*erpa*workshop

Trusted Digital Repositories for Cultural Heritage

ERPANET Workshop Report, Rome
17–19 November 2003

FINAL REPORT

Trusted Repositories for Preserving Cultural Heritage

Accademia nazionale dei Lincei, Rome, Italy
17-19 November 2003

Accademia Nazionale dei Lincei

Contents

Introduction **4**
Workshop Setting 4
Workshop Structure **5**
Introduction 5
Quality Standards for Creating Digital Resources 6
Actions for capture 7
Storage and repositories: methods and procedures 10
Emerging technologies and applications: case studies and conclusions 12
Conclusions **14**
Appendix One: Workshop programme **15**
Appendix Two: Workshop Participants **22**
Appendix Three: Workshop Speakers **19**

Introduction

Trusted digital repositories present issues unique to the digital environment as well as issues encountered in more traditional repositories. In fact, the establishment and management of a trusted digital repository build on both traditional approaches and innovative research on the electronic environment. In the past few years, great progress has been made in the intellectual and practical definition and realisation of trusted digital repositories. The OAIS model has served as a catalyst, providing common ground for much research and helping the emergence of more unified approaches to digital repository development. Ongoing work is still required in order to resolve issues, such as continuing access, standards development and adoption, but significant strides have been made towards building an effective digital repository model that can be adopted and adapted by different institutions. As institutions and users grow more assertive and confident in their use of electronic resources, digital collections are finally gaining widespread recognition of their status.

On the basis of the concepts and practice developed at the international level and by the main research projects, this ERPANET workshop's aim has been to identify and discuss key scientific, technical, management and policy issues for adequately building and successfully implementing trusted repositories for preserving cultural heritage in different contexts.

Workshop Setting

The workshop was held in Rome, 19-21 November 2003, in the magnificent Corsini Palace. Thanks to the close and productive collaboration with the Centro Linceo Interdisciplinare (Accademia Nazionale dei Lincei) the initiative took place in a prestigious setting and benefited from the excellent organisation of the Accademia Nazionale. Simultaneous Italian and English translation was available during the workshop and experts from the public sector, academia and the research arena came from different countries and traditions and from different digital cultural heritage preservation environments.

The initiative was divided into four work sessions, plus an introductory session, in which Robert Kahn, director of one of the most prominent research centres for digital resource creation, and Tito Orlandi, workshop coordinator on behalf of the Centro Linceo, took part. Kahn and Orlandi discussed the main components and critical issues involved in creating and managing repositories for the preservation of digital resources. In the ERPANET tradition, each session was followed by small working group discussions (*break-out sessions*) about some of the key issues raised by the speakers.

Workshop Structure

Introduction

Maria Guercio, who introduced the presentations in his capacity as ERPANET co-director and workshop organiser, pointed out that the issue of trusted repositories is one of the most complex within the preservation function. This is due both to the insufficient expertise achieved in this field by the institutions traditionally in charge of preservation and to the number of activities and requirements that have to be developed in order to ensure the adequacy and security of repositories. Because of the huge area this topic covers, the workshop identified and discussed only some of the key issues that require investigation: standards for the layout of qualified digital content, criteria for resource acquisition, methods and procedures for archiving and repository management, application technologies and platforms.

The first issue, the content quality of the digital resources chosen for permanent preservation, has, of course, many different characteristics depending on the kind of preserved heritage and on the specific goals of the institutions in charge of the preservation function. With regard to the other issues, examples and discussions focused mainly on the methodological aspects shared across widely different sectors, establishing the basis for a joint comparison and verification. Each workshop session was conducted with reference to actual cases and experiences, which provided material for the analysis and discussion of the various aspects.

The complexity of the topic deserves close analysis and raises many questions, which were partially addressed in the workshop. However, a definitive answer will only be achieved through many more initiatives, many more shared experiences and research projects. Some of the questions – especially those with the most organisational impact – were addressed from the very beginning of the workshop and were discussed primarily in the break-out sessions:

1. Is it possible and/or necessary to agree on one definition of a trusted digital repository?
2. How do different communities see trusted digital repositories?
3. What current implementations may be rightfully called ‘trusted digital repositories’?
4. How are the concepts of reliability, authenticity and trustworthiness interpreted in different contexts and why?
5. How should the roles and responsibilities of the many stakeholders be addressed?
6. Is the potential of trusted digital repositories currently being adequately exploited?
7. What issues have not yet been addressed in trusted digital repository implementations and research?

The two introductory presentations provided an initial answer to some of these questions, especially to those specifically targeting methodological and overall organisational issues. Tito Orlandi addressed the issue of quality control and content standardisation and Robert Kahn conveyed a very rich picture encompassing all principles, components and functions of digital repositories.

