Issues in Digital Preservation: Towards a New Research Agenda

Jean-Pierre Chanod[1], Milena Dobreva[2], Andreas Rauber[3], Seamus Ross[4], Vittore Casarosa[5]
[1]Xerox Research Center Europe - Grenoble, FR
[2]The University of Strathclyde - Glasgow, UK
[3]Technical University Wien, AT
[4]University of Toronto, CA
[5] ISTI-CNR, Pisa, IT

Abstract

Digital Preservation has evolved into a specialized, interdisciplinary research discipline of its own, seeing significant increases in terms of research capacity, results, but also challenges. However, with this specialization and subsequent formation of a dedicated subgroup of researchers active in this field, limitations of the challenges addressed can be observed. Digital preservation research may seem to react to problems arising, fixing problems that exist now, rather than proactively researching new solutions that may be applicable only after a few years of maturing.

Recognising the benefits of bringing together researchers and practitioners with various professional backgrounds related to digital preservation, a seminar was organized in Schloss Dagstuhl, at the Leibniz Center for Informatics (18-23 July 2010), with the aim of addressing the current digital preservation challenges, with a specific focus on the automation aspects in this field. The main goal of the seminar was to outline some research challenges in digital preservation, providing a number of "research questions" that could be immediately tackled, e.g. in Doctoral Thesis. The seminar intended also to highlight the need for the digital preservation community to reach out to IT research and other research communities outside the immediate digital preservation domain, in order to jointly develop solutions.

Introduction

Digital preservation affects every single citizen in the information society. It covers a whole spectrum of issues, from the long-term access to and use of personal digital objects, to the complex area of information objects' lifecycle management in big institutions from the engineering, governmental, research and cultural heritage sectors. E-government, e-science and e-culture more and more depend on proper storage and access to huge collections of digital resources, which should not be affected by the changes in the technological environment.

Given its relevance, digital preservation has become an active research field, combining expertise from a range of backgrounds including experts from cultural heritage and memory institutions, legal experts, scientists and engineers working intensively with scientific and

technical data in a broad range of disciplines and, of course, computer science and IT experts. This collaborative and interdisciplinary work has increased our understanding of the problem and has led to the study of different approaches, to the definition of international standards (e.g. OAIS and PREMIS), to the development of tools and systems that provide solutions to specific problems.

Despite its rapid progress, digital preservation has not yet reached a level of maturity similar to that of the constituent research domains, and a consistent theory of preservation is not yet in place. Due to the highly interdisciplinary nature of the challenges, as well as due to its rapid evolution into an independent research field of its own, digital preservation research runs the risk of becoming a "self referential" domain, missing essential expertise and input from more traditional sub-disciplines in each of the various disciplines involved.

Recognising that it would be particularly beneficial to bring researchers and practitioners with various professional backgrounds related to digital preservation together, and to allow them to exchange ideas and experiences, a seminar was organized in Schloss Dagstuhl, at the Leibniz Center for Informatics (18-23 July 2010), with the aim of addressing the current digital preservation challenges, with a specific focus on the automation aspects in this field. Ultimately, the ideal solution to the preservation challenge should be an (almost completely) automated environment for the creation of preservation metadata and the definition of a preservation workflow.

Given the broad range of topics falling under the umbrella of Digital Preservation, the specific focus of this seminar was on IT issues in this domain. This was also reflected in the members invited to participate, resulting in a strong technical background from which to explore potential solutions and relationships to other disciplines. The ultimate goal was to develop a research agenda for digital preservation that concentrates on the increasingly technical challenges and to foster new approaches to digital preservation, complementing the broader research agendas existing in this field.

Participants were asked to provide their views and possible answers to some of the following questions.

- What is the real place of automation in digital preservation?
- How to integrate better ongoing research in information retrieval, content management, modelling, simulation, natural language processing and other areas to the automation in digital preservation?

