Project no. 034762

DigitalPreservationEurope (DPE)

Instrument: Coordination Action

Thematic Priority: IST-2005-2.5.10

Access to and preservation of cultural and scientific resources

Research Roadmap

Due date of deliverable: 31.10.2007

Actual submission date: 31.10.2007

Start Date of Project: 01 April 2006

Duration: 36 Months

Name of Organisation of Lead Contractor for this Deliverable:

FernUniversität in Hagen

Names of Partners Engaged in this Deliverable: GU, TUW, SB, NANETH, VUFC, HUF

Final

Project co-funded by the European Commission within the Sixth Framework Programme

(2002-2006)

_____

Dissemination Level: **PU (Public)**

Legal Notices

Catalogue Entry

Title	Research Roadmap
Creator	FernUniversität in Hagen
Subject	Information Technology; Science; Technology--Philosophy; Computer Science; Digital Preservation; Digital Records; Science and the Humanities.
Description	The DPE Research Roadmap provides a concise reflection of the core research issues for digital preservation. Based on a crosswalk of ongoing activities and existing research agendas, DPE recommends digital preservation research into 10 focus areas.
Publisher	HATII at the University of Glasgow, includes FernUniversität in Hagen, Fondazione Rinascimento Digitale, Ministero Per I Beni E Le Attività Culturali, Národní knihovna Ceske republiky, Nationaal Archief van Nederland, Statsbiblioteket, Technische Universität Wien, Vilnius University Faculty of Communication.
Contributor	FernUniversität in Hagen
Contributor	HATII at the University of Glasgow
Contributor	Nationaal Archief van Nederland
Contributor	Technische Universität Wien
Contributor	Ministero Per I Beni E Le Attività Culturali
Contributor	Statsbiblioteket

Title	Research Roadmap
Contributor	Niedersächsische Staats- und Universitätsbibliothek Göttingen
Date	25/09/2006 (creation)
Type	Text
Format	Adobe Portable Document Format v.1.3
Resource Identifier	ISSN [to be provided by GU]
Resource Identifier	ISBN 978-1-906242-09-1
Resource Identifier	DOI [to be provided by GU]
Language	English
Rights	© DigitalPreservationEurope Partners

Citation Guidelines

Document Version Control

Version	Date	Change Made (and if appropriate reason for change)	Initials of Commentator(s) or Author(s)
0.1	25/09/2006	Initial Version of Research Roadmap	hb
0.2	10/10/2006	Input from Wiki Draft	hb, sr, ml, hh, cr, rn, ss
0.3	15/10/2006	Added small Cognition section	hb
0.4	06/11/2006	Added Current Research Directions, some structural changes, additions and modifications	hb, mh, jk, rl, sr, rn, ss
	06/11/2006	Phone discussion	sr, hb
	07/12/2006	Voicemeeting	sr, cr, rn, hb
	11/12/2006	Voicemeeting	sr, hb, rr, rn, cr, hb, ml, mm, cc, ss
	19/12/2006	Interview Package finalized	sr, rr, jk, rn, cc, ml, cr
	22/12/2006	Research Analysis Matrix	rr, cr, rn, jk, hb
0.5	29/12/2006	Revised structure, updated contents	hb
	03/01/2007	Skype Conference Call	sr, rr, hb
0.6	05/01/2007	Added project descriptions, started with executive summary	sr, hb
0.7	07/01/2007	Added DCMI and Semantic Web	hb
	09/01/2007	Schiphol Meeting	sr, hh, rr, hb
0.8	16/01/2007	Final revision of the structure, added references, executive summary, minor additions	hb, sr
0.9	18/01/2007	Elaborated on introduction, discussion about current status and work to be done	hb, mh
1.0	19/01/2007	Incorporated most of Hans' suggestions for the research issues, added crosswalk section, and agenda summaries	hh, hb, mh

1.1	20/01/2007	Added further agenda summaries, moved Links to other Disciplines into Appendix A, deleted Cultural Heritage section	hb
1.2	22/01/2007	Added further agenda summaries, corrected references, completed sota section, added draft executive summary, some revisions	hb, mh, rr
1.3	23/01/2007	Added remark about only looking about research agendas, further contributions, added missing sota sections	rr, mh, hb
	12/ß3/2007	Skype Conference Call	Sr, hh, hb
1.4	13/03/2007	Updated Netherlands, added Baltic States.	hh, zb
1.5	16/03/2007	Added Italy, Austria, some corrections (AE => BE).	mm, rn, hb
1.6	20/03/2007	Added Czech Republic, updated Estonia	jh, rr
1.7	23/08/2007	Language changes	ew
1.8	17/10/2007	Inserted reference to UVC/Multivalent in 4.2, added iRODS, NARA, and analysis to UK science report. brief intro to online survey	hb
1.9	23/10/2007	Extended section on persistent identifiers	dh
2.0	25/10/2007	Completed online survey section	hb

Document Quality Control

Version QA'd	Date	Recommendations (and if appropriate reason for change)	Initials of Quality Assurance Person
2.0	29.10.2007		EW

Document Change Commentator or Author and Reviewer List		
Author Initials	**Name of Author/Reviewer/Responsible QA Person**	**Institution**
hb	Holger Brocks	FernUniversität in Hagen

sr	Seamus Ross	HATII at the University of Glasgow
ml	Maurizio Lunghi	Fondazione Rinascimento Digitale
hh	Hans Hofman	Nationaal Archief van Nederland
cr	Colin Rosenthal	Statsbiblioteket
rn	Robert Neumayer	Technische Universität Wien
ss	Stefan Strathmann	Niedersächsische Staats- und Universitätsbibliothek Göttingen
mh	Matthias Hemmje	FernUniversität in Hagen
jk	Jurate Kupriene	Vilnius University Faculty of Communication
rl	Rimvydas Lauzikas	Vilnius University Faculty of Communication
rr	Raivo Ruusalepp	Nationaal Archief van Nederland
cc	Chiara Cirinnà	Fondazione Rinascimento Digitale
mm	Maurizio Messina	Biblioteca Nazionale Marciana
zp	Zibute Petrauskiene	Vilnius University Library
jh	Jan Hutar	National Library of the Czech Republic
ew	Emily Witham	HATII at the University of Glasgow
dh	Dominic Heutelbeck	FernUniversität in Hagen

Approved By with DPE (signature) (SRoss)	Date

Accepted by at European Commission (signature)	Date

1. EXECUTIVE SUMMARY

After almost two decades of setting digital preservation research agendas there is little evidence of actual progress in the development of solutions. Our heritage may now be at greater risk because many in our community believe that we are making progress towards solving the preservation challenges. A major cause of this failure has been that the community has consistently widened their research focus. As a community we failed and we need to understand why.

To identify the remaining fields and challenges for our future research and development work in Digital Preservation we must first ask what we have already achieved. If we take OAIS [41] as an example we can surely claim that we have already achieved a very high level conceptual model of the Digital Preservation application and research domain. Work has also begun on a few early Digital Preservation testbeds. Nevertheless these achievements only aim at supporting the explicit management of more or less static digital preservation processes and static and homogenous Information Object Media and Information Package Types.

In addition to this at least half-a-dozen metadata models from the background of Digital Libraries, the Internet, the Web, and Knowledge Technologies are currently under development. These metadata models are intended to support Digital Preservation applications but are not yet interoperable and integrated within a technical framework, are not yet standardised themselves or operating on the basis of appropriate basic standards. Nor do they fully take into account the current state of the art in computer science such as the cutting edge knowledge in computing architectures and resource management technologies, information extraction such as linguistic and mm analysis technologies, semantic representation, annotation and processing technologies, effective context and workflow management technologies or finally efficient human computer interaction and human-human collaboration support technologies..

The Research Roadmap aims to contribute to the planning of our future research and development in Digital Preservation by means of analysing the state of the art in Digital Preservation research and existing research agendas on a global level. In addition to this the Roadmap will analyse the state of the art and future markets of computer science and information and communication technology, the needs and demands of the Digital

Preservation user communities and their leading experts. It is hoped that this analysis will pave the way for the development of conceptual, technological and application oriented building blocks for advanced solutions to support Digital Preservation applications in a variety of possible scenarios.