Orlandi pointed out that content creators and editors have so far not paid enough attention to the complex issue of assessing, in the new digital environments, the amount and characteristics of the information that identifies and accompanies records in the transition from their traditional environment to that of digital creation and preservation. The first step to take would be to define the stages of the digital resource ‘publication’ procedure clearly, with specific reference to information recognition in relation to its representation. Similarly to what happens in the traditional publishing

world, when reproduction takes place (and therefore throughout the stages needed to ensure digital preservation) it is necessary to identify which elements have to be retained exactly in their original form and which ones may be modified. As the speaker pointed out, an explicit theory on these issues does not even exist in the paper-publishing world, which is simply guided by best practice. Introducing digital tools in this area might generate unfounded trust, unless this trust is paired with awareness of the many key issues that make the problem particularly relevant. One of these issues, for example, is the conceptual distinction made between the two traditional functions, which are partially in contradiction to each other, of the printed edition: 'on the one hand, the reproduction of an authentic text, on the other hand the access to such text on the part of the readers. In the digital environment, the preservation of the information contained in the record and its transmission to a specific addressee.'[1] The most common, and most serious, mistake made in this area is to underestimate the multiplicity of levels of analysis and management of the digital record in relation to its visualisation. Orlandi pointed out that there is no single solution: it may be managed by centralised companies, with large funding, which deal with huge homogeneous *corpora*; or it may rely on distributed work that only requires agreement on formal standards, which ensure interoperability and re-use in different ways and environments.

The choice between centralising functions in large public institutions and creating a reliable network of research and service centres was one of the discussion topics of the workshop and was addressed by many speakers. In the second introductory presentation, Robert Kahn provided the participants, as mentioned above, with an exhaustive examination of the structure, components, functions, management and description systems of digital repositories and preserved resources. In particular, Kahn focused on the management system and architecture of repositories requiring identification of their basic technical components according to the criteria of efficiency and quality: digital objects, unique identifier of structures and data containers developed as a distributed Internet service ('distributed identifier service on the Internet'), able also to locate online repositories holding digital objects, and metadata harvesting tools. The importance of creating a system of federated repositories is seen as a critical success factor both for heterogeneous repositories and for systems of integrated repositories. Among other issues, Kahn addressed the key relevance of metadata management with regard to both their identifying and descriptive function and the conditions of object access and use.

Quality Standards for Creating Digital Resources

As already mentioned in Tito Orlandi's introductory presentation, the quality of the repository and of digital creations and reproductions is based on principles and standards that still have to conform to each sector's specific needs. The first session's speakers analysed this issue in relation to some specific disciplinary and sectoral environments: the creation of electronic publishing digital repositories (Peter Robinson, *The Commedia of Dante Alighieri in Many Versions: An Electronic Edition of Dante's* "Commedia," Manfred Thaller, *From National Cultural Heritages in the Digital World to a Digital European Cultural Heritage*); the production of standardised and structured digital content for archaeological research (Paola Moscati, *Markup Languages for the Preservation and Improvement of Archaeological Information*); certification, authentication, and scientific recognition of publications and results of research activities in the electronic environment (Raul Mordenti, *Publishing of Humanistic Research in the Digital Environment*); new cognitive structures needed for rendering traditional hypertext models in digital form (Claudio Leonardi and Emiliano Degl'Innocenti (*Fili e trame di Arianna nella massa digitale*); and new forms of publishing and the transformations caused by new objects, such as *e-books*, in the relationship between cultural production, use and publishing market (Gino Roncaglia, *From Printed Books to E-books: (Cultural) Costs and Benefits*).

In this first session, Dino Buzzetti (*Digital Libraries and Complex Digital Objects)* focused on the problem of the adequacy, in terms of exhaustiveness and functionality, of the new digital objects in comparison with the traditional ones. Buzzetti examined the issue from the point of view of both

[1] See Tito Orlandi report, http://www.erpanet.org

completeness of information and the requirements needed for a correct reading and interpretation of the available texts. Semantics and syntax today acquire an entirely new role and new potential thanks to mark-up languages. New key issues and new questions, however, have to be addressed, starting with the need to develop scientifically adequate methods for the use of mark-up languages, and in particular the possibility to elaborate rules and tools for automatic processing aimed at promoting the 'formal semantics' of mark-up languages and a representational method in various disciplinary environments in relation to different issues. These issues are:

- characteristics of schemas and DTDs to be identified;
- development of representational forms able to distinguish structures from data models, (understood as algorithms that may be performed using the available data);
- the need to combine the goal of being exhaustive with the goal of being functional;
- and the ability to develop differentiated architectures based on the characteristics of the complex preserved digital documents (born-digital resources, and digital resources formatted for preservation or distribution).

There is also a complexity that cannot be ignored and that has to do with the identity of the digital object, which does not necessarily coincide with the unique identification of the represented document, both because reproductions do not always contain the linked functions, and because different reproductions of the same object become necessary for different functional goals (for example, in XML format and/or in PDF format). Overall, every presentation in the session acknowledged what Dino Buzzetti had discussed, that is, the fact that the content of digital repositories, independently from their characteristics and specific goals, will inevitably include multiple objects and variants linked to an invariable element represented by the informational content. Therefore, in the digital world far more than in the traditional environment, variability and invariability represent the key problem to be addressed with regard to the identity of the document.