- How to improve the metadata lifecycle management in digital preservation?
- What are the general and specific application scenarios in digital preservation?
- Is a common digital preservation framework possible? How the various application domains change requirements to digital preservation?
- How to achieve fully deployed services for various users in the digital preservation area?
- What is the place of automation in a consistent theory of preservation?

Rather than running the traditional form of presentations, the seminar was organized around break-out sessions discussing research directions on selected areas. Each of these was introduced by a short keynote statement scoping the respective fields. Each topic was

subsequently discussed in two break-out sessions with interim plenary reporting and discussion. To make the results of the seminar more concrete and of immediate use, the groups of participants were given the following topics and were asked to define possible research challenges, identifying a set of immediate research questions, suitable for being the subject of a Doctoral Thesis.

- Preservation Ready Systems – Digital Preservation and Enterprise Architecture
- Beyond Metadata? Information Retrieval/Mining/Visualization/Context
- Storage Technologies and Protocols
- Policy and Rule Management
- Ethics, Privacy, Security and Trust
- Evaluation and Benchmarking in Digital Preservation
- Application domains

In the next section we will briefly summarize the main outcome from each group. A more detailed description of the outcome of each break out session, provided by the group rapporteur, can be found on the Seminar web site:
http://www.dagstuhl.de/no_cache/en/program/calendar/semhp/?semnr=10291.

Main research questions

Session 1: Preservation Ready Systems – Digital Preservation and Enterprise Architecture

If we consider, in the most general sense, an Information System, i.e. a system conceived and created based on the assumption that inputs, processes and outputs will deal with information, we can see that in general issues of digital preservation have been considered "after the fact", mainly as a way to preserve the output of the system. At the same time, digital preservation research has focused mainly on the development of Digital Preservation Systems (like OAIS), and on scenarios where Digital Preservation Systems are expected to interoperate with "external" Information Systems. These are the common scenarios in Cultural Heritage and in other areas such as Science, Technology and Medical publishing, where "Information Archiving" is a well defined concept.

If we consider now scenarios like Management Information Systems, Engineering Systems, Health Care Systems, and other domains where all the information of the system needs to be permanently available and is subject to frequent changes by the production processes, then it appears that the concept of "Information Archive" makes little or no-sense, even if digital preservation is clearly very much relevant. In these scenarios it would be desirable to identify digital preservation requirements from the beginning and to embed the implementation of those requirements into the system itself, making it a "Digital Preservation Ready" system.

In conclusion, there is a need for new knowledge to be produced (or motivated) by the digital preservation community about how to give the capability of being "DP Ready" to those Information Systems where digital preservation is difficult or impossible to delegate to external Digital Preservation Systems. The research challenges related to this topic should become part of the concerns of the Enterprise Architecture, contributing to the building of

systems that in addition to being reliable, usable, interoperable, etc., are also "preservable or, in other words, systems that are resilient against changes in technology.

Indicative research questions

- What are the principles to use to identify cases and evidences of potential DPR scenarios where ignoring DP can be a threat or a cause of loss in data, information or knowledge; or a loss of opportunity; or an economic loss; or a cause of degradation or even of loss of life? What are the principles to use to prevent those scenarios?
- What are the principles that can assure longevity of the information/knowledge base of information systems, apart from those systems' primary purposes?
- We already have a fairly large body of knowledge for the preservation of digital information. But what about the preservation of digital processes?
- How can we compare and align the actual DP references, such as OAIS, with other existing references already common in Information Systems environments, such as COBIT , ITIL (governance of IT infrastructures), MoReq (Records Management), etc.? Can we define, as a complement/alternative to OAIS, a kind of "MoRep-DP" (an ontology of generic requirements for DP to be taken in consideration in processes of Enterprise Architecture)?
- What strategic win-win scenarios and moves can the DP community envisage to approach and complement other relevant existing reference and standardization communities (such has for example the OASIS) with the DP concern?