DPE Recommended Research

- Restoration

- Conservation

- Management

- Risk

- Significant Properties of Digital Objects

- Interoperability

- Automation

- Context

- Storage

- Experimentation

CONTENTS

Introduction

After almost two decades of setting digital preservation research agendas there is little evidence of actual progress in the development of solutions. Our heritage may now be at greater risk because many in our community believe that we are making progress towards solving the preservation challenges. A major cause of this failure has been that the community has consistently widened their research focus. As a community we failed and we need to understand why.

1.1. PURPOSE

Despite the frequent discussions concerning the direction of research in digital preservation no concise and well-developed strategy representing a clear consensus has emerged. To address this need the DPE Research Roadmap's objective is to provide a concise overview of the core issues which have to be addressed in future digital preservation research based on an extensive crosswalk of existing preservation research agendas.

1.2. AUDIENCE

This report is targeted at funding bodies and policy makers, but also at researchers working in the area of digital preservation.

1.3. SCOPE

Building on a consideration of research agendas developed during the last 16 years, the DPE Research Roadmap provides

- an assessment of the state-of-the-art in digital preservation
- a review of the selected reports on digital preservation research
- an overview of core research issues identified by DPE.

Furthermore, it will be validated and supported by

- interviews with experts in the field of digital preservation
- an accompanying online survey, giving the opportunity for public comment.

1.4. WORKING PRACTICES

The DPE Research Roadmap was created by

- carrying out desk-based research, studying literature and examining ongoing and previous activities

- focused discussions in meetings and conference calls, but also via email and instant messaging

- conducting a series of interviews (ongoing)

- posting an online survey on the DPE website[1].

1.5. STRUCTURE OF THE REPORT

Following this introduction, Chapter 3 provides a thorough analysis of existing research agendas on digital preservation, summarizing their proposed research directions and discussing their impact. Chapter 4 analyses the crosswalk of research agendas and focuses on the core issues of digital preservation research as identified by DPE. The main report concludes with a summary of the findings and recommendations for further actions.

The DPE Research Roadmap is accompanied by two appendices which together provide extensive coverage of the state-of-the-art, ongoing activities and links to research conducted in related domains (Appendix A), detailed description of ongoing initiatives on Digital preservation, (Appendix B) A Summary of the research agendas reviewed.

A final edition of this report will also contain a thorough analysis of the feedback gathered through the expert interviews and the online survey, reflecting their findings and implications in a systematic way.

---

[1] http://www.digitalpreservationeurope.eu/delphi-survey/

2. ANALYSIS OF PREVIOUS RESEARCH AGENDAS

For the crosswalk analysis of previous research agendas we have selected 12 documents which have been published over the last 16 years and elicited their core recommendations on digital preservation research.

Below is a list of the documents reviewed and the abbreviations assigned to them, a more detailed summary of these reports and their contexts can be found in Appendix B.

- **UEI** – *Understanding Electronic Incunabula: A Framework for Research on Electronic Records* [9] by Margret Hedstrom, 1991.

- **PDI** – *Preserving Digital Information* [17], edited by John Garrett and Donald Waters, 1996.

- **DPNU** – *An Investigation into the Digital Preservation Needs of Universities and Research Funders* [15] by Denise Lievesley and Simon Jones, 1998.

- **SoDP** – *The State of Digital Preservation - An International Perspective* [5] contains articles by various authors, 2002.

- **IAT** – *It's About Time: Research Challenges in Digital Archiving and Long-term Preservation* [3] was published by the NSF in 2003.

- **I2S** – *Invest to Save* [4] was prepared for the NSF-DELOS working group on digital archiving and preservation in 2003.

- **eScience** – *e-Science Curation Report* [16] by Philip Lord and Alison Mcdonald was published in 2003.

- **Cyber** – *Revolutionizing Science and Engineering Through Cyberinfrastructure* [18] was created by the Blue-Ribbon Advisory Panel on Cyberinfrastructure of the NSF in 2003.

- **DigiCult** – *The Future Digital Heritage Space: An Expedition Report* [6] was published as a DigiCULT thematic issue in 2004.

- **Erpanet** – *Electronic Resource Preservation and Access Network*[2] was a European Commission funded project which ran from 2001 until 2004.

- **Warwick** – *Digital Curation and Preservation: Defining the research agenda for the next decade* [1] reports on the Warwick workshop held in 2005.

---

[2] http://www.erpanet.org

- **DRR** – *Digital Repositories Roadmap - Looking Forward* [29] by Rachel Heery and Andy Powell, 2006.

- **OSI**– *E-Infrastructure Strategy for Research: Final Report from the OSI Preservation and Curation Working Group* [44] by Neil Beagrie, 2007. Note: Part of this report is based on the findings of [1], which are not repeated here.

2.1. ANALYSIS MATRIX

Each source was examined by at least two members of the DPE consortium in order to extract its core arguments. Since the focus was on digital preservation research, other issues have been ignored. Based on the interim findings, the identified issues have been clustered into five categories, which are characterised as follows:

- **Digital Object Level** – Fundamental issues in digital preservation are raised at digital object level. There has been significant research done on file formats which has produced working solutions for handling the preservation of static digital resources as bit streams, but new types of digital resources emerge regularly and there has been little research on complex and dynamic objects to date.

- **Collection Level** – The collection level is concerned with the longevity of digital collections. While most of the digital preservation research has, thus far, focused on individual types of digital objects, in practice digital repositories are dealing with collections of digital objects and their metadata. Interoperability between different archiving institutions is most often required at the collection level.

- **Repository Level** - The management of digital repositories is based on the understanding of the organisational requirements for long-term preservation of digital material. While digital repositories are being set up, based on OAIS and other standards, research and best practice in managing a digital repository remains a growing area.

- **Process Level** - Preservation processes are established in various contexts. Current preservation practices at most digital repositories still map traditional preservation techniques for physical materials onto digital content. The deluge of digital content that can be expected in the near future necessitates automation of ingest, preservation and delivery processes.

- **Organisational Environment** – The organisational environment is characterised by the creation and use of information, including handling legal aspects such as policies and intellectual property rights. Collaboration between archiving institutions, but also

with content creators and distributors are necessary to improve the efficiency and effectiveness of digital preservation. But collaboration might also be interdisciplinary, drawing upon research in related domains.

The results from the crosswalk analysis have been entered into a matrix, with similar issues being grouped under the most appropriate category. Some groupings and categorisations were debated, but agreement was reached in each case. To simplify the visual representation of the crosswalk analysis matrix (see Table 1), each research issue raised is represented by a "+" in their respective group. Hence "++" in column "*IAT* "and row "*Migration*" depicts the occurrence of two individual recommendations for research into migration in the report *It's About Time* [3]. The absence of a "+"indicates that that issue was not covered by the report specified in the column.

Table 1: Simplified Crosswalk Analysis Matrix

	UEI	PDI	DPNU	SoDP	IAT	I2S	eScience	Cyber	DigiCult	Erpanet	Warwick	DRR	OSI[3]
	1991	1996	1998	2002	2003	2003	2003	2003	2004	2001-2004	2005	2006	2007
Digital Object Level													
Migration		+		+	++	+							
Emulation				++	+						+		
Experimentation		+				+							
Registries and repositories					+++	++++				+	++		+
Complex Objects	+			+	++	+		+	++	++	++		
Significant properties			+	+	++	+	+						
Authenticity				++	+++++					+	+		
Acceptable loss					+			+					
Collection Level													
Interoperability				+	+++	+					+		
Metadata		+			++++								
Management					+++	+							
Standardisation			+		++		+	+		+	++	+	
Media Types					+		+		+				
Repository													

---

[3] Findings of the Warwick report [1] are not repeated here.