The development of XML, that is, of a mechanism for streamlining information, a method for serial representation that allows at the same time to convert the document to text, to bit strings, while also identifying the internal/external informational auto-reference elements, is an important new step in this process, which, of course, still presents many issues that are not yet made clear enough. Examples include the need for functional and scientific verification of the characteristics and qualities of the informational elements used and usable.

Equally important and complex is the goal, pointed out in particular by Paola Moscati, of safeguarding the integrity of the materials, in relation to both their specific complexity and the complexity of the scientific activities of verification, evaluation, and comparison that digital cultural heritage preservation has to ensure. As Manfred Thaller emphasised, cultural resources succeed only when they manage to influence scientific research and, therefore, contribute to its development. It is necessary then that the new generation of scholars pay particular attention to the creation phase of digital objects. It is also impossible to underestimate the huge impact that the new creation, management and use environment will have on the cost issues for resource creation and access. Even the responsibility system in this environment is destined to change, through an increased number of actors and fragmentation of documentation and verification activities, which will no longer be restricted to traditional preservation institutions only (archives and libraries).

Actions for capture

The second session of the workshop tackled the question of preservation and of the problematic relationship between the creation of digital repositories and the need for long-term access to qualified, integral and authentic information. This session and the following ones all addressed the complex issues at stake both from the theoretical point of view of what principles and methods are needed for the realisation of preservation models and from the practical point of view of analysing

examples of implementation. This approach enabled the workshop participants to evaluate how much progress practitioners have actually made in the field in the past few years, and to look at the state of digital preservation and at the achievements of some institutions.

The session focusing on methods and tools for permanent acquisition and preservation of digital resources opened with a case study introduced by Seamus Ross and Adam Rusbridge (*Ingest strategies for digital libraries*), who discussed the University of Glasgow's attempt at creating a digital archive of scientific materials presented at conferences. The presentation focused mainly on key issues and shed light on the project's methodology, discussing the identification of workflows, the need to render complex objects, and the identification of and compliance with descriptive requirements, in particular the metadata necessary to ensure preservation and access. Also discussed was the key issue of the preservation function's feasibility, especially when the implications of manual resource acquisition and the cost/benefit relationship are at play, with specific reference to the time (and therefore human resources) necessary for the correct acquisition of the materials to be preserved. In this case the speakers specifically pointed out the difficulty of dealing with unpublished materials, which are heterogeneous and varied, also due to the application tools used for their creation (published works put together by specialised staff are, instead, more homogeneous). It is also difficult to trace the provenance of unpublished materials, since they are often unstructured and not accompanied by documents showing their origin and supporting their proper description and management. The presentation offered a broad and rather unsettling overview of the problems that repositories in charge of long-term preservation will have to face when managing digital archives that have not undergone proper preparation at their origin and have not been set up for both management and scientific descriptive metadata. Repositories will have to conduct preliminary work, search for the necessary supplementary information through unstructured channels, in some cases retrieve obsolete hardware to execute programs under their native platform, re-establish links among objects that have lost their interrelation information over time, and overcome the difficulty of using proprietary software that is no longer available. The issues identified are a good example of the difficulties that archivists and librarians will increasingly have to face in the coming years, in an environment where traditional problems, such as lack of order and of available information, will be compounded by the issue of having to identify ways to read and access the resources.

The workflow model identified by Rusbridge and Ross was a first opportunity to tackle one of the key issues of the workshop: the management activities essential in the acquisition of digital heritage. These activities are selection, registration, acquisition, virus risk assessment and management, evaluation, physical preparation for acquisition (disk structure, file content, etc.), description, and archiving, i.e., assessment and use of appropriate preservation methods. The speakers paid specific attention to the issues of metadata identification and extraction, and to the different degrees of difficulty presented by the formats and application platforms used at the moment of creation. The presentation raised many research and discussion issues, such as ways of ensuring correct object representation, the possibility of developing a descriptive grammar, and the possibility of identifying and using tools which, although originally created for other purposes, may be able to extract attributes and properties from files. It clearly emerged that there is a need for having file format and metadata archive registers available, as essential components for any preservation strategy. It was also pointed out how difficult it is to assess costs and how important it is to have very well prepared technical staff and metadata available. These metadata have yet to be adequately developed, based on shared metadata classification and preservation rules.

The attempt to develop a systematic approach, in the archival documentation environment, for the creation of digital repositories was presented in rich technical detail and with great attention to methodological issues by Andrew Wilson, who discussed the year-long experience of the National Archives of Australia (*Access across time: How the National Archives of Australia preserves digital records*). NAA's strategy will certainly raise increasing attention on the part of the archival community, both for its realistic and innovative take on the issue and for the systematic effort and theoretical consistency shown in the archives' experience. What the Australian project is interested in is not simply the development of a digital archive, but the creation of the preservation function in

a digital environment. Wilson envisages the preservation function as characterised by the independence of the formats utilised and by the adequate development of the evaluation, transfer, description and search activities. The speaker pointed out that preservation simply cannot be ensured just by the acquisition of the original application systems, which, due to market restrictions, are only sold in formats with reduced application capabilities. Experts have to focus on performance more than on objects since, in comparison with what happens in the traditional environment, digital preservation issues apply not as much to the objects as to the capability of ensuring their use.