Session 2: Beyond Metadata? Information Retrieval, Mining, Visualization, Context

Under this general area we may identify four major topics of interest: (i) what to preserve; (ii) the nature of Digital Objects; (iii) the conceptual modelling of Digital Objects and Digital Preservation; (iv) techniques applicable to Digital Preservation. It is also possible to define three recurring challenges across these topics: (i) the fluidity of preserved information; (ii) the preservation of context; (iii) scalability issues. For each challenge, the table below provides more detailed elements under each topic of interest.

Areas **Challenges**	**What to preserve**	**Nature of Digital Objects**	**Conceptual Modeling**	**Techniques**
Fluidity	Process/Dynamics	Volatile/endangered parts of Digital Objects	Evolution of ontologies Ontology of relations	Rate of change in metadata
Context	Intent Usage Production Other processes	Network of objects - annotations Preserve DO vs.DB - decomposition of DO into elements	Ontology relations - Similarity - Linking - Relations Diversity/Community	Statistical annotation assignment Diplomatics
Scalability	Recursion of preservation	Network of objects Preserve DO vs.DB	Diversity/Community Diplomatics	Extraction/ Collection of Knowledge Diplomatics

Indicative research questions

In all questions below it is assumed as a pre-requisite (a) to define an appropriate metadata set for preservation purposes or to demonstrate a mapping onto an existing metadata set and (b) to define the models and the metadata in such a way as to enable either minimization or streamlining of human input.

- Define a modelling framework for capturing the intended and the actual usage of digital objects, such that this may be documented in the course of creating or using the object respectively. A suitable classification scheme of usages is needed.
- Define modelling frameworks for specifying or describing processes of producing a digital object, as well as methods for automatically establishing or documenting these.
- Define modelling frameworks for specifying or describing processes of using a digital object, as well as methods for automatically establishing or documenting these.
- Develop a framework for representing situational collections of digital objects. These collections may be unstructured or they may have a structure induced either by untyped links or by specific relations, giving rise to networks of objects. Both the composition and structure of a collection may vary with time. An ontology of relations between digital objects is needed.
- Develop diplomatics-driven semi-supervised methods for metadata extraction from and annotation of digital objects.

Session 3: Storage Technologies and Protocols

Broadening in some way the scope of this general area, several main topics of interest have been identified: (i) self sufficiency; (ii) "imperfect" digital preservation; (iii) information handling; (iv) storage technologies; (v) distributed Systems and protocols; (vi) new frontiers. For each one of those we provide here a list of possible research challenges, with a very brief explanation.

Self-Sufficiency

Self-explaining, self-correcting, self-replicating code

When exchanging a digital item as data, it has been always necessary to "explain" such data in some way to receivers, both in terms of its data format and the underlying semantics. Adding properties such as "self-explanatory code", "self-correction" and "self-replication" would mark a change from passive to active data, although it remains totally unclear how, if at all, to achieve such properties in general terms. The limits of self-explanation should be explored, for example assuming the use of embedded textual or XML-based descriptions into binary data. A major difficulty regarding this research topic is the issue of its evaluation.

Self-contained, embedded devices for writing, querying and reading contained data

Rather than depending on a complex and changing ecosystem of hardware and software to provide for writing, querying and reading a certain type of data, it may be interesting to have a self-contained, embedded device which acts as an interface for humans, and which provides these services in a closed manner. A central aspect is the autonomous maintenance of data integrity within the device, ensuring that data is constantly checked and its detected health reported to the user.

Imperfect Digital Preservation

Perception-based graceful degradation in digital preservation
Depending on its content, a digital item provides for a perception-based graceful degradation (e.g. the loss of a few pixels in a digital photograph is quite acceptable). This requires models for measuring the perceived loss of information for video and audio content, which leads to the question of how to design file formats with perception-based error resilience / graceful degradation. At a later stage, it may be interesting to consider whether it is possible to transpose perception-based graceful degradation from the audio-visual content to symbolic data beyond text-based summarization.

