managed at the regional level by the department for Regional Environmental Information System (S.I.R.A. – Sistema Informativo Regionale Ambientale).

## **Chapter 5: Details of the Interviews**

Initial contact was made with Cecilia Poggetti, an archivist who works for a consultancy working with the ARPAT since 2002 on the management of the information systems. Mrs Poggetti and her team supported all the activities by contacting the ARPAT staff and maintaining constant contact with ERPANET. She contributed directly to the case study and further contacted Eng. Giuseppe Ancilli, responsible for the Technical service of the General Direction S.I.R.A. and Dott. Monica Caponeri (S.I.R.A.).

Contacts took place from March to July 2004.

## **Chapter 6: Analysis**

This section presents an analysis of the data collected during the case study. It is organised to mirror the sequence of topics in the questionnaire.

- Perception and Awareness of Digital Preservation
- Preservation Activity
- Compliance Monitoring
- Digital Preservation Costs
- Future Outlook

### **Perception and Awareness of Digital Preservation**

Generally speaking, interviewees reported that inside the ARPAT there is little perception or awareness of the importance of digital preservation and the problems linked to it. There are a few exceptions to this, all of which are positioned within the S.I.R.A. – the main unit devoted to managing the information system of the ARPA – and which arise from employees function and work scope. This awareness is quite relevant and strategic in the environmental information system department since the recovery and homogenisation of the historical environmental database is crucial to the information system itself. The ability to preserve such databases through time, which essentially contain the organisation's archival records and metadata, is considered fundamental and strategic to the efficacy of the system as an information system.

#### *The main problems*

Although the absence of awareness makes it difficult to identify individual problems, some issues have nonetheless been identified as problematic. A particularly crucial issue for the ARPAT is the existence of a vast range of formats used in the different types of work it undertakes. A typical example of this is in the audio files produced to survey and attest acoustic pollution phenomena. Despite recognition of this problem, a complete list of the possible varieties of formats of production/reception of digital documentation has yet to be prepared; however, staff reported it to be an project for the near future.

#### *Asset value and risk exposure*

As the ARPAT houses thousands of irreplaceable types of records, their loss would have a great consequence. Staff recognise that a complete list of all the various formats managed at the ARPAT would be very useful not only to ensure the entire collection is properly managed and preserved, but also to assess *how many* and *which* types of information can only be available in digital form. Many of the ARPAT's born digital records are incompatible with a paper environment, for example, audio files. In such cases, the added value of the digital information is constituted by its very existence.

Another example of added-value is the increased functionality and simplified management of potentially huge amounts of information when held in digital format. The so-called environmental 'cadastre' are database where it is possible to re-elaborate huge quantities of informative elements, automatically gathered, that is only possible to manage digitally. The environmental 'cadastre' constitute added value from the original scope of the data collection purposes. For example, the data to monitor

air quality are obtained by specific stations and are finalised to take decisions on limiting or reducing environmental pollution. Whilst the original purpose is relative to real time decisions, continually alimenting such a database has the consequence of populating an extremely valuable tool also useful for future means. Another interesting example is that of environmental installations and permits. These data, originally created for administrative purposes to manage the authorisation permits, add to the knowledge base regarding pollution-emitting sources.

At the moment, the ARPAT's technical records are preserved on paper-based medium such action may result in a loss of information or could compromise the readability of the information. All administrative documentation is systematically reproduced and preserved on paper. These activities have been designed with short term preservation of the records in mind; given the fact that the ARPAT itself has only been in existence for a few years, at the moment there is no thinking nor general concern for long term preservation despite the recognised value in the long term preservation of its resources.

Although the ARPAT has not yet developed a preservation strategy for digital information, the risks of losing information whatever the format are known and identified as such:

- Legal risks: The ARPAT can be requested to intervene in litigation for violations to norms in environmental matters
- Decreased efficiency in management of operations: To properly perform environmental control, it is necessary to confront the daily data with the previous ones
- Historical consequences: To elaborate provisions on the changing environmental conditions, it is necessary to have as much historical information as possible

### *Regulatory Environment*

The ARPAT is not subject to any particular legislation that obliges it to preserve its technical documentation. Extensive legislation is being drafted concerning the use and preservation of administrative records which the ARPAT is aware they will have to observe when it comes into force, see the ERPANET case study on the CNIPA<sup>6</sup> for further details.