In the light of these considerations, NAA's proposed solution has two basic components: maintain a master copy in the form of a simple bit string for each acquired source and build an active preservation intervention to recreate the performance. This kind of intervention, developed through experience with audiovisual materials, requires a preliminary analysis of the essential document components that need to be preserved. In all instances, it is necessary to define open and well-documented formats, which might also be managed through open application tools, for the migration of the materials to be preserved. These formats are based on XML mark-up languages. Wilson also presented the application software developed by the Australian archival administration, XENA.[2]

Robin Dale (*Global registry of digital formats*) addressed another key issue in creating and managing digital repositories with relation to acquisition procedures, i.e., the registration of digital formats. Dale's presentation was very detailed and filled with topics for further discussion, as well as with actual accounts of the experiences of some important institutions in this field. The speaker pointed out that it is crucial to develop and implement a system for identifying and preserving file formats ('a format is a fixed, byte-serialised encoding of an *information model*'). Digital repositories depend heavily on the knowledge and correct use over time of file formats in most part of the management stages (acquisition, storage and access). It is essential to be clear about what the scope of a global registry system is, particularly if we think of how complex its function is ('maintain persistent, unambiguous bindings between public *identifiers* for digital formats and *representation information* for those formats'[3]) with regard to the fragmentation and constant evolution of technology, and also with regard to the need to capture a multiplicity of representation information for each format, i.e., identify relevant technical syntactic and semantic properties, both in descriptive terms and in terms of process ('services and systems using format as input or output'), as well as of management ('provenance'). There are several reasons for supporting the need for such a registration system: need for identification, validation, transformation, characterisation, risk assessment and presentation. The model currently being developed addresses these issues through a network of institutions that have created an ad-hoc committee formed by universities, research centres, and prestigious archives and libraries. The committee has defined an ontology of formats and identified the basic properties for their characterisation. Furthermore, the committee has identified the key services of a registration system: approval on the part of the institutions, with specific attention to the level of technical revision and to the transparency and access levels (public disclosure), inclusive update activity, notification to users of format modifications and evolution (for example, of obsolescence risks), development of format import policy, description and representation information of each format, and transfer to external repositories of the whole registration system, or part of it. Supplementary activities include a format identification service for a specific digital object, through comparison of the profile attributes of the source examined, a validation system able to verify the format of a specific digital object by analysing its attributes, a conversion service from the original format to a new requested format, and a service of automatic

[2] XML Electronic Normalising of Archives. The presentation showed how the Australian solution works: a double check-sum system aimed at authentication, one for the original wrapped bit stream, the other for the normalised version, is in use. In the Australian approach, the actions that concern object management and are necessary for reconstructing and understanding the object's context and use are documented. Attention is also paid to the modalities of XML coding, which aims at developing schemas managed through plug-in modules that define the essence of the documentary source.

[3] Robin Dale report, http://www.erpanet.org

metadata extraction for some formats. The presentation identified some of the key issues of the function addressed:

- The risk that excess granularity, established in a discretionary way (for example, through the identification of local versions and profiles), might cause an abnormal growth of formats for registration.
- The need for a trustworthy and sustainable registration system, so as to encourage the deposit of proprietary information and justify costs, which depend on the system's planned quality and degree of authority.
- The need to define the operative registration system development model, with regard to the double possibility of an independent intervention on the part of the users or a centralised intervention exclusively performed by the repository's technical staff.
- The need to define a management model for a sufficiently reliable system, taking into consideration all the risks linked to its existence within an already operational organisation and evaluating the compatibility of global objectives with the possible national limitations of the repository.

Robin Dale concluded her presentation by saying that the complexity of the function, the many requirements to be met and the important issues to be solved should not discourage the implementation of a system that will assume an increasingly important role, especially in relation to complex long-term preservation needs, as a tool suited to ensure the permanence of resources.

Storage and repositories: methods and procedures

Fynette Eaton (The development of the Electronic Records Archives Program at the US National Archives: a case study), presented the decade-long experience of the National Archives in Washington, one of the longest and most prominent experiences in the field of digital preservation. First, the speaker gave an accurate description of the characteristics of the preservation function developed by NARA and of its evolution over the past decade. She specifically addressed the investigation of theoretical and methodological issues in digital resource management and preservation that has been made possible through the direct participation of the American institution in the research projects InterPARES 1 and 2, through collaboration with the San Diego Supercomputer Center, and through involvement in the development of the ISO OAIS standard in partnership with the National Aeronautics and Space Administration. The presentation then focused on the new operative strategies in the ERA project, designed with the specific goal of creating a dynamic and scalable system for the long-term authentic preservation of digital resources in different formats, in a way that ensures the highest degree of independence from existing specific hardware and software infrastructures. Fynette Eaton pointed out that in the project each of the following archival activities is implemented:

- Acquisition and storage in a secure environment
- Arrangement
- Description
- Access, giving priority to the Web environment
- Selection.

It clearly emerged from the presentation how important the three-year research effort and the decision of assigning highly qualified staff to the project are. The presentation addressed the effort

of analysis and interaction with the market, the scalability of the project, and the capability of easily managed interfaces that may be personalised, as well as informal and temporary documents. These aspects of the project were considered innovative elements in a very complex project and were later discussed among the workshop participants.