Enabling longevity through redundancy
To ensure redundancy for very large amounts of data on the scale of petabytes, potentially stored on a large number of disks (10.000+ units) in a distributed scenario, redundancy has to be managed automatically in a transparent fashion by an autonomous system. Developing a mathematical foundation for such a system would allow the simulation of different parameterizations and measurements of key metrics, such as the Total Cost of Ownership, bandwidths for simultaneous archival and restoration during live operation, Quality of Service.

Summarization and Forgetting in digital preservation
Summarization and aggregation of preserved data over time maybe a desirable function of digital preservation, for example in cases where over time a huge number of commercial transactions lose their value as individual transactions, but retain their value in terms of overall trends and proportions. In other cases it might be desirable to prioritize data in order to be able to select information that might be forgotten "more easily" if a loss could potentially be steered. Also the possibility of "forgetting data for a while" might be desirable, for example when data first loses its commercial importance, only to gain historical value at a later point in time.

Information Handling
Secure, gradual release of confidential information over time
There are many cases in which confidential information may be released after a certain period of time (e.g. political decisions, personal information).In digital preservation it would be desirable to actually ensure that confidential information is represented in such a way that confidentiality is provably maintained over a defined period of time, and rescinded afterwards. This research topic has close ties with cryptography, potentially depends on the use of Trusted Computing infrastructure, and potentially has to handle the preservation of keys and cryptographic methods.

Provable deletion of information
Proving the deletion of a (preserved) digital object to a third party can become necessary, very often to satisfy legal requirements. Giving proof of a deletion itself is a non-trivial task that may depend on the support of the underlying technical infrastructure and also has to cope with issues such as non-erasable media. Here again, for guaranteeing the deletion of the right digital item, the use of cryptographic methods like Digital Rights Management or digital signatures may be necessary.

Storage Technologies
Visualizing magnetization of magnetic media

In efforts related to Digital Archaeology, it would be desirable to have methods for visualizing the magnetization of magnetic media, thereby switching media access from the magnetic to the visual domain.

Enabling actual long-term storage technology
Current storage technology is optimized for aspects such as access speed and throughput, but not for long-term storage of data. In digital preservation, it would be desirable to have the possibility of dynamically choosing whether to prioritize access speed over durability of storage or vice versa, assuming that a suitable storage technology is available.

Total Cost of Ownership models for comparing storage technologies in digital preservation
Different types of storage technology have varying Quality of Service parameters and Total Cost of Ownership. It would be desirable to have actual models for simulating a specific digital preservation solution, for computing both the resulting TCO and QoS parameters that could be expected, and for comparing different solutions. Adequate models could also help testing alternative solutions with respect to other parameters (e.g. redundancy, throughput, storage capacity, risk of data loss).

Distributed Systems and Protocols

Distributed peer-to-peer archival system
In complex and dynamic environments where data to be preserved is stored in a redundant manner, it would be desirable to have a distributed system which can handle the addition and removal of nodes with storage capacity, and which manages the redundant storage of data according to given policies. The problem of distributing contents over large networks has been successfully addressed in the domain of Peer-to-Peer, and could be extended to manage the automatic distribution of data and the management of redundancy in digital preservation.

Synchronisation of distributed storage data in case of bit errors
In distributed storage systems that may easily carry data in the range of Petabytes, nodes should be able to mutually correct errors occurring in a node (e.g. bit flips due to cosmic rays) without actually needing to exchange large portions of data. In part, this is a problem of synchronizing data in different storage locations, which has been addressed by protocols such as rsync and others, but not fully addressed in the context of P2P systems or for large volumes of data.

New Frontiers

DNA as data carrier
Biological life itself can be considered as a role-model for the preservation of data, since data in DNA has been "preserved" successfully for millions of years. There are potential opportunities to learn how life itself has solved the preservation of DNA-based representation of information, where interesting properties such as self-repair are built-in. This is still quite a speculative area, and there are opportunities for cross-domain cooperation with Biology and Bio-Informatics.

Digital preservation in the age of Quantum Computing
Due to the advent of Quantum Computing, the question arises whether it is going to have effects on digital preservation itself, e.g. regarding the preservation of quantum computing

programs, or preserving the result of quantum computing computations. These also are highly speculative questions, as it is unclear whether such effects are actually present.