### **Preservation Activity**

The activities so far performed at the ARPAT have not yet encompassed activities explicitly concerning digital preservation: so far only a project for the reorganisation and automation of the documentary system has been started. This project derives from management needs, specifically to speed up the working processes of the organisation and to recover existing documentation.

At the beginning of this project, there was a belief that just introducing and systematically using digital documents and the registry system would solve the management problem. Only later did it become clear that to achieve digital document and records management, significant organisational and cultural changes were required. The project currently in place is thus being complemented with an archival study to propose methodological solutions for the creation and management of the

---

<sup>6</sup> Available from the ERPANET website: [Hhttp://www.erpanet.org/studies/H](http://www.erpanet.org/studies/H).

organisation's documents and records and interviewees indicated that it will tackle preservation issues at a later stage.

The project has been structured with a double approach: in a first instance the documentary system will be studied and reorganised via the introduction of archival tools such as a classification scheme and a retention schedule, independently of the types of documents, records, and file formats used. In the second instance, the specific problems linked to digital preservation will be tackled, including the use of digital signatures and interoperability of systems.

The ARPAT is unique amongst the ARPAs in implementing this type of project; thus their approach may well act as a model for other ARPAs in the future.

### *Policies and Strategies*

Despite recognition of the long term value of their resources, the ARPAT has yet to develop policies and strategies regarding the long term preservation of their digital resources. This may well be due in part to the relatively young nature of the organisation and their current focus on implementing effective documents management. As digital preservation issues are to be considered at a later stage, perhaps policies and strategies for digital preservation will follow.

Collaboration amongst the various ARPAs on several common matters is very active, and collaboration in this area as well would be desirable. The ARPAT thinks that the possibility to test strategies and tools developed will contribute an increased value to the project and will provide more opportunities to confirm or introduce corrections to their choices so far.

### *Selection*

At the start of the project, a classification scheme and a retention schedule that defines the rules for managing a variety of typologies<sup>7</sup> of files and classified through the classification scheme, was realised for the provincial departments, and has been experimented with and validated for all the periphery structures of the ARPAT. The schemes have been realised following the methodology and the models elaborated at the Scuola Superiore della Pubblica Amministrazione (SSPA)<sup>8</sup> and the most interesting national experiences. Since acceptance of these tools has to happen by decree, both classification scheme and retention scheme will not be implemented until the decree is issued. Furthermore there is a study taking place aiming at drafting the manual that will contain rules on preservation and management of the document management system.

In the ARPAT there is a specific group (Group of Responsibility) who is responsible of implementing and maintaining these tools: this group has been the reference point for the project since the beginning of the project, although their responsibilities will be transferred to a specific internal department when the required organisational changes have taken place. There is in fact already training taking place for people inside the organisation to provide them with the competencies to further advance the project and the updates and maintenance activities that will periodically be performed on the tools developed (i.e., responsible staff for the registry system and for the archive, for each area identified).

---

<sup>7</sup> According to the retention schedule the preservation period is established regardless of the type of support.

<sup>8</sup> Scuola Superiore della Pubblica Amministrazione, [Hhttp://www.sspa.itH](http://www.sspa.itH).

### *Preservation*

As outlined earlier, the ARPAT is not yet applying any digital preservation strategy and has not yet really fully perceived a great need to do so, largely because its documentation is still very recent and the majority of its documentation is still in active use. Furthermore, its legal requirements for preservation of administrative records are met by printing the records to paper and there is no legal requirement for preservation of the technical records. Since no strategy or planning is in place for preservation, ARPAT has not made any decisions regarding technologies or media, although they purport to be aware of the basic options available to them.

For what concerns formats, the ARPAT currently uses pdf, proprietary Microsoft Office formats (doc, xls, etc.) and .csv for its documents, and the ESRI<sup>9</sup> products format for cartography data. The databases have already been migrated from Sybase and Access into an Oracle database, primarily because of a perceived need for increased functionality. Some of the technological decisions taken (such as the Virtual Machine Ware environment for the machine's virtualisation or to virtually operate obsolete applications and operating systems) might facilitate middle term preservation through emulation but the impression is that these activities do not have an adequate planning at the moment

Responsibilities for preservation have not yet been delineated, but there is a feeling that the responsibilities should be assigned to the territorial structures and co-ordinated at central level. Accordingly, the interviewees felt that the function should be assigned, just as happens today with the management of the informative and documentary systems, to the administrative offices of the structures.

### *Access*

The ARPAT has no specific deposit areas for digital preservation but there is redundancy of copies for security reasons.

Access to information preserved on the server is only possible through user authentication. The archiving format is the original format used at the moment of creation.