Level													
Tools and architectures		+			+		++						+
Benchmarks		+			++++		+						
Hardware Issues								++++			++		
Storage		+		+	+	+++		+		+	++		
Trust		++	++	+++	+	+++			++	++++			
Scalability					+++			++					
Sustainability				+	+++					+	++++	+	
Planning		+		+			++			+	+	+	
Repository Management							+			++	+		
Cost					+		+			+	+		
Process Level													
Access				+		+		+	++		++		
Automation				++	+++++	++++	+		+++++		+++++		
Monitoring				+									
Organisational Environment													
Creation and use	++	+					+						+
Legal Issues		+	+	+++			+	+		+			+++
Collaboration		+++		+++	+			++					++

2.2. RESEARCH TOPICS AND ISSUES

This section illustrates some of the topics and issues raised in the research agendas under review to provide an overview on the characteristic research topics deemed relevant in the past 16 years, not to judge or compare individual agendas. Since complete coverage is beyond the scope of this report, we restrict ourselves to brief descriptions of some exemplary issues for each group within a category.

Digital Object Level

At digital object level, *migration* and *emulation* (*virtualisation*) of software and documents are recommended to overcome technological obsolescence. Research into practical e*xperimentation,* for instance, deals with accelerated aging of systems and software. Whereas *registries and repositories* should be established for software, file formats and representation information like schemas, ontologies or concept spaces. Research into *complex objects* comprises best practices for multimedia and dynamic formats, but also the specification of models and methods to represent temporal and procedural relationships. Research into

Significant properties requires formal models for the selection and appraisal of digital objects including the identification of their "preservable" aspects. Establishing the *Authenticity* of digital objects covers the certification and authentication of preserved digital objects for example by digital signatures, and the establishment of persistent identifiers. *Acceptable loss* corresponds to the need for defining acceptable levels of information loss, i.e. not everything can be preserved.

Collection Level

Research into *interoperability* is required to promote the exchange of content and services between repositories. Standards for descriptive and management information are suggested under the *metadata* heading, fostering the mapping between concept spaces at collection level. Effective means for the validation of content, but also for refreshing aging media are required from the *management* perspective. This also includes methods for verifying collection completeness and anomaly detection. *Standardisation* relates stable ways to represent sound, image, video and other object components, but also to long-term metadata viability. It also comprises standardised ways for operating with and describing research data. Research into *media types* addresses the evaluation of the effectiveness of different formats, data models and metadata schemas.

Repository Level

Recommendations for transaction systems managing rights and authentication, methods and protocols for interoperability between architectural layers with a repository system, for example OAIS compliant archives are given under the heading *Tools and architectures*. Metrics for measuring the quality and fidelity of the preserved digital objects and metrics for assessing the costs and effectiveness of preservation functionality are the prerequisite for *benchmarks*, allowing the adoption quality markings for data and data repositories. *Hardware issues* refers to standardised interfaces for pluggable storage hardware systems, processing power and error correction codes for commercially available hardware. Capacity and scalability with respect to massive quantities of data are the main issues related to *storage*, but salvage and rescue of archival media are also deemed relevant. *Trust* is one of the most prominent issues, requiring standards for the audit and certification of digital archives. *Scalability of solutions* in the main refers to distributed and GRID storage, but also scaling down to meet the needs of smaller institutions. *Sustainability* refers to sustainable funding and business models for digital repositories, including market analyses. Research into fail-safe mechanisms, the specification of preservation strategies, and the provision of appropriate tool support are covered by *planning*. *Repository management* is concerned with the selection

of technologies, quality control and the formalisation of specifications for management and policies. *Cost* models are proposed to drive the selection process, but also to provide a comprehensive financial overview on the life-cycle of digital objects.

Process Level

New *access* models should be developed in partnership with designated user communities. This includes multilingual support and visualisation of complex data sets in addition to the development of semantic descriptions and ontologies. *Automation* requires formal models for the representation of preservation processes. Automated ingest, metadata extraction and automatic migration are amongst the most prominent recommendations. The *monitoring* and management of ongoing processes, be they automated or manual, is another research issue at process level.

Organisational Environment

The *creation and use of information* covers changes in archival practice, indicating that preservation starts with owners and providers. Additionally, it comprises educational materials and guidelines for researchers. *Legal issues* correspond to copyright and intellectual property, but also to policy issues such as regulations for the legal deposit of electronic publications. Interdisciplinary *collaboration* is aimed at promoting work across borders, e.g., to draw upon research in related disciplines or to improve consistency across sectors. Partnerships with content creators and distributors are also endorsed.

2.3. DISCUSSION

It is clear then from this cross-walk, that many research issues have been raised in the past 16 years. However actual progress in finding digital preservation solutions has been slight. Many plans have been developed, but widely applicable solutions to digital preservation problems remain the exception rather than the rule. There is no single explanation for this failure but a combination of reasons which led to the current unacceptable situation. Below, some of the issues we observed which may have contributed to the lack of progress, despite the considerable effort and funding.

- **lack of common understanding** – There is no common understanding of the precise definition of digital preservation and its allocated responsibilities within the larger picture of information and knowledge management.

- **loss of focus** – The research focus in digital preservation has consistently widened, incorporating issues already covered by other domains. One example of this is

research into access, which is a well established research topic in the Digital Libraries community.

- **lack of practical experience** – Despite several projects stressing the real-life applicability of their approach, only few have found their way into productive environments. This might be due to the fact that many problems only occur in practice, and are not experienced in a project environment.

- **fragmentation** – Many projects, mainly due to their limited funding, are only able to address a subset of research issues. In contrast the research agendas analysed here prescribe a holistic approach to digital preservation research. Without any coordination between the projects, potential synergies are lost.

- **frictional losses** – Competition between institutions and projects at national and international level impedes advancements, because existing solutions are ignored in favour of additional proprietary solutions. This problem mainly stems from the various funding schemes.

- **IPR impediments** – Those rights which restrict the retentions of objects by libraries and archives, in part resulting from mismatched balancing of public and private interests.

- **lack of training** - The lack of adequate knowledge transfer by training and professional skills development.

We are aware that in examining the research agendas alone and not relating them explicitly with the practical work that has been carried out, we are presenting a partial assessment. The fact that a topic is re-current in these agendas does not necessarily mean that work on the topic has not been done.

Furthermore despite the evaluation of the current state of digital preservation research appearing somewhat frustrating, hope remains strong for its success. The theoretical basis of digital preservation research is sound. Now the shift in focus needs to return to resolving the fundamental issues and apply the results in practice.

3. ONLINE DELPHI SURVEY

The DPE online delphi survey was designed to draw on professional community expertise to assess the current state of the art in digital preservation research and identify the challenges of that need to be met in the near and medium term. Participants of the online survey were given the opportunity to answer five questions about each of five themes, but also an opportunity to provide general comments.

The five research themes were

- **Theme 1: Digital Objects and Collections** – concerned with the longevity of digital objects and collections.

- **Theme 2: Management of Digital Repositories** – concerned with the development of our understanding of the organisational requirements for long-term preservation of digital material.

- **Theme 3: Key Characteristics and Long-term Management of Digital Objects** – deals with fundamental issues in digital curation such as authenticity, context and acceptable loss, but also with the automatic extraction of information from the digital objects themselves which enable their management and use.

- **Theme 4: Emerging Research Domains** – aims at eliciting new innovative methodologies for digital preservation and curation, drawing on experiences from related domains such as computer science and engineering, cultural heritage or cognition.

- **Theme 5: Measurement and Experimentation** – focuses on ensuring future accessibility of digital objects, especially on the role of experiments for objective evaluation of preservation strategies.

3.1. QUESTIONS

To keep the DPE online Delphi survey easily accessible, the intuitive questions were identical for each theme:

- **1**: In your view, in the past five years what have been the major innovations to arise from digital preservation research under this theme?

- **2**: What do you envisage could/should be achieved in this research area within the next 5-10 years?

- **3**: What are the major impediments that need to be addressed in order for the community to produce innovative research results that can be translated into practical applications?

- **4**: Where can you see areas in which Europe could demonstrate research leadership that would facilitate digital preservation and curation?

- **5**: What types of international collaboration(s) would be necessary to promote innovation and in what research be effectively realised under this theme?

- **6**: Please add any other information, advice, etc. you consider would help DPE to shape, deliver, and ensure take-up of this Research Roadmap.

3.2. FEEDBACK

Feedback to the online survey was unexpectedly low, especially for themes 3-5. Despite announcements in various mailing lists related to digital preservation, only few viable answers to the themed questions have been provided. Without any statistically significant amount of input, our analysis is restricted to brief qualitative summarizations of the findings and recommendations at theme-level.