Autonomous Agents as data carriers
Computer viruses can be considered as prototypes for artificial life, which focus on their diffusion and replication, thus ensuring the redundancy of their code as data. Applying this concept to digital preservation, the idea is to utilize autonomous, distributed agents to preserve portions of data, similar to biological cells. In addition to being highly speculative, this idea raises a number of legal and ethical problems.

Session 4: Policy and Rule Management

It is the current state of the art to explicitly formulate policies and rules for digital preservation and express them separately from the underlying system in order to make them more manageable, adaptive and auditable. They describe how digital preservation is ensured by an institution and digital preservation systems must enforce them. Digital preservation policies and rules also provide important contextual information for evaluating the authenticity of preserved objects and for communicating service expectations and requirements between stakeholders.

Indicative research questions

Some of the research questions and challenges about policy and rule management are the following.

- How different methods for deriving policy compare to each other (e.g. policies based on digital preservation or domain specific standards, based on theoretical works, based on natural language statements, derived from observations and measurements of system and human behaviour)? Are the policies derived with different methods comparable to each other? How can these different methods for deriving policies be refined or supported?
- Can policies play an effective role as boundary objects in decision making and negotiations between stakeholders? Can policies serve as sufficient documentation? Can policies effectively model the results of negotiations? Can policies support negotiations between stakeholders? Is there a minimum set of decisions needed in order to specify the policies for a digital preservation environment?
- What are the differences between preservation policies for dedicated preservation environments and those for production systems with preservation capabilities? Are domain-specific languages required or are generic policy languages sufficient for digital preservation? How content-specific does a policy need to be? What are the categories of policies that are required at different levels of granularity? How can different compliance levels be expressed?
- How should policies be expressed in order to facilitate the definition and automation of preservation processes? Can we develop controlled vocabularies or even ontologies for expressing policies? What should be the level of granularity and the atomic operations of those ontologies?
- How should policies that change over time be managed and versioned so that they can be appropriately associated with the digital objects of the time? How can this

process be supported? Can policies and rule sets be found which are resilient to change?

- How can we prove that a policy is consistent and complete? How can we compare and harmonize different policies of two cooperating organizations? How can policy conflicts that prevent successful preservation be detected and resolved? How should policies and rules be formulated so that we can reason about them?

Session 5: Ethics, Privacy, Security and Trust

The topics of ethics and privacy are closely connected with security and trust, and their interplay provides a number of interesting research challenges and questions. While these will need solutions to be connected to ethical and legal frameworks, the need was felt to understand the technical implications and possibilities in order to provide guidance on the risks as well as potential mitigation strategies from a technical perspective.

Indicative research questions

A (by far non exhaustive) list of research questions includes.

- How can we model and detect different categories and level of permissible collection and usage?
- How can we develop various representations of digital information with different levels of machine-interpretability?
- How can we develop methods for ensuring integrity and controlling access
- How can we integrate security measures into digital preservation?
- How can we use standard digital signature models and key management methods over long periods of time?
- How can we support security strategies (e.g. user roles, access rights) over very long time periods?
- How can we reconcile the rights of individuals across accounts and systems?
- How can we express specific levels or types of trustworthiness in a trusted system?
- How can we identify the “original” among multiple instances of digital objects?
- How can we define and mitigate specific data leakage risks?
- How can we manage sensitive data and define and mitigate confidentiality risks?
- How can we define and mitigate forgery risks?
- How can we detect evidence of forgeries?
- How can we develop functional requirements for Digital Rights Management aware of digital preservation?
- How can we detect inappropriate reuse of intellectual property?
- What are ethical considerations related to information emerging from aggregates but not from individual items (e.g. data mining, data correlation)?
- How can we develop controlled release regimes based for example on time periods, trigger events, use context, documentary context?

