

An Introduction to ERPANET



DPC Forum
23 June 2004

Seamus Ross
Director HATII, University of Glasgow &
Principal Director ERPANET

Introduction

ERPANET works to enhance the preservation of cultural and scientific digital objects through raising awareness, providing access to experience, sharing policies and strategies, improving practices, investigating sectoral preservation behaviours, and conducting research.

ERPANET EU FP5 Project

- ERPANET aims to do the following:
 - raise awareness of preservation needs and strategies;
 - appraise and evaluate information sources and developments in digital preservation and increase the accessibility of this work;
 - provide an EU-wide enquiry and advisory service on preservation issues, practice and technology;
 - organise 15 developmental seminars and workshops across EU (so far 20 leading organisations and more than 100 professionals contributed)
 - stimulate research in digital preservation;
 - promote awareness among software producers of the preservation needs of the user community; and
 - research to improve method and practice.

The Partners

Institutions:

- The Humanities Advanced Technology and Information Institute (HATII)
- Schweizerisches Bundesarchiv
- Nationaal Archief van Nederland
- Institute for Archival and Library Science (Università degli Studi di Urbino)

The Directors and Co-ordinator

- The Directors
 - Niklaus Bütikofer (Schweizerisches Bundesarchiv)
 - Maria Guercio (ISTBAL, Università di Urbino)
 - Hans Hofman (Nationaal Archief Nederland)
 - Seamus Ross (HATII, University of Glasgow)
- The Co-ordinator
 - Peter McKinney (HATII, University of Glasgow)



ERPANET & HATII -- Seamus Ross

5

Organisations Need Help



Maria Guercio and John McDonald at the Preservation Policy and Procedures co-sponsored with the Archives de France – Le Centre des Archives Contemporaines, Fontainebleau, January 2003.

- Off-the-shelf policy statements
- Business cases & strategies
- Digestible guidance on technologies and their preservation implications
- Improved models (reference, costs, standards, functional requirements)
- Simple Guidelines on digital survival
- Guidance on creating data repositories
- IPR support and guidance

6

OBJECTIVES

Thematic Workshops & Seminars

- Exploring core topics with experts and specialised institutions
- Increasing the level of awareness within the communities of digital information providers, practitioners, experts and preservers
- Developing basic recommendations
- Providing documentation of new research directions and results
- Provide skills development opportunities to the community



Speakers at the Lisbon Seminar on *The Selection, Appraisal and Retention of Digital Scientific Data*, Weber Amaral, William Anderson, Terry Eastwood, John Faundeen, Pedro Fernandes, Luigi Fusco, Francoise Genova, Myron Gutmann, Gail Hodge, Jürgen Knobloch, Meredith Lane, Seamus Ross, Kevin Sch?rer, Alex Szalay, Peter Weiss

Workshops & Seminars

- Workshops & Seminars, June 2002 – October 2003
 - *Digitisation, Conservation and Quality Control* (Toledo, June 2002)
 - *XML for digital preservation* (Urbino, October 2002)
 - *Introduction to the OAIS Model* (Copenhagen, November 2002)
 - *Preservation Policy and Procedures* (Fontainebleau, January 2003)
 - *Long term preservation of databases* (Bern, April 2003)
 - *Preserving the Web*, (Kerkira, May 2003)
 - *Metadata in Digital Preservation*, (Marburg, September 2003)
 - *ERPANET 'DSpace' Installation Workshop*, (Glasgow, September 2003)

WORKSHOPS & SEMINARS

- Workshops & Seminars, November 2003 – June 2004
 - *Digital repositories: audit and certification*, (Rome, November 2003)
 - *Preservation of Scientific Data Sets* (Lisbon, December 2003)
 - *Audit and Certification* (Antwerp, April 2004)
 - *File Formats* (Vienna, May 2004)
 - *Persistent Identifiers* (Cork, June 2004)
 - Business Models (Amsterdam, September 2004)
 - Preserving Digital Art (October 2004)
 - Workflow (Budapest October 2004)



Peter Robinson and Manfred Thaller at the ERPANET and the Accademia dei Lincei workshop on *Trusted Repositories for Preserving Cultural Heritage* (Rome, November 2003)