The problems faced by the ARPAT in terms of access are copyright issues (especially for cartography and orthophoto privacy) and security and access privileges. Once again these problems will be better addressed at a later stage in the project.

### **Compliance Monitoring**

No plan for monitoring the activities has been prepared so far.

### **Digital Preservation Costs**

For what concerns funding, the ARPAT has already allocated a budget to the activities of archival and organisational analysis as well as for the technological part. The ARPAT has a clear knowledge and control of the amount of money so far used for the project.

The project is moving step by step from the organisational challenges to the archival analysis and will finally approach the challenges poses by digital preservation. The

---

<sup>9</sup> An IT society.

budget in use is the one of the S.I.R.A., not a special budget for digital preservation. It was not possible to gather knowledge on the future intentions of management towards the possibility of funding in the next months/years projects more directly linked with digital preservation.

### **Future Outlook**

The ARPAT feels that their strategy for the project will be adequate for the next few years. By organising this project with a natural evolution from the immediate goal of reorganising the document management system towards the analysis of further problems also concerned with managing records, the ARPAT wishes to implement what is essentially a complex and lengthy project.

Interviewees were of the opinion that should more money become available, this would allow for a more systematic and comprehensive approach to tackling all of the necessary issues. It is the intention of the ARPAT to provide more training internally and to deepen the analysis of the specific needs of the agency, so as to elaborate a preservation strategy that also explicitly delineates responsibilities and actions.

Once the project has been finalised according to this vision, the ARPAT feels that it will have set the basis and will have established procedures and strategies to conduct the ordinary activities needed to preserve digital information over time.

## **Chapter 7: Conclusions**

Overall, the ARPAT, thanks to the S.I.R.A., is more aware that it believes that digital preservation is a big concern, especially with regards to resources only available through its databases and especially crucial for environmental concern and monitoring. The practices they have established for the document management illustrate staff awareness that only a scalable approach, solid in its archival and organisational basis, and extremely grounded on the analysis of the starting situation, could give way to implementing a solid framework that could be applied to other types of digital information generated and held within their organisation and that of their sister agencies (ARPAs). The ARPAT is eager to collaborate with similar agencies and if the project results in a solid approach encompassing the organisational, archival, and preservation issues, it may take a leading role amongst other similar agencies, being the only agency of this type currently involved with such concerns.

It was clear to the interviewees that this constitutes only an initial, first step of a long procession that needs significant further considerations and staggered implementation. There are no real policies or strategies for the long-term preservation of digital resources yet and the impression is that the ARPAT's management should be more interested and involved in the activities needed since at the moment long-term preservation is not considered a managerial objective. This would result in greater stability and continuity of funds and the achievement of a greater awareness and commitment in the long term.



**Appendix 1: References**

The portal of the ARPAT can be found at <http://www.arpat.toscana.it/>.

The structure of the ARPAT is described at  
[http://www.arpat.toscana.it/struttura/st\\_chi.html](http://www.arpat.toscana.it/struttura/st_chi.html) (in Italian).

Information about the S.I.R.A. can be found at  
[http://www.arpat.toscana.it/struttura/st\\_sira.html](http://www.arpat.toscana.it/struttura/st_sira.html) (ARPAT: la struttura dell'agenzia.  
Settore tecnico 'Sistema informativo regionale ambientale - SIRA).

**Appendix 2: List of acronyms used**

ARPAT Agenzia Regionale per la Protezione Ambientale della Toscana Regional Agency for the Environmental Protection of Tuscany).

S.I.R.A. Sistema Informativo Regionale Ambientale (S.I.R.A., Regional Environmental Informative System).

SSPA Scuola Superiore della Pubblica Amministrazione (Superior School of the Public Administration).

## CONTACT DETAILS

### **ERPANET Coordinator**

George Service House  
11 University Gardens,  
University of Glasgow  
Glasgow, G12 8QQ,  
Scotland

Tel: +44 141 330 4568  
Fax: +44 141 330 3788  
Coordinator@erpanet.org

## ERPANET STAFF

### **directors**

Seamus Ross, Principal Director  
Niklaus Bütikofer, Co-Director  
Mariella Guercio, Co-Director  
Hans Hofman, Co-Director

### **coordinator**

Peter McKinney

### **editors**

Andreas Aschenbrenner  
Georg Büchler  
Joy Davidson  
Prisca Giordani  
Francesca Marini  
Maureen Potter

[www.erpanet.org](http://www.erpanet.org)