Session 6: Evaluation and Benchmarking in Digital Preservation

Evaluation and benchmarking is a crucial aspect in digital preservation, considering that in practical terms there can be no quality assurance without evaluation, no marketplace without comparability and no competition without a metric. A pragmatically viable approach is to define the main aspects that need to be measured and address them on a prioritisation basis.

Indicative research questions

Specification of criteria

- How can we decide on the completeness of a set of criteria at a certain level?
- How can we address the fact that significance of certain criteria may **change over time** and that new criteria will be added to the set in the future, e.g. when the context of usage changes, the user community shifts, etc.?

Properties modelling

- Can we define classes of objects concerning their properties relevant for digital preservation?
- How can we describe this in a way that is preservable in itself?
- How can we describe an object according to these properties?
- What is the minimum set of properties needed to define a similarity metric between two objects of a certain class?
- Is it possible to automatically discover new properties, given an object and its set of properties?

Measurements and calibration

- How can we know that the measurements are correct, having available a measurement device for comparing two objects?
    - formally proof the correctness of the method used in the device
    - multiple devices, pick a second (third) method - statistical approximation
    - self-correction and adjustment over time
- How can we achieve (and reflect) confidence in a measurement?
- How can we model uncertainty in measurements and handle it?
- Is it feasible and sufficient to employ crowdsourcing for approximating and/or verifying measurements for certain criteria?

Benchmarking and ground truth

- How can we define an annotated benchmark set and metrics to compare different measurement devices in scenario X?
- How can we define a benchmark to compare preservation (-ready) environments?
- Is it feasible to create sets of reference objects for certain classes that adequately represent their class?
- Based on a set of extracted properties, can we construct a minimum representation (a minimally simple test object that contains all properties)?
- Can we create the equivalent of a *colour card* as it is being used in photography, to embed in objects and/or processes for quality assurance?

    - What are the desired properties of such an artefact?
    - How can we generate these and embed them?
    - Can we define preservation processes that can take them into account?
- Can we create links between extracting the semantics and extracting the content/formatting (more or less like a checksum used in communication protocols)?

Preservability and risk assessment

- How can we define a (exhaustive) list of factors that influence the preservability of a certain object?
- Can we define a unified scale of risk assessment and/or preservability of a certain object?

Supporting quality assurance

- How can we enrich processes (production, ingest) to produce information that supports present and future Quality Assurance?
    - What is the minimum set of properties that have to be defined?
- How can we address aspects such as dependency management and change propagation over time while securing preservation requirements in a system with known purpose, function, boundary conditions, test data, expected outcomes, etc.?

Human experience and reference points

- Can we model the behaviour and interactions between systems in such a way that allows us to assess faithfulness of preservation of a piece of/or a system?

Feedback and Learning

- Can we create a preservation (-ready) system that incorporates (user) feedback about preservation quality and learns from mistakes to improve quality?
- Can we feasibly incorporate user feedback into access modules and feed that information into an improvement cycle?
- Can we address the relation between objective measurements and subjective judgement using such crowdsourcing and approximation?

Session 7: Application Domains

Computer games can represent the “exemplar application” overarching all the other application domains. We might say (with some optimism) that if we can preserve computer games, then we can preserve everything digital. They present a number of technical and legal challenges:

- Games are complex digital artefacts
- Games are interactive, and interaction must run at original speed
- Games are usually copyright protected with different measures
- Games are of hybrid legal nature
- Games are complete applications, bound to an original system (hardware, operating
- system, libraries), for which they have been programmed
- Games are at the same time audiovisual works and software

Indicative research questions

Given their complexity, computer games can be seen as indicators for future developments in digital preservation in a number of application domains. Some of the research challenges given by computer games are the following.