Theme 1: Digital Objects and Collections

Progress has been made in packaging standards such as METS and XFDU, file format characterisation and format registries such as PRONOM or GDFR. Within the next 5-10 years, digitisation is expected to be a preservation solution for endangered audiovisual materials. Furthermore, common understanding of accreditation needs will evolve, and globally resolvable persistent identification of archived digital objects will be established. Systems will be less expensive and less complex to run, and also exhibit improved scalability. More usage of open-source software is expected. Current impediments are scattered implementation efforts and resources, high costs for maintenance but also the lack of available ICT research expertise and universities, the lack of international standards and the lack of synergies between different countries and organizations have been mentioned. At European level, sustained research into technological preservation solutions, sustained support for innovative organisational approaches, and research in downsized solutions for local uptake by smaller curation entities is recommended. Conferences and international events might raise awareness on the problem of standards and foster cooperation between institutions. It has been stressed that effective international partnerships between practitioners and researchers need to be established. Concentration of expertise through collaborations around testbeds is also advised, in addition to the equal involvement of players of varying size to ensure that solutions scale to the actual needs.

Theme 2: Management of Digital Repositories

A good uptake of OAIS, an understanding of the urgency of preservation planning, the need for trusted repository organisations, and the widespread emergence of practical repositories such as PORTICO, the BL's DOM or highly distributed solutions like LOCKSS are seen as the major innovations over the past five years. Further elaborations in the practical implementations of OAIS repositories are expected for the next 5-10 years, in addition to improved training and best practices the creation of an internationally trusted non-profit organisation dealing with IPR, DRM and content licensing. The need to capture context and associated metadata is seen as major impediment, whereas the current focus on OAIS and certification is considered to inhibit practical actions (in contrast to statements by other contributors). Competition at international/institutional level and lack of communities prevent the sharing of knowledge and best practices. European leadership requires action, particularly in national archives and libraries, but also in scientific data centres. Additionally, the cross-fertilization of fundamental (testbeds) and applied research (local implementations) in areas such as preservation planning, technology watch, preservation actions, etc. is stressed as important for demonstrating research leadership in digital preservation and curation. International collaborations are described as helpful, but expensive. Exchange should be established at practitioner level, i.e. between communities of implementers and between implementers and testbeds. A final recommendation suggests having something like "OAIS-lite" to foster progress in digital preservation.

Theme 3: Key Characteristics and Long-term Management of Digital Objects

The major innovations in this theme are depicted as the specification of metrics and criteria for audit and certification services. Over the next years, methods for establishing integrity, authenticity and trust for digital objects such as digital signatures, persistent identifiers, and audit trails are expected to be established. More robust, less expensive and easy-to-use applications based on open-source solutions are considered to gain in importance. Generic solutions and approaches are impeded by organisational and sector disparity, but also due to a lack of financial investments. Methods for establishing integrity, authenticity and trust (see above) are seen as major vehicles for the demonstration of European leadership in the long-term management of digital objects. Additionally, more research and testing might result in the launch more applied solutions ("beta programmes"). To do so, more EC projects in the area of digital preservation should be funded, but the collaboration between institutions for sharing information and experiences also needs to be improved.

Theme 4: Emerging Research Domains

More coordination of efforts among partners, countries and non-profit institutions regarding emerging research domains should be established in the near future. This means that political, social and personal conflicts between countries, institutions and individuals need to be overcome. Focusing on their own interests, organisations are often very protective of their own culture and knowledge assets. The creation of centres of excellence for digital preservation and curation might be an instrument to overcome these organisational gaps and to demonstrate consolidated European perspective. This should be supported and complemented by more research and joint projects.

Theme 5: Measurement and Experimentation

Regarding measurement and experimentation, more cooperation is regarded as the main achievement for the next decade. Again, financial barriers seriously impede the deployment of tools and testbeds. Conferences and demo centres appear as suitable means for demonstrating European leadership in this area. Hence international collaboration is fostered by specific themed events, including workshops and round table meetings.

3.3. DISCUSSION

Without a solid statistical basis, the results of the online delphi survey inevitably remain inconclusive. Overall, the survey input is consistent with the results from the analyses of previous research agendas; hence the interpretation of the answers to the five themes leads to comparable conclusions.

The goal of the DPE online delphi survey was to involve the community in the very definition of the core research issues in digital preservation. But unfortunately, the community suffers from "online survey fatigue", i.e. too many surveys by numerous projects (EC and international) related to heterogeneous aspects of digital preservation and e-infrastructures have been going around. Even within the DPE survey results, feedback to the higher-order themes was considerably lower.

Overall, the low amount of feedback indicates that within its context, the online survey has failed as an instrument and we need to resort to alternative means of community validation. For instance, the DPE user forum[4] provides an excellent community platform for the discussion of research issues, but also roadmap sessions at various conferences appear to be more appropriate. The first roadmap session was held at iPRES 2007 conference[5] with over 100 delegates attending. After listening to an introduction to the research roadmap delegates

---

[4] http://www.digitalpreservationeurope.eu/forum/phpBB2/viewtopic.php?t=8

[5] http://ipres.las.ac.cn/

were asked to spend around 20 minutes list the research issues they consider to be of key importance. Following this members of the audience presented their thoughts and discussions followed. The lists of themes gathered from this session were in both English and Chinese and are currently undergoing translation before analysis can be completed. Further sessions are planned, for example at the EVA/MINERVA conference in November 2007, and will be organized and reported on in cooperation with WP1 and WP6.

Note: The DPE online delphi survey was experimental, designed and implemented by the partners involved as an additional means to explore and assess the (future) digital preservation landscape. It was not specified in the DPE technical annex.

4. DPE RECOMMENDED RESEARCH

The analysis of the last 16 years of effort in the area of digital preservation and the input from the online delphi survey support our claim that while much work has been done on the periphery, the central problems of digital preservation have yet to be resolved. We consider that although access, digital rights management, users and other aspects make digital preservation useful, this is outside the scope of core research.

4.1. RESTORATION

Digital objects are preserved as bit streams on some type of storage media. Despite great care, the integrity of these digital objects might be compromised, perhaps due to some external event. Forensic methods can be used to recover data from physically damaged media. However if the type of the salvaged objects is unknown, and we do not know whether the bit stream represents a program, a file, a group of files or a scientific data set this is not an adequate solution. Rendering such objects and making their content known remains a significant challenge for digital preservation and still requires much research into the reconstruction of digital objects.

Future research to support restoration can take advantage of methods and findings in information and communication theory. One such area of work is the encoding algorithms of multimedia bit streams that have to live with loss of parts of the stream during transmission. These can, depending on the effort spent during the encoding process, partially or even fully recover from such situations by building on error correction mechanisms that have been long established in the field of coding theory.

4.2. CONSERVATION

Digital collections need to be safeguarded to allow for continuous consultation of the original sources by future generations. To cope with the obsolescence of technology, strategies such as migration or emulation have emerged. Virtualisation approaches to data interpretation such as Lorie's Universal Virtual Computer (UVC) or the Multivalent Document Model[6] extract data from the bit stream and return it to the user in an understandable way. If programs are written against the UVC, only the implementation of the UVC is required to run the interpreters (or media adapters) and thus read the data on future computers. But little practical evidence has been gathered so far on their real-life application.

Results from database management and knowledge representation research can be utilised to automatically represent constraints related to outdating of platforms and/or data formats. It can also be used to support automatic identification and transcoding of content that is under the risk of being no-longer renderable because of change in technology. Furthermore, content harmonisation, information extraction and semantic annotation technologies that support the transformation of proprietary and outdated formats into state of the art meta formats such as XML-, RDF- and OWL-based formats are at least continuously emerging or even already existing for various document and media storage formats.

4.3. MANAGEMENT

The management of digital repositories is based on the understanding of the organisational requirements for long-term preservation of digital material. However, little is know about how to assess the current state of a digital collection with respect to the obsolescence of software, hardware environment and file formats. Effective mechanisms, including appropriate support tools, for monitoring and preservation planning are required as integral parts of repository management.