Sponsoring organisations

- The Biblioteca Nacional España
- Caja Castilla la Mancha
- Università degli Studi di Urbino,
- The Royal Library, Denmark,
- Archives de France – Le Centre des Archives Contemporaines (Fontainebleau),
- Schweizerisches Bundesarchiv
- Trivadis AG,
- Ionian University Kerkira,
- Archivschule Marburg,
- Accademia dei Lincei
- Ministero per l'università e la ricerca (Italy)
- Cambridge University and Library
- MIT (USA)
- CODATA
- City Archives of Antwerp
- National Library of Austria
- University of Cork



Participants at the Italian Presidency meeting co-sponsored with ERPANET *Futuro delle Memorie Digitali e Patrimonio Culturale*, October 2003.



Participants at the *File Formats* Seminar in Vienna, May 2004

Nearly 100 Speakers since June 2002

- Speakers included (sample, see erpanet website for full list):

Palle Aagard, IT and Telecom Agency, Denmark; Rolf Apweiler, European Bioinformatics Institute; Kevin Ashley, University of London Computer Centre; Carlo Batini, AIPA, Authority for IT in the Public Administration, Italy; Neil Beangie, DPC; Julián Bescós, Informática El Corte Inglés; George Bokos, Ionian University Corfu; Filip Boudrez, DAVID project, City Archives of Antwerp; Hartmut Burghard, Multimedia - EU Bookshop; Niklaus Bütikofer, Swiss Federal Archives; Andrew Charlesworth, University of Bristol; Vassilis Chrysikopoulos, Ionian University Corfu; Salvatore Costa, Ministry of Finance, Italy; Julia Craig-McFeely, Digital Image Archive; of Medieval Music, Oxford University & Royal Holloway University of London; Robin Dale, RLG; Michael Day, University of Bath; Wendy Duff, University of Toronto; Fynette Eaton, NARA; Peter Emmerson, Emmerson Consulting; Pierluigi Feliciati, Italian National Archives / MINERVA; Victoria Fenner, English Heritage; David Giarretta, CCLRC Rutherford Appleton Laboratory; Marie-Claude de la Godelinails, Institut National de la Statistique et des Etudes Economiques; Lars-Erik Hansen, Swedish Social Insurance Administration; Stephan; Heuscher, Swiss Federal Archives; Hans Hofman, National Archief van Nederland; David Holdsworth, CEDARS / University of Leeds; Stephan Jäermann, Swiss Federal Archives; Peter Keller-Marxer, Swiss Federal Archives; Steve Knight, National Library of New Zealand; Greg LaMotta, NARA; Julien Masanès, Bibliothèque nationale de France; Enrica Massella Ducci Teri, Italian Public Administration; John McDonald, John McDonald Consulting; Antonio Mené, Italian Chamber of Deputies; Denis Minguillon, CNES, France; Reagan W. Moore, San Diego Supercomputer Center; Heike Neuroth, Göttingen State and University Library; Christine Pétillat, Centre des Archives Contemporaines; Hans Petschar, Porträtsammlung, Bildarchiv und Fideikommissbibliothek, Austrian National Library; Terje Pettersen-Dahl, National Archives of Norway; Joël Poivre, Archives de France; Maureen Potter, Digital Preservation Testbed; Andreas Rauber, Vienna University of Technology; Enrico Rendina, Consorzio Roma Ricerche; Marco Rendina, Consorzio Roma Ricerche; Benoit Rlandey, Quetelet Center; Bill Roberts, Tessella; Seamus Ross, HATII, University of Glasgow; Thomas Schirli, Statistisches Amt des Kantons Basel-Stadt; Dietrich Schüttler, Phonogrammarchiv, Austrian Academy of Sciences; Gabriella Scipione, CINECA; Najla Sample, Edinburgh University Library; Thomas Severiens, Institute for Science Networking; Elizabeth Shepherd, University College London; Annette Siegenthaler, Die Deutsche Bibliothek; Jean-Pierre Tell, Centre des Archives Contemporaines; Malcolm Todd, National Archives, UK; Remco Verdegem, Digital Preservation Testbed; Ellis Weinberger, Cambridge University Library; Andrew Wilson, National Archives of Australia; Deborah Woodyard, British Library; Thomas Zürcher Thier, Swiss Federal Archives