- Games are applications, whose cultural value consists to a good portion in the way the user interacts with it.
    - How can historical interactions/ user behavior be preserved?
    - How can the applications be preserved, so that future users can be enabled to interact with the application environment in a way the historic users did?
- There is a clear tendency, that games were distributed more and more in parts rather than in one application. Also the service based model, in which the application is running on a server and is sending only a video stream over the internet back to the user, went into business recently.
    - Are we facing a turn from a download model towards a service based model of distribution?
    - What strategies must be developed to preserve artefacts, which are distributed in such a way?
- Digital Rights Management systems become more and more worrying for archivists. For example one of the latest trends is that the game will only run when connected on-line with a special dedicated server from the Intellectual Property holder.
    - How can DRM systems be developed, which satisfy the needs of the IP holders to protect their IP and offers special features to allow preservation, too?
- Considering the capabilities of mobile devices, mixed/augmented reality games applications will become more important.
    - How can augmented/mixed reality environments be preserved?
- Since the early days of the Web, the gamer community set up great resources of metadata and game archives on a user generated basis. This could become an example, which might be applied to other domains as well.
    - How can user generated content be used in digital preservation in a systematic way and at institutional level?

Conclusions

This preliminary exercise has produced a number of proposal and ideas ranging from short to medium term research questions to long term and highly speculative research challenges. These ideas need to be organized in a coherent road map, with priorities and a judgement of relevance. In many cases these ideas have highlighted the need for the digital preservation community to reach to other research communities for cross-fertilization and joint development of solutions.

Despite the wide range of topics addressed in Section 2, there is still a number of topics that are essential in the context of digital preservation and need the attention of the relevant research communities. These include:

- Business models
    - Object level, value of information
    - Systems level: cost models, market analysis, investment cycles
    - Organization: risk analysis, investment trade-off, investment cycles

- Theory of Information and effects on digital preservation
    - Shannon-equivalent for semantics
    - Object models, de-composing information, mash-up objects, object networks
    - Processes
    - Facets of Information
- Standards
    - Standardization bodies
    - Prescriptive standards vs. evolving standards
    - Industry standards
- Digital-preservation-stable hardware
    - hardware that runs stably, optimized for operational longevity rather than speed, limiting number of required system migrations

In conclusion, we can say that while the field of digital preservation is still young, it has matured to a level of considerable complexity and specialization. In order to solve the challenges ahead of us, however, the preservation community needs to ensure it remains open and manages to attract professionals from different backgrounds, including but definitely not limited to, computer science experts, to jointly address the challenges that our information society is facing.

At the same time Computer Science has to acknowledge that in order to achieve a sustainable computing environment there is the need for digital preservation to be included as an integral part of the system design and development process, in addition to established concepts like system operation, maintenance, reliability, etc. Once this level of IT system maturity is reached, digital preservation will come "for free" as part of systems operation, rather than as a separate add-on. Until then, we will need to continue investing considerable efforts to mitigate the risks threatening the long-term availability of digital information.

Appendix 1 – List of Participants

Jose Barateiro , Polytechnic Institute of Lisbon/INESC-ID - Lisboa
Christoph Becker , TU Wien
Jose Borbinha , Polytechnic Institute of Lisbon/INESC-ID - Lisboa
Vittore Casarosa , ISTI-CNR - Pisa
Jean-Pierre Chanod , Xerox Research Center Europe - Grenoble
Panos Constantopoulos , Athens University of Economics and Business
Milena Dobreva , The University of Strathclyde - Glasgow
Muriel Foulonneau , Henri Tudor Research Center
Rudolf Gschwind , Universität Basel
Michael Hartle , TU Darmstadt
Yunhyong Kim , The Robert Gordon University - Aberdeen
Andreas Lange , Computerspiele Museum (im fjs e.V.) - Berlin
Cal Lee , University of North Carolina- Chapel Hill
Jens Ludwig , Universitätsbibliothek Göttingen
Alexander Mehler , Universität Frankfurt am Main
Natasa Milic-Frayling , Microsoft Research UK - Cambridge
Gillian Oliver , Victoria University of Wellington

Vassilis Plachouras , Athens University of Economics and Business
Andreas Rauber , TU Wien
Stefan M. Rueger , The Open University - Milton Keynes
Peter Stanchev , Kettering University - Flint
Stephan Strodl , TU Wien
Dirk von Suchodoletz , Universität Freiburg