Methods from configuration management of software installations can be utilised to represent dependencies between platforms, tools, formats, and technologies. If coupled with appropriate metadata representations and knowledge technologies that allow for automated reasoning about dependencies within configurations of platforms, tools, and formats, this can support the management of digital repositories even in automatic or semi-automatic ways. Finally standardisation of appropriate change management protocols between independent software vendors of various types of platforms, tools, and formats seem important in this area. Research needs to focus on the planning, enacting, executing, managing, and monitoring of organisational processes for Digital Preservation. Methods from workflow management and support and collaboration support can be utilised.

4.4. RISK

In essence, digital preservation is a risk management problem. Hence decision making instruments are needed which will help to translate the uncertainties involved in digital preservation into quantifiable variables if possible, allowing one to take appropriate measures balancing costs and benefits. Risk mitigation in digital preservation might benefit from the research in business engineering and computer science which applies advanced automated

---

[6] see [43] for a description of Multivalent and an assessment of UVC

techniques such as categorisation and machine learning for the assessment of current risks and the selection of corresponding strategies.

Appropriate metrics for various types of risks as well as for their economic and other consequences have to be defined. Algorithms should also be developed that support the measurement of various types of risks on the basis of such metrics. Methods of information visualisation can be utilised to make these risks more easily understood by humans. Furthermore, methods of knowledge representation and reasoning can be utilised to represent these risks in an explicit, machine readable form that can be automatically processed and analysed by applying methods of machine learning and automated reasoning.

4.5. SIGNIFICANT PROPERTIES OF DIGITAL OBJECTS

Preserving digital objects requires an understanding of the significant properties that need to survive in order to keep the object usable, meaningful, authentic and reliable. In part this depends on the chosen file format of the objects, but most significant properties are determined by the business context in which they were created and used. In order to be useful for future users they need to understand the creator's intention behind the digital object. As this intention is reflected in the way the object is shaped and structured these implicit characteristics need to be identified so they can be preserved through time.

Although methods are currently being developed to enable this, not much practical experience has been gained in this area. Additional fundamental research and practical experiments, covering the many different types of digital objects, are needed to gain a thorough understanding of the underlying issues.

Research in this area may benefit from existing results in information extraction and annotation as well as from organisational and business process context representation and automated processing.

4.6. INTEROPERABILITY

While most of the digital preservation research to date has focused on individual types of digital objects, in practice repositories deal with vast collections of digital objects and their metadata. Persistent business models and standardisation for digital objects and metadata formats and interaction protocols represents the key to preservation of service and content and achieving interoperability between generations of systems.

Trust is also an essential issue for digital repositories, but metrics and criteria for defining audit and certification services are at best in their infancy. Within a federated environment

this trust must extend to the reliability and quality of service offered by the institutions which constitute the aggregated service infrastructure.

4.7. AUTOMATION

The level of automation in digital preservation is still very low. One of the reasons for this may be the lack of a systematic approach to digital preservation. In order to be able to cope with the huge volumes of the digital objects now requiring preservation, our appraisal, acquisition, preservation action and metadata extraction processes must be automated. More efficient and cost-effective work models and tools are required that are scalable to different types of digital objects and organisational environments.

4.8. CONTEXT

The context of a digital object is the representation of known properties associated with it and the operations that have been carried out on it. The context of preservation has a major impact on the way digital objects should be managed and preserved through time. To enable proper preservation management and policies a thorough understanding of that environment will be necessary. This may include information concerning available technologies, legal/societal requirements, existing knowledge and user requirements. This contextual emphasis would mean that there is no universal or ideal preservation strategy. Little research has been done in this area so far.

To exploit this contextual knowledge, reasoning mechanisms might be employed to assess the current state of a digital object, providing guidance in taking rational decisions. With machine-interpretable semantics and appropriate configuration, digital objects might even become self-aware, able to take rational, autonomous actions based on their perception of the context representation. Self-aware digital objects might be able to detect the obsolescence of their content format and pro-actively invoke corresponding migration services.

4.9. STORAGE

Although Digital preservation is strongly reliant on archival and library traditions, it also borrows heavily from other related disciplines. Indeed, it has often been the case that developments and innovation in other subject areas has driven digital preservation research forward. In particular current issues with the scalability and capacity of storage solutions may be resolved by developments emerging in science communities such as GRID computing and

eScience initiatives and serve as future building blocks in distributed digital preservation infrastructures.

4.10. EXPERIMENTATION

Experimentation will play a fundamental role in ensuring the future accessibility of digital objects through the objective evaluation of preservation strategies and for research in all the areas mentioned above. Digital objects have to be characterised and associated with appropriate preservation plans. The deployment of testbeds will enable the development of evidence-based methods. Experimentation will also help understand users' interactions with digital repositories and evaluate the impact of new support mediums such as information visualisation techniques for collection and access management.

5. CONCLUSION AND RECOMMENDATIONS

After 16 years of research and the continuous refinement of a research agenda, very little progress has actually been made during this time. The research agenda that we propose reflects this and stresses the need to address the research themes that support the concept of a Collective Memory of the future. The virtual Digital Library infrastructure that will underlie this Collective Memory should include seamless and cost-effective Digital Preservation support for distributed, collaborative archival and preservation teams.

The first step should be to examine and utilise emergent trends in next generation ICT infrastructures and architectures such as semantic-enabled information infrastructures, grid-based resources, service infrastructures and service oriented architectures. The remaining technological challenges related to utilising such semantic-enabled and grid-based infrastructures for a establishing a future Collective Memory platform should also be identified. This will allow Digital Library infrastructures that inherently include transparent, effective and efficient Digital Preservation support.

Other priorities in achieving this vision are: to enable cost-effective access to Digital Preservation technologies, to support metadata management and metadata brokering including context and semantics, to enable the integration of Digital Preservation services and information collections as grid services, to support the dynamic definition of Digital Preservation support workflows refining OAIS and utilizing web service architectures, and to enable advanced data and information access support services refining OAIS in its respective fields.

Furthermore research activities should focus on developing innovative and value-added services which support collaborative and distributed archival and preservation teams in the creation and management of a Collective Memory. In particular services to support such tasks as the selection and integration of relevant resource collections, the selection and adaptation of Digital Library and Digital Preservation services as required, the definition of preservation-specific processes for information selection/extraction, information annotation and processing, information syndication and publishing, the semantic annotation of content (based on RDF, RDF Schema, OWL), collaboration services (based on CSCW Technologies), advanced services for context-based personalized access, and dynamic workflow support (based on BPEL, BPML) should be designed, implemented and evaluated appropriately.

Adequate standards which refine and implement all areas of OAIS (based on XML, RDF, OWL, SOAP, WSDL, BPEL, BPML, XML Signature, WSS SAML) for organizing a

Collective Memory should be developed and made operational in reference architectures, frameworks, and implementations of application solutions.

Finally, flexible and dynamic distributable architectures for a Collective Memory platform must support on-demand retrieval, enactment, activation, and access to Digital Preservation resources. They should also support the dynamic definition, execution and control of tailored information access processes, such as contextualized publishing based on information enrichment and resource based contextual information visualization, and the ad-hoc tailored configuration of Digital Library services into an application solution.

In conclusion we must, as a community, examine and utilize emergent trends in next generation ICT infrastructures and architectures, focus on developing innovative and value-added services which support the work of collaborative and distributed archival and preservation teams, develop and make operational standards which refine and implement all areas of OAIS and finally develop flexible and dynamic distributable architectures for a Collective Memory platform must support on-demand retrieval, enactment, activation, and access to Digital Preservation resources.

6. REFERENCES

[1] David Giaretta, Heather Wevaer: *Digital Curation and Preservation: Defining the research agenda for the next decade*. Report of the Warwick Workshop, 7 & 8 November 2005.

[2] Seamus Ross, Margaret Hedstrom: *Preservation research and sustainable digital libraries*. Int. J. on Digital Libraries 5(4): 317-324 (2005).

[3] *It's About Time: Research Challenges in Digital Archiving and Long-term Preservation*. April 12-13, 2002. Sponsored by the National Science Foundation and the Library of Congress.