Robin Dale then tackled the key issue of digital repository certification and audit (*Certification and audit*), acknowledged by professionals, and by the best scholars, as one of the toughest challenges in digital preservation initiatives. In opening her remarks, the speaker reminded the audience that the Taskforce on archiving of digital information stated in 1996 that a network of reliable and trustworthy organisations is 'a critical component of the digital archiving' and that certification is an essential process for the creation of 'an overall climate of trust about the prospects of preserving digital information'.

Many issues emerge when trying to define the concept of certification and its application in the digital environment for the organisation of trusted digital repositories. According to Robin Dale, independence of the certification process and institutions' authority are essential components. Not every institution that deals with digital heritage will be able to become a certified repository, but will have to buy certified services from third parties, as indicated in the RLG-OCLC report. These third parties will be able to set up and guarantee the use of adequate functional and information models conforming, for example, to the requirements of the OAIS standard. They will also guarantee the advanced and widespread use of standards, security procedures, and financial and technical sustainability of the implemented functions. In the past few years, a working group (http://www.rlg.org/longterm/certification.html) has been focusing on the characteristics of and requirements for certification. This group is supported by important international institutions, mostly North American and British (US National Library, NARA, aerospace agencies, universities and scientific research centres, Internet Archive, RLG, DPC, OCLC) and has identified various typologies of elements that may enable the evaluation and development of a certification system. These elements cover repository attributes, adopted policies, and functions implemented. The group is also working towards identifying a reference structure and towards drafting a documented report on the essential components mentioned above. There exist a series of still unexplored relevant issues: Dale mentioned the existence of different degrees of certification, the possibility of self-evaluation or cross-evaluation processes, the kind of preservation ensured (should it be at the bit level or should it include all functions for the complete rendering of the digital object?), the validity of self-certification (on which the speaker has doubts), and the evaluation of the process costs in relation to its efficiency. With regard to the preservation model, the speaker pointed out, among the minimal requirements identified, the need for correctly archiving objects (i.e., with their context metadata) and for making them searchable and accessible.

Henk Harmsen (*Dissemination of scientific data: the DARE community*) presented the actual case of a digital repository developed in a scientific environment for the dissemination of preserved information. The repository was developed as part of a research and development project conducted by the Dutch government in agreement with national academic institutions. They had the goal of implementing an example and identifying the conditions necessary for each research institution to have its own digital repository, open-access and qualified for a type of preservation that is going to be sensibly different from the one in the traditional environment. The speaker focused mainly on the many difficulties that arise from such an enterprise, with regard to both the management of materials on the Web, and the ways of cooperating in the scientific environment. Other issues concern the development of publication tools, which will have to be based on open standards (such as, OAI, XML and PDF) and will have to use overall communication and sharing strategies adequate to the managed materials, on the basis of criteria aimed at ease of use and flexibility, and consistent with the planned functions. Although it developed an open access system, the DARE project chose a centralised preservation model, assigning all preservation tasks to the Koninklijke Bibliotheek. This decision, however, did not imply that the institutions producing the digital objects would give up their responsibilities: involving the producers in the preservation process is in fact considered as a requirement for success in the project, and the concept of preservation is part of the open-access model. There are still unresolved issues in this area, as it

also emerged from the discussion that followed the presentation. For example, it is still difficult to convey the importance of preservation, of open access (which so far seems to be implemented almost exclusively through actual electronic publications), and of the responsibility and timeframe for the selection of materials.

Emerging technologies and applications: case studies and conclusions

The case studies introduced in the final session provided the opportunity to present and discuss experiences, as well as to return to and further analyse the many issues emerging from the previous sessions.

The results of the LOCKSS programme, presented by Victoria Reich (*Custodianship of cultural web content: the LOCKSS Program*), specifically highlighted the need for supporting traditional preservation centres in this difficult transition phase, also using the possibilities explored by LOCKSS. Among preservation institutions, libraries in particular need help with Web-based content (electronic theses and dissertations, e-government records, grey literature, various documents and publications, and other Web materials). Institutions may be helped through the development of a simple and easy-to-use model, low in cost and reliable, mostly thanks to the use of open-source products and the implementation of procedures based on redundancy of the collected materials, as well as on verification of, and comparison among, the various captured content versions and on the distribution/sharing of the activities, which imply internal forms of continuous validation of materials and guarantee against loss.

This programme, too, presents issues that cannot be ignored:

- It requires authorised editor access.
- It requires certified versions of the captured material.
- It does not allow preservation of the dynamic Web.
- It requires a large amount of archiving memory.
- It has not yet developed a relevant system of alliances.
- It cannot currently go beyond the http environment.