Group B at the *File Formats Seminar* in Vienna, May 2004

ERPANET & HATII -- Seamus Ross

11

BENEFITS of Workshops & Seminars

- Creation of authoritative sources of information
- Fora for information and preservation professionals that can be extended into their own and other communities
- Exchange of experiences; successes and failures
- Post-event analytic report recommending areas for research, development and service provision following each event



Mr. Allemann, Trivadis AG Switzerland co-sponsor of the ERPANET Bern Workshop on *Long term preservation of databases*, April 2003.

ERPANET & HATII -- Seamus Ross

12

Some Lessons learned

- Community hungry for practical case studies and reports of real world experiences
- Lack of good examples makes it difficult to identify speakers with expertise, technical knowledge of a high enough calibre and suitable presentation skills
- Audience has a high and increasing skill-base
- Discussion with experts through break-out sessions in hot demand
- Cross-domain audience helps to overcome the barriers between domains
- Evaluations show that ERPANET events are in demand and participants in one event intend and do return to others



ERPANET Bern Workshop on *Long term preservation of databases* at the Schweizerisches Bundesarchiv, April 2003.

ERPANET & HATII -- Seamus Ross

13

Content Development Instruments

- Literature Evaluation Mechanisms
- Commentaries
- Case Study Instruments
- Case Study Reports
- Evaluation of erpaEvents
- Workshop and Seminar Reports
- Materials from all Workshops & Recordings of most



Alex Szalay, The Johns Hopkins University (USA) speaking at the ERPANET Lisbon Seminar on *The Selection, Appraisal and Retention of Digital Scientific Data*, December 2004.

<http://www.erpanet.org>

ERPANET & HATII -- Seamus Ross

14

erpaCOMMENTARIES

- Authoritative online information services on Digital Preservation
 - Analytic Reviews of published and online literature
 - including assessment of 100 journals
 - Assessment of technical guidelines and relevant standards
 - Projects reviewed



Participants at the ERPANET Seminar on Metadata for Digital Preservation, Marburg (September 2003)

erpaTools

- Policy Building Framework
- Costing Digital Preservation Strategies
- Digital Object Risk Management Methods
- Selecting Technologies
- Ingest Strategies
-



Michael Day of UKOLN speaking at the ERPANET Seminar on Metadata for Digital Preservation, Marburg (September 2003).

ERP AePrints

- Ensuring long term accessibility to digital preservation literature, both gray and published
- Creating a single point of accessibility
- Better profiling of the literature



Andrew Wilson of the National Archives of Australia speaking at the ERPANET Seminar on *Metadata for Digital Preservation*, Marburg (September 2003).

erpaAdvisory



Robert Kahn (Corporation for National Research Initiatives) and Tito Orlandi (Centro Linceo Interdisciplinare) at the ERPANET and the Accademia dei Lincei workshop on *Trusted Repositories for Preserving Cultural Heritage* (Rome, November 2003)

- Provides the community with access to answers to digital preservation expertise
- Building a Resource of Authoritative Answers to Digital Preservation Questions
- Making accessible examples of best practice
- Currently under-utilised—we are investigating why

erpaStudies – their aims



Stuart Weibel of OCLC at the grave of George Boole following the ERPANET seminar on *Persistent Identifiers*, Cork, June 2004

- Build a picture of current methods matched against their context of use underpin the definition of best practices;
- Accumulate and make accessible information about how approaches to digital longevity;
- Produce materials for seminars and workshops;
- Enable comparisons of approaches across sectors;
- Enable the development of assessment tools;
- Identify issues for further research.

Case Studies Parameters

- Case studies in public and private sectors
- Aim to achieve spread across organisational size and location
- Diversity of organisational type, activity, regulatory framework, and culture
- Perception across records and resources
- Outcomes:
 - Vertical View: reviews of how different sectors approach digital preservation
 - Horizontal View: looking across a range of topics what are best practices across domains.