[4] *Invest to Save: Report and Recommendations of the NSF-DELOS Working Group on Digital Archiving and Preservation*. Prepared for: National Science Foundation's (NSF) Digital Library Initiative & The European Union under the Fifth Framework Programme by the Network of Excellence for Digital Libraries (DELOS), 2003.

[5] *The State of Digital Preservation: An International Perspective*. Council on Library and Information Resources Washington, D.C., July 2002.

[6] *The Future Digital Heritage Space: An Expedition Report*. DigiCULT Thematic Issue 7, December 2004.

[7] *Memorandum on the long-term accessibility of digital information in Germany*. March, 2006.

[8] Joint NDIIP/JISC Digital Preservation Workshop. 7-9 May 2006, Washington, DC.

[9] Margaret Hedstrom: *Understanding Electronic Incunabula: A Framework for Research on Electronic Records*. The American archivist, 54, S. 334-355, 1991.

[10] Thomas Severiens, Eberhard R. Hilf: *Entwicklung eines Beschreibungsprofils für eine nationale Langzeit-Archivierungs-Strategie* (National long-term Preservation Policy). November 2005.

[11] John McDonald: *Towards a National Digital Information Strategy: A Review of Relevant International Initiatives*. Information Management Consulting and Education, DRAFT, March, 2006.

[12] Monika Krimbacher, Michael Neuhauser, Martina Vogl: *Survey On The Long-Term Preservation Of Digital Documents* In European Libraries 2005. Report from the eContent project reUSE.

[13] Ingeborg Verheul: *Networking for Digital Preservation: Current Practice in 15 National Libraries*. IFLA Publication, München, Saur, 2006.

[14] Kevin Ashley: *Digital Preservation Research: A review of the challenges, An Expedition to European Digital Cultural Heritage Collecting, Connecting - and Conserving?*, Salzburg, 2006.

[15] Denise Lievesley, Simon Jones: *An Investigation into the Digital Preservation Needs of Universities and Research Funders*, BLRIC Report no. 109, 1998.

[16] Philip Lord, Alison Mcdonald: *e-Science Curation Report*, JCSR Report, 2003.

[17] John Garrett, Donald Waters (co-chairs): *Preserving Digital Information: Final Report and Recommendations*, Commission on Preservation and Access and the Research Libraries Group, 1996

[18] *Revolutionizing Science and Engineering Through Cyberinfrastructure*. Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure, January 2003.

[19] Andrew Fitzgibbon, Ehud Reiter: *Memories for life: Managing information over a human lifetime*. written as part of the UK Computing Research Committee's Grand Challenges exercise.

[20] *Grand Challenges in Computing Research*. Grand Challenges 2006 Conference Report, 22-24 March.

[21] Tony Hey, Anne Trefethen: *The Data Deluge: An e-Science Perspective*.

[22] K. J. Peach : *The Impact of e-Science*.

[23] Fran Berman: *Cyberinfrastructure - Issues and Challenges*.

[24] *Management im virtuellen Forschungsraum* (in German). Wissensmanagement Special Ausgabe 01/05.

[25] Kieron O'Hara, Wendy Hall, Keith van Rijsbergen, Nigel Shadbolt: *Memory, Reasoning and Learning*. Morris, R., Tarassenko, L. and Kenward, M., Eds. Cognitive Systems: Information Processing Meets Brain Science, chapter 10, pp. 236-260. Elsevier, 2006.

[26] Michael P. Papazoglou, Paolo Traverso, Schahram Dustdar, Frank Leymann, Bernd J. Krämer: *Service-Oriented Computing: A Research Roadmap*. Francisco Cubera, Bernd J. Krämer, Michael P. Papazoglou, Eds. Service Oriented Computing (SOC), Dagstuhl Seminar Proceedings, 2006.

[27] *Technology Planning for Business Competitiveness*. Emerging Industries Occasional Paper 13, August 2001.

[28] Ronald N. Kostoff, Robert R. Schaller: *Science and technology roadmaps*. IEEE Transactions on. Engineering Management 48, 132–143 (2001).

[29] Rachel Heery, Andy Powell: *A Digital Repositories Roadmap: looking forward*, April 2006.

[30] *INNOVANET: Roadmap for Systematic Innovation and Scientific Discovery*. Deliverable 6, November 2003.

[31] *European Roadmap on Research Infrastructures (Report 2006)*. European Strategy Forum on Research Infrastructures (ESFRI), September 2006.

[32] *Rules for making the roadmap*. European Strategy Forum on Research Infrastructures (ESFRI), 2006.

[33] Electronic Publications, Electronic Documents and Virtual Information (IFLA), http://www.ifla.org/VI/6/dswmedia/en/bib_elec.htm

[34] Research Agendas (NHPRC), http://www.archivists.org/news/NHPRCinfo2.asp

[35] Preserving Digital Information, http://www.rlg.org/ArchTF/ref.html

[36] Technology Roadmapping - Resources, http://www.ifm.eng.cam.ac.uk/ctm/trm/resources.html

[37] Neil Beagrie: *National Digital Preservation Initiative: An Overview of Developments in Australia, France, the Netherlands, and the United Kingdom and of Related International Activity*, copublished by the Council on Library and Information Resources and the Library of Congress, 2003.

[38] John McDonald: *Toward a Canadian Digital Information Strategy: A Review of Relevant International Activities*, Information Management Consulting and Education, Version 2.0, March 2006, http://www.collectionscanada.ca/cdis/012033-800-e.html

[39] Ray Edmondson: *Audiovisual Archiving: Philosophy and Principles*, UNESCO, 2004

[40] D. J. Rayers: *Metadata in TV production: Associating the TV production process with relevant technologies*, BBC R&D White Paper, WHP 046, 2002.

[41] *Reference Model for an Open Archival Information System (OAIS)*. Blue Book. Issue 1. January 2002. This Recommendation has been adopted as ISO 14721:2003.

[42] Thomas A. Phelps, Paul B. Watry: *A No-Compromises Architecture for Digital Document Preservation*. ECDL 2005: 266-277

[43] Neil Beagrie: *E-Infrastructure Strategy for Research; Final Report from the OSI Preservation and Curation Working Group*, Office for Science and Innovation, 2007.

[44] K. Schroeder: *Persistent Identification for the Permanent Referencing of Digital Resources - The Activities of the EPICUR Project: Enhanced Uniform Resource Name (URN) Management at Die Deutsche Bibliothek*, In The Serials Librarian; Volume 49, Issue 3; Mai 2006 : page 75-87.

[45] Schroeder, Kathrin: *Zum Nutzen von "Persistent Identifiers": von CARMEN-AP4 über EPICUR zum Normalbetrieb*, In Zeitschrift für Bibliothekswesen und Bibliographie Jg. 52 (2005), H. 3-4, page 181- 187.

[46] M. Lautenschlager, I. Sens (2003): *Konzept zur Zitierfähigkeit Wissenschaftlicher Primärdaten*, In Information : Wissenschaft und Praxis 54, 2003, Heft 8, page 463-466. http://www.mad.zmaw.de/fileadmin/extern/Publications/Paper-Konzept-Primdaten.pdf

A   Ongoing Activities

A.1   National and International Efforts

The nestor report by Thomas Severiens and Eberhard R. Hilf on a national long-term preservation policy [10] served as the basis for the characterisations of ongoing activities in digital preservation. Another overview of recent developments has been published by Neal Beagrie [37].

Australia

Due to its federal organisation there is no national strategy for digital preservation in Australia.

In 1996 the National Library of Australia (NLA) started the PANDORA project[7], a collection of techniques (for example the TRIM software suite for word processing formats), policies and proprietary workflows for digital preservation. Human experts are responsible for the selection of material, according to a documented catalogue of criteria.

The PADI (Preserving Access to Digital Information[8]) project, also driven by NLA, aims at international cooperation between national and international institutions which are actively engaged in digital preservation. It provides a subject gateway containing collections of links and information dealing with various aspects of digital preservation.

Its objectives are: to facilitate the development of strategies and guidelines for the preservation of access to digital information; to develop and maintain a web site for information and promotion purposes; to actively identify and promote relevant activities; and to provide a forum for cross-sector cooperation on activities promoting the preservation of access to digital information.