The i-Tor project, discussed by Henk Harmsen in his second presentation (*i-Tor: tools and technologies for open repositories*) also focused on the possibilities and advantages of standards and open and interoperable systems, and of the architecture implemented within the OAI model for the development of a tool for the digital object acquisition required by the DARE project, which was discussed earlier. The issue of the definition of shared metadata schemas, which should also be sufficiently specific and adequate to the complexity of the documentary material, was discussed in depth and then also effectively addressed in the presentations of Sergio Bernardi on the use of DOI technology (*Digital voluntary deposit and preservation: the use of DOI technology*) and of Marco Veneziani (*MAG: a metadata set for managing and preserving Italian digital resources*). In all the cases described, the speakers focused, although using different approaches, on the key role of metadata for the unique identification of resources. Different approaches emerged: the DOI solution aims at permanent location, while the i-Tor solution aims at creating objects in standard formats. The Italian Digital Library adopted a middle-ground solution for the management and preservation of digitised objects, although, at least in this preliminary phase, its preference seems to be for permanent location, rather than for the creation of a surrogate in standard format.

Stephan Heuscher (*Today's design of tomorrow's trust in digital archives. Considerations from the ARELDA project*) tackled the issue of authenticity in digital archives/repositories in long-term

preservation projects, through considerations arising from the experience of the ARELDA project, conducted by the Swiss national archives.

The speaker first clearly defined the difference between digital libraries and digital archives (where the information value coexists with an additional evidential value), then listed the essential characteristics that guarantee the reliability of the documentary heritage in traditional environments (integrity, intelligibility, authenticity and accessibility). In digital documents, these characteristics present some contradictory elements, which at the current stage have not yet been resolved by the available technical infrastructures, which can guarantee data security only for a limited time, as happens with digital signatures. Heuscher pointed out that, among the difficulties raised by the authentic preservation of digital resources, a particularly relevant issue is that of loss of authenticity characteristics due to the inevitable migration processes. This issue can only be addressed through full documentation of the preservation process, definition of rigorous and automatic procedures, retention of all context data that may guarantee verification, intelligibility, and use of the resources over time, and development of adequate infrastructures that are sustainable at the financial and management level.

Conclusions

Although the workshop's presentations addressed most key issues, many of them still require further discussion and analysis, as it clearly emerged from the break-out sessions at the end of the second and third day, as well as from discussions throughout the workshop. The fruitful and innovative debate that took place helped to identify some recommendations aimed at promoting research and positive experiences in an area that is becoming increasingly important for the creation and management of digital resources. In brief, some of the conclusions reached by the workshop participants are:

- the need to raise awareness of the relevance of the preservation issue among stakeholders, with the goal of turning – as soon as possible – the occasional voluntary deposit of materials in trusted repositories into a good habit;
- the importance of developing schemas and tools that facilitate the deposit of objects and their description through the implementation of automatic metadata extraction systems;
- the relevance of flexibility criteria for the increasingly complex functions that apply to the preservation of and access to digital heritage, requiring new mediation levels;
- the need to develop:
    - a broad range of standards;
    - solutions integrated with other registration systems (of formats and metadata schemas);
    - metadata extraction tools;
    - migration tools;
    - a minimum required level for description attributes;
    - deposit agreements resulting from an accurate negotiation and clearly identifying responsibilities.

Among the specific recommendations developed through the workshop, the need to prompt preservation institutions to be more involved in activities emerged as one of the most prominent. These activities may range from illustrating the benefits of new services to developing levels of service consistent with the quality of the deposited objects, from establishing priorities in the production of standards and in the normalisation processes to setting up networks of repositories able to support and integrate with each other.

Appendix One: Workshop programme

erpaworkshop

Trusted Repositories for Preserving Cultural Heritage

Rome, 17-19 November 2003

Accademia dei Lincei – Centro Linceo Interdisciplinare, Fondazione Ezio Franceschini
Villa Farnesina, 230 Via della Lungara, Rome

Monday 17 November

09:00	*Registration*		
09:30	**Welcome address**		
SESSION ONE **Introduction**			Chair: Maria Guercio
09:30	Introduction and welcome	Maurizio Brunori (Centro Linceo Interdisciplinare, Rome, Italy)	
10:00	Archives and information	Tito Orlandi (Centro Linceo Interdisciplinare, Rome, Italy)	
10:30	Trusted digital repository: an introduction	Robert Kahn (Corporation for National Research Initiatives, CNRI, Virginia, USA)	
11:00	*Coffee break*		
SESSION TWO **Quality Standards for Creating Digital Resources**			Chair: Tito Orlandi
11:30	Fili e trame di Arianna nella massa digitale	Claudio Leonardi, Francesco Santi and Emiliano Degl'Innocenti (Fondazione Ezio Franceschini)	
12:00	Presenting a text in many versions: an electronic edition of Dante's Commedia	Peter Robinson (De Montfort University, Leicester, UK) and Prue Shaw (University College London, UK)	
12:30	Digital libraries and complex digital objects	Dino Buzzetti (Università degli Studi di Bologna, Bologna, Italy)	
13:00	*Lunch*		
15:00	From national cultural heritages in the digital world to a Digital European Cultural Heritage	Manfred Thaller (University of Köln, Germany)	
15:30	Markup languages for the preservation and improvement of archaeological information	Paola Moscati (Centro Nazionale delle Ricerche, Rome, Italy)	
16:00	*Coffee break*		