Case Study Practices

- Research by questionnaire and interview (normally by telephone)
 - A team of six research assistants in four countries
- Examined how other projects had conducted surveys to inform our own design
 - The Pittsburgh Project and InterPARES
- Designed focused template which we hope other projects will adopt when conducting similar studies
- Interviews with three staff members in each organisation sought: Business Manager, IT Manager, Archivist/Records Manager

Case Studies: The Sample

- Drawn from 400 organisations we have contacted
- Current participation rate: just under 20% of the 400
- Many initially said 'yes' and then said 'no'
 - “currently developing a digital preservation programme and at an early stage”
 - “no time and capacity for participation”
 - “too early to participate”
 - Pulled out when seeing questionnaire: common excuse, no one able to complete form
 - Dropped out after reading report (e.g. BNF, International Court of Justice)—occasionally on legal advice

Focus of Investigations

- Preception and awareness of risk posed by digital materials
- Understanding how digital preservation affects the organisation
- Identifying what actions have been taken to prevent loss of digital materials
- Develop an understanding of the way preservation activities are monitored (policies or programmes)
- Understand how organisations plan future preservation requirements (e.g. costs)
- What technologies are in place

ERPANET & HATII -- Seamus Ross

23

Perception and Awareness

- Awareness varied across both public and private sector institutions (e.g. broadcasters high possibly reflecting reuse recognition, drug companies high but for economic and compliance reasons)
- Some sectors did not see the benefits or were trying to identify where responsibility should lie (e.g. publishers)
- Values attributed to digital materials varied:
 - Reuse, evidential value, and risk reduction—top reasons
 - Cultural and historical value given lowest priority

ERPANET & HATII -- Seamus Ross

24

Policy, Strategy and Standards

- Few organisations had developed policies
- Where there were policies interviewees acknowledged they were not implemented across organisations
- Standards and guidelines developed by other communities were not attractive
- Few organisations articulate preservation strategies as part of system and software specifications

Selection, Preservation & Costs

- Selection of digital information for preservation
 - Solid theoretical awareness of issues and the need to determine retention, but little application of appraisal
 - A small number of interviewees were being pressed 'to keep everything and let the technology solve the problem'
- Preservation & Storage
 - Respondents noted the complexity, diversity and size objects
 - Proliferation of formats seen as a problem
 - Most sectors showed little impetus to develop longevity solutions
 - Pragmatic approach is the most popular (e.g. migration on demand)
- Digital Preservation Costs
 - Very limited understanding of costs
 - Poorly predicted and not part of revenue expenditure

Case Studies: General Findings (1)

- Diversity of strategies but few have them
- Activity is fragmentary
- Variation in awareness of risk (many only following our survey)
- Responsibility rarely taken at corporate level
- Recognised need for better sharing of knowledge
- Collaborative action rare (broadcasting is an exception)
- Organisations are waiting for external solutions
- Practices tend to be: *incomplete, ad hoc, and individual*

Case Studies: General Findings (2)

- Variation in the value of information depends upon recognition of business/organisational dependency or risk associated with information
- Selection criteria from the paper world have been transferred to the digital environment
- Preservation and storage poorly understood
- Costs difficult to justify in the corporate environment
- Strong sectoral cooperation leads to increased confidence and action (e.g. broadcasting)
- Need better sharing of knowledge
- Lack of a single point of contact—no one knows who knows or is responsible

Case Studies: Objectives Achieved

- Accumulated materials to illustrate methodologies and practices
- Created a basis for establishing practical guidelines
- Raised awareness among more than 400 institutions and individuals
- Stimulate sharing of information and experiences
- Added objective understanding to a profile which has often been based on single analyses and anecdotal

What are we doing with the research results

- Attempting to use our knowledge to address some of *their* problems
 - e.g. Building Tools
 - Off-the-shelf policy statements
 - Business cases and strategies
 - Digestible guidance on technologies and their preservation implications
 - Construction of improved models (reference, cost...)
 - Guidance on repository design, maintenance, audit
- Conducting Further Research, for example
 - ERPANET's Portable Object Ingest Project (POIP)
 - Digital Curation Centre's research on ingest and extraction of semantic metadata

VISIT The ERPANET WEBSITE:

<http://www.erpanet.org>