There exists a memorandum of understanding with the British Digital Preservation Coalition (DPC), the State and University Library in Göttingen (SUB), the EU project Erpanet and the German nestor project.

---

[7] http://www.nla.gov.au/policy/plan/pandora.html
[8] http://www.nla.gov.au/padi

Denmark

In 2003 the Danish Ministry of Culture published a report on preservation of Denmark's cultural heritage "Bevaring af kulturarven". Among other things the report covers the current status on preservation strategies for physical as well as digital objects[9] [10].

In the spring of 2006 the Danish Ministry of Culture established a group containing representatives from the major archives, libraries and museums in Denmark with a focus on digital preservation. The aim is to coordinate the national efforts and maintain international contacts with regards to digital preservation. Among the efforts will be collection and dissemination of knowledge and expertise on digital preservation, coordination and initiation of projects as well as support participation in international activities.

The State and University Library and The Royal Library are developing, under the aegis of Denmark's Electronic Research Library, a preservation service for Danish institutional repositories.

The State and University Library and The Royal Library have developed and implemented a web-archiving system[11] for the Danish portion of the World Wide Web. The system features a distributed bit-preservation system based on a message-passing architecture. Future plans include making an open source version of this software available.

France

The French National Library (Bibliotheque national de France, BnF) prescribes workflows for archiving printed and digital documents. The archival system is centralised, implementing the binding national archiving policy. Other institutions involved in digital preservation are CNES (Centre Nationale d'Etudes Spatiale) which has been involved in the development of the OAIS model and CNRS (Centre Nationale de la Recherche Scientifique). CNRS together with the national pre-print archive CCSD, is participating in the ETOL project (European Thesis Online) aiming at the long-term preservation of dissertations.

The HAL project (Hyper Article En Ligne) currently under development deals with the interlinking of archived material with current literature collections, and also integrating the BnF workflows.

---

[9] http://kum.inforce.dk/graphics/kum/downloads/publikationer/bevaring_af_kulturarven.pdf

[10] http://www.kum.dk/sw5954.asp

[11] http://www.netarkivet.dk/

Germany

The German Federal Ministry of Education and Research (BMBF) is funding "nestor - Network of Expertise in Long-Term Storage of Digital Resources"[12] to bring parties together who are concerned with the preservation of digital resources. The aim of nestor is to foster a cooperative infrastructure in which various specialists can work on a solution to the problem of digital preservation.

BMBF also funds kopal[13] a cooperative project implementing a operating preservation system. The goal of the kopal project is to develop a technological and organisational solution to ensure the long-term availability of electronic publications. Thereby, the transparent integration into existing library systems and the reusability through memory institutions play a critical role.

There are several other research projects on digital preservation in Germany. Most of which are listed in the nestor "Projects section"[14].

Italy

The Digital stacks project, funded in 2006 by the Digital Renaissance Foundation (Fondazione Rinascimento Digitale) and the National Central Library of Florence of the Ministry of Cultural Affairs (Biblioteca Nazionale Centrale di Firenze) is one of the main national projects on digital preservation. It aims to test the storage of a considerable amount of data, 10 terabytes, to establish a sustainable contribution to the viability of digital resources.

The Digital stacks project is setting up the following test infrastructure: 1) Ten rack-mounted personal computers have been installed in accordance with multi-site architecture principles: five at the National Central Library of Florence, and five at the National Central Library of Rome; 2) An open source operating system has been selected; 3) Data replication relies on an open source disk synchronization utility to avoid hardware dependences; 4) A third site will act as a dark archive for disaster recovery purposes, based on a different technology to enhance the security of the overall system; 5) Five copies of every file are replicated: two in Florence, two in Rome and one copy in the dark archive.

---

[12] http://www.digitalpreservation.de

[13] http://kopal.langzeitarchivierung.de

[14] http://www.langzeitarchivierung.de/modules.php?op=modload&name=PagEd&file=index&page_id=16

This architecture is scalable and easy to maintain: the storage can be increased without any particular hardware, software or vendor constraint. During 2007, the project will extensively test this technical infrastructure and will also take into account further aspects of digital preservation, to integrate the available tools in the field of preservation strategies. Certification of the Digital stacks will also be an important part of the project.

Another important Italian project concerns the conservation of public administration documents, following up the Digital Administration Code of March 7th, 2005. The objective was the elaboration of guidelines for the long term preservation and management of public documents from different sectors. In 2006 the technical work group of the project, created by CNIPA. The national centre of IT for Public Administration, published a document where they established some of the necessary conditions to make digital preservation possible. More information can be found on the CNIPA web site[15].

Austria

There are several key institutions involved in digital preservation research in Austria, among them the Austrian National Library (ONB) and the Vienna University of Technology (TUW). The ONB is an important strategic stakeholder, mainly active in the preservation of its photograph archive and access to digital newspapers and access to Austrian electronic theses and dissertations. The Vienna University of Technology is active in the areas of Web Archiving, Home Archiving, and Preservation Planning. TUW also participates in several major national and international research projects. The Austrian Broadcasting Association (ORF) has been active in the digitizing of its video archive, and participates in the PRESTOSpace project. The 'Oesterreichische Computergesellschaft' (OCG) has a working group 'Digitale Langzeitarchivierung' which brings together research institutions and industry partners to coordinate national initiatives and set directions for future research in digital preservation. The 'Ludwig Boltzmann Institut' is works on the archiving and publication of multimedia art, specifically for the holdings of Ars Electronica. Salzburg Research is a partner in the Digicult project and has core expertise in the fields of digitization and the delivery of cultural heritage.

Great Britain

While numerous national and local projects for digital preservation have been conducted in Great Britain, the short term nature of their funded has led to a limited take up of their findings.

In 2001 the Digital Preservation Coalition (DPC) was established for developing strategies for the implementation of digital preservation, resulting, e.g., in guidelines dealing with various aspects. Digital preservation projects being undertaken by DPC members are listed in the DPC website http://www.dpconline.org/graphics/join/projects.html.

JISC (The Joint Information Systems Committee of the UK) funds a programme for digital preservation and document management. Each JISC programme consists of a number of related projects, which JISC funds within further and higher education institutions and JISC services. These projects are listed in JISC website: http://www.jisc.ac.uk/whatwedo/projects.aspx

Czech Republic

The Czech Republic has a substantial record of achieving remarkable results in the area of digital archiving recognized worldwide. In 2005 the National Library of the Czech Republic was awarded the first UNESCO/Jikji Memory of the World Prize for its contribution to the preservation and accessibility of its documentary heritage. Large national projects have been functioning for over ten years, combined with other regional, subject oriented, and institutional projects. These individual projects have produced millions of digitized pages and accompanying metadata. They now face the challenge of fostering closer cooperation and coordination among these projects under the framework of the Czech Digital Library, which conceptualizes a new national integration of the different library repositories in this country with repositories of other cultural heritage institutions. The main goal of the Czech Digital Library was to address the urgent necessity to preserve and make permanent access to collections of traditional, digitized, and digitally created documents as an important part of the Czech cultural heritage.

Thanks to large national grant projects funded by the Ministry of Culture, digitization projects started in the early '90s and completed the archiving of our .cz domain in 2000. They have about 3 million digitized pages and about 156 million archived web documents. These documents are accessible via the three subsystems, Manuscriptorium, Kramerius and WebArchiv. Manuscriptorium aims to build a virtual research environment for the realm of historical book resources. Manuscriptorium represents the current state of our achievements[16]. "Kramerius" project focuses on preservation of and accessibility to endangered "modern"

---

[15] http://www.cnipa.gov.it/site/it-IT/Attivit%c3%a0/Dematerializzazione/

[16] http://www.manuscriptorium.com/Site/ENG/

periodicals, books, and other documents[17]. WebArchiv has two main objectives: first, to catalogue some online publications as a part of the national output and to integrate their records into the Czech National Bibliography database; and second, to store these publications in a deposit system so as to ensure long-term access to them and to preserve the whole .cz domain by ongoing harvesting[18].