16:30	From printed books to e-Books: (cultural) costs and benefits	Gino Roncaglia (Università della Tuscia, Viterbo, Italy)
17:00	Publishing of humanistic researches in digital environment	Raul Mordenti (Università di Tor Vergata, Rome, Italy)
17:30	**Break-out session for discussion on session 2**	
18:15	*End of session 2*	

Tuesday 18 November

SESSION THREE **Actions for Capture**		Chair: Seamus Ross
09:30	Ingest strategies for digital libraries	Seamus Ross (Director HATII) and Adam Rusbridge (ERPANET Researcher)
10:15	Access across time: how the National Archives of Australia preserves digital records	Andrew Wilson (National Archives of Australia, Australia)
11:00	*Coffee break*	
11:30	Global registry of digital formats	Robin Dale (Research Libraries Group, USA)
12.15	**Break-out session and discussion on session 3**	
13:15	*Lunch*	
SESSION FOUR **Storage and Repositories: Methods and Procedures**		Chair: Hans Hofman
14:30	Certification and audit	Robin Dale (Research Libraries Group, USA)
15:15	Repository design and creation: a case study	Fynette Eaton (US National Archives & Records Administration, USA)
16:00	*Coffee break*	
16:15	Dissemination of scientific data: the DARE community	Henk Harmsen (DARE Project, the Netherlands)
17:00	**Break-out session and discussion on session four**	
17:45	*End of session 4*	

Wednesday 19 November

SESSION FIVE **Emerging Technologies and Applications (Case Studies)**		Chair: Maria Guercio
09:00	Custodianship of Cultural Web Content; the LOCKSS Program	Victoria Reich (Director LOCKSS programme, Stanford, USA)
09:30	i-Tor: tools and technologies for open repositories	Henk Harmsen (NIWI-KNAW, the Netherlands)
10:00	DSPACE	Nick Wainwright (HP Laboratories, Bristol, UK)
10:30	Today's Design of Tomorrow's Trust in Digital Archives: considerations from the ARELDA project	Stephan Heuscher (Swiss Archives, Switzerland)
11:00	*Coffee break*	

11:30	Digital voluntary deposit and preservation: the use of DOI technologies	Gabriella Scipione (CINECA Project, Bologna, Italy)
12:00	MAG: a metadata set for managing and preserving Italian digital resources	Marco Veneziani (Lessico Intellettuale Europeo, CNR, Italy)
12:30	Lunch	
14.00	**Breakout session and discussion on session 5**	
14:45	*Conclusions of the workshop*	

Appendix Three: Workshop Speakers

Tito Orlandi	Archives and Information
Robert Kahn	Trusted digital repository: an introduction
Claudio Leonardi	Fili e trame di Arianna nella massa digitale
Peter Robinson, Prue Shaw	Presenting a text in many versions; an electronic edition of Dante's *Commedia*
Dino Buzzetti	Digital libraries and complex digital objects
Manfred Thaller	From national cultural heritages in the digital world to a digital cultural heritage
Paola Moscati	Markup languages for the preservation and improvement of archaeological information
Gino Roncaglia	From printed books to e-Books: (cultural) costs and benefits
Raul Mordenti	Publishing of humanistic researches in digital environment
Seamus Ross, Adam Rusbridge	Ingest strategies for digital libraries
Andrew Wilson	Access across time: how the National Archives of Australia preserves digital records
Victoria Reich	Custodianship of cultural Web content; the LOCKSS programme
Robin Dale	Global registry of digital formats
Robin Dale	Certification and audit
Fynette Eaton	Repository design and creation: a case study
Henk Harmsen	Dissemination of scientific data: the DARE community
Henk Harmsen	i-Tor: tools and technologies for open repositories
Nick Wainwright	DSPACE

Stephan Heuscher Today's design of tomorrow's trust in digital archives: considerations from the ARELDA project

Sergio Bernardi Digital voluntary deposit and preservation: the use of DOI technologies

Marco Veneziani MAG: a metadata set for managing and preserving Italian digital resources

Appendix Two: Workshop Participants

Sara Adamo	Università degli Studi di Milano	Italy
Alessandra Amati	CNIPA	Italy
Piero Attanasio	CINECA Project	Italy
Blandine Blukacz-Louisfert	Records and Archives Unit	Switzerland
Robin Boast	Museum of Archaeology and Anthropology, University of Cambridge	UK
Marco Boschetti	CSI Piemonte	Italy
Filip Boudrez	DAVID Project researcher	Belgium
Maurizio Brunori	Centro Linceo Interdisciplinare	Italy
Paolo Buonora	State Archive of Rome State Archivist	Italy
Dino Buzzetti	University of Bologna	Italy
Daniela Canali	Doctoral researcher (CNR)	Italy
Sheila Carey	Canadian Heritage Information Network	Canada
Rita Carrarini	Istituto centrale per la patologia del libro	Italy
Gloria Cirocchi	Parliamentary Library documentalist	Italy
Claudia Colagiovanni	Gruppo ENEL	Italy
Giovanni Conso	Accademia Nazionale dei Lincei	Italy
Ugo Contino	Library Automation Group (CASPUR)	Italy
Paola Coppola	Università di Roma Tor Vergata, librarian	Italy
Robin Dale	Research Libraries Group (RLG)	USA
Roberto Dallari	Parliamentary Library Counsellor	Italy
Michael Day	UKOLN, University of Bath	UK
Joy Davidson	University of Glasgow ERPANET British Editor	UK
Maria Teresa De Gregori	Università di Roma Tor Vergata, Assistan librarian	Italy
Gaspard De Jong	Netherlands Institute for Scientific Information	The Netherlands
Assunta Di Febo	Istituto centrale per la patologia del libro	Italy
David Donald	Glasgow Caledonian University	UK
Giorgio Dori	ICT Consultant	Italy
Fabio Del Giudice	Banca di Roma	Italy
Fynette Eaton	US National Archives & Records Administration (NARA)	USA
Lesley Ferguson	Royal Commission on the Ancient & Historical Monuments, NMRS Curator	UK
Daniela Ficini	Università di Roma Tor Vergata, Librarian	Italy
Paola Gargiulo	Library Automation Group (CASPUR)	Italy
Ivan Grossi	Consultant	Italy
Monica Grossi	University of Urbino	Italy
Maria Guercio	University of Urbino ERPANET Co-director	Italy
Henk Harmsen	Netherlands Institute for Scientific Information, Assistant Director	The Netherlands
Stephan Heuscher	Swiss Archives	Switzerland
Peter Horsman	Archiefschool, Research Coordinator	The Netherlands
Robert Kahn	Corporation for National Research Initiatives	USA

Jaap Kloosterman	International Institute of Social History Director	The Netherlands
Branislav Kovacevic	Open Society Archives	Hungary
Germana La Favia	Banca di Roma	Italy
Peter Lazaridis	Aristotelian University, Librarian	Greece
Claudio Leonardi	University of Florence	Italy
Liga Locane	Latvia State Archives	Latvia
Irene Lourdi	National & Kapodistrian University of Athens, Archivist & Librarian	Greece
Sigrun Magnusdottir	University of Akureyri, Director of Information Services	Iceland
Gianfranco Maiozzi	Ministero delle Comunicazioni, Librarian	Italy
Peter McKinney	University of Glasgow ERPANET Coordinator	Italy
Enrica Masselli Ducci Teri	CNIPA	Italy
Giovanni Michetti	University of Urbino	Italy
Raul Mordenti	Università di Tor Vergata	Italy
Paola Moscati	Centro Nazionale delle Ricerche (CNR)	Italy
Paola Franca Munafò	Istituto centrale per la patologia del libro, Responsabile della didattica	Italy
Samir Musa	Università degli Studi di Urbino *"Carlo Bo",* ERPANET Italian Content Editor	Italy
André Müller	Himmel Blau, Researcher	Switzerland
Erik Oltmans	Koninklijke Bibliotheek	The Netherlands
Tito Orlandi	Centro Linceo Interdisciplinare	Italy
Maria Clara Peira	CSI Piemonte	Italy
Ortrun Peyn	Society of Antiquaries of London, Head of Library Cataloguing	UK
Garcia Raúl	National Direction of Archives of Cuba	Cuba
Victoria Reich	LOCKSS Programme director	USA
Kelly Richmond	American Academy in Rome	Italy
Peter Robinson	De Montfort University , Leicester	UK
Gino Roncaglia	Università della Tuscia	Italy
Seamus Ross	University of Glasgow ERPANET Director	UK
Adam Rusbridge	University of Glasgow	UK
Raivo Ruusalepp	Estonian Business Archives, Ltd, Consultant	Estonia
Ellis Sada	Università Cattolica del Sacro Cuore, Librarian director	Italy
Gabriella Scipione	CINECA Project	Italy
Goutam Kumar Sanyal	Parbatya Bouddha Mission (PBM), Programme Adviser	Bangladesh
Dirk Scholz	NESTOR c/o Bayerische Staatsbibliothek	Germany
Dr Prue Shaw	University College London	UK
Dorothea Steiner	Assistant in research project	Germany
Andrea Strambio De Castillia	Centro per la cultura d'impresa	Italy
Maria Teresa Tanasi	Centro di Fotoriproduzione Legatoria e Restauro	Italy
Jean-Pierre Teil	French National Archives, Head of Constance programme	France
Manfred Thaller	University of Köln	Germany

Emmanuel Tetteh Thompson	Love and Peace Development Foundation	Ghana
Brian Tingle	California Digital Library, Content Management Designer	USA
Esther Trippel-Ngai	United Nations Office in Geneva, UNOG Registry, Records and Archives	Switzerland
Roberta Valente	HR Consulting, Knowledge Manager	Italy
Marco Veneziani	LEXICON	Italy
Francesca Ventura	Gruppo ENEL	Italy
Stefano Vitali	State Archives of Florence, State archivist	Italy
Nick Wainwright	HP Laboratories	UK
Andrew Wilson	National Archives of Australia	Australia
Sara Zanisi	Centro per la cultura d'impresa	Italy
Margaret Zito	FAO Library	Italy