Three different repositories have hosted the separate data of Manuscriptorium, Kramerius, and WebArchiv. From 2007 onward, the digital data of these three subsystems will physically share one large repository. They plan to integrate Kramerius and WebArchiv in 2008 under the umbrella of one strong international and tested operational system and to develop Manuscriptorium to cater to the specialised needs of researchers by bringing together fragments of historical documents dispersed in different countries and continents.

The Netherlands

The key players in digital preservation in the Netherlands are the National Archives and the National Library (Koninklijke Bibliotheek / KB). Other institutes that are working in the area of digital preservation are DANS and IISG.

The National Library has an e-Depot, based upon the OAIS model, that has been developed in collaboration with IBM. This archival system for digital publications which is continuously extended in various projects. Furthermore it has a National Preservation Office (NPO), which maintains contracts with the major publishers located in the Netherland, for which it provides preservation services. Additionally, KB has driven the European NEDLIB project (1998-2001), which published guidelines for digital repositories.

The National Archives is and has been involved in different international projects, for instance the ERPANET project (2001-2004), the InterPares project (1998-2006), the Delos project (2004-2007) and more recently PLANETS and Digital Preservation Europe. One of the main projects for the development of techniques and strategies for digital preservation the National Archives has driven, was a testbed for evaluating digital preservation strategies (Testbed Digitale Bewaring; 2000-2003). Finally the National Archives is involved in international standardisation work through ISO in the area of records and archives management. The particular ISO committee, TC46/SC11, has published standards for records management (ISO 15489:2001) and metadata for records (ISO 23081-1:2006 and ISO 23081-2:2007).

---

[17] http://kramerius.nkp.cz/kramerius/Welcome.do?lang=en

Baltic States

Estonia – The development of a strategy for digital preservation of cultural heritage in Estonia began in 2003 with the formation of a working group. In 2004 the Minister of Culture approved The National Strategy and The Action Plan for the Digital Preservation of Estonian Cultural Heritage, 2004-2007. Competence centres for digitisation, metadata management and preservation have been established to help the memory institutions to achieve and implement common standards for the digitisation and description of cultural heritage. The next strategy has been approved for 2007-2010 where the main aim is to develop a common digital storage facility for digital cultural heritage materials from all memory institutions.

Some of the most significant digitisation projects to date are the "Red Book of endangered cultural heritage objects" which aims to study the physical condition of objects in all memory institutions in order to establish preservation priorities, and the "Red Book of Estonian Publications 1535-1850", dedicated to the preservation of document heritage[19]. Other recent projects include: DEA– Digitised Estonian Newspapers 1821-1944[20], SAAGA[21], 17th century imprints in the Estonian Historical Archives[22] and AIS (Archival Information System)[23].

The National Library of Estonia participated in the EC funded project ReUSE and has established a digital repository "Digar" for the collection, preservation and dissemination digital publications. The National Archives of Estonia has started the development of its digital archive management system software. The first stage of this has been the development of a high-level model and architecture for the system. The National Broadcasting Company has received significant funding to create, store and keep its TV programmes in digital form.

Latvia - The Latvian cultural policy model is organised around the Ministry of Culture but there is no separate National policy and strategy for digital preservation of cultural heritage. The National Board of Culture established in 1995 is the most important advisory body to the Ministry of Culture. National Programme Culture was developed in 2000. Documents for further culture sector development are for example: 'Long Term Cultural Policy Guidelines 2006-2015', and 'Medium Term Strategy 2006-2008 for the Cultural Sector'.

---

[18] http://en.webarchiv.cz/

[19] http://www.nlib.ee/PunaneRaamat/otsieng.html

[20] http://dea.nlib.ee

[21] http://www.eha.ee/saaga/index.php?lang=eng

[22] http://www.eha.ee/plakatid/

[23] http://ais.ra.ee/ais/

The most significant Latvian projects are 'Museum Collections on the Internet'[24], HERITAGE –1: Preservation of Latvian Periodicals[25], Poster in Latvia 1899-1945 and Poster in Latvia 1945-2000[26], Latvia in the 16th -18th Century Maps[27], VIRMUS (Virtual Museum)[28], History of the Culture of Latvia in Images and Latvian History of Civilization in Images. Part A: Portraits[29],.

Lithuania - Conception of digital preservation of cultural heritage was enacted by the State government in 2005. In the same year the Council for Digital Preservation of Cultural Heritage at the Ministry of Culture was established. However there is no national strategy for Digital Preservation in Lithuania. Digital preservation activities are implemented mostly in libraries, as well as in museums, archives and academic institutions.

ARUODAI[30], a significant digital preservation project, is the electronic compendium of sources on Lithuanian spiritual culture: ethnology, language and history. This repository will provide the means to preserve and analyse language, folklore and ethnological, archaeological, and historical data. BARIS[31] is active in the digitalization of the historical documents of the Vilnius Church Province. The register of cultural heritage valuables[32], Information system of Lithuanian museums (LIMIS)[33] and MUSICALIA (part of the EUREKA project ECH:TOPICC)[34], provide direct access to the digital archive of the descriptions and images of the musical manuscripts preserved in Lithuania. Creation of the integrated virtual information system of the libraries (under creation). The main goal of the project is to give people of Lithuania and beyond the opportunity to use the full text documents stored in Lithuanian libraries, museums and archives interactively and free of charge. The project is implemented by National Library of Lithuania, Department of the Archives and Museum of Art.

---

[24] http://www.meandrs.lv/main.jsp?1171795987688&language=en

[25] http://www.lnb.lv/digitala_biblioteka/index_anglu.htm

[26] http://www.lnb.lv/digitala_biblioteka/Plakati/main_anglu.htm

[27] http://www.lnb.lv/digitala_biblioteka/kartes/main_anglu.htm

[28] http://www.muzejs.lv/en/about_virmus.htm

[29] http://www.lnb.lv/retumiprojekts/eng/main.htm

[30] http://www.aruodai.lt/

[31] http://www.kf.vu.lt/baris/

[32] http://www.heritage.lt/inside1.php?i=17

[33] http://www.muziejai.lt/Panorama.en.htm

[34] http://www.musicalia.lt/index.en.htm

United States of America

There are various institutions and organisation active in the area of Digital preservation in the USA are, the most prominent being the Library of Congress (LOC) and the Online Computer Library Centre (OCLC). In 1998, the LOC started deliberating on a national policy for digital preservation. An example for an institutional policy was developed in 2000 at the Columbia University Library[35].

The National Archives and Records Administration (NARA) is an independent Federal agency that preserves our nation's history and defines us as a people by overseeing the management of all Federal records. NARA ensures continuing access to essential evidence that documents the rights of American citizens, the actions of Federal officials, and the national experience[36].

The international PREMIS project (PREservation Metadata Implementation Strategies) focuses on metadata concepts for digital preservation. Its main goal is to evaluate strategies for the storage and description of digital objects within an archive and also the exchange of objects and their associated metadata between archives and archival systems. PREMIS finished in 2005, publishing its final report comprising a metadata dictionary with instructions.

The National Digital Information Infrastructure and Preservation (NDIIP) plan is currently being executed, building on results from, amongst others, the METASCHOLAR workshop and LOCKSS. LOCKSS (Lot Of Copies Keeping Stuff Safe) developed a peer-to-peer proxy network for persistent access to electronic journal articles which observes copyright regulations. In 2004 the LOCKSS project signed contracts for strategic partnerships within an integrated project. Each partner is responsible for specific aspects, e.g., CALTEC (California Digital Library at the University of California) for the development of web archival tools for the long-term preservation of political and government documents. Various government agencies, libraries, archives, research institutions, universities and commercial companies are also involved.

In 2002 the e-government act was extended by implementation specifications for the long-term preservation of government documents.

---

[35] http://www.columbia.edu/cu/libraries/services/preservation/dlpolicy.html

[36] http://www.archives.gov/

iRODS (i Rule Oriented Data Systems) is an adaptive grid middleware which has been developed since 2004 at the San Diego Supercomputer Center (SDSC)[37]. Funded by the National Science Foundation (NSF) and NARA, the iRODS development builds on expertise gathered in the application of the SRB (Storage Resource Broker)[38] data grid management system. iRODS follows the rule oriented programming paradigm (ROP) for encoding customisation of data management functionalities in a declarative fashion. Processes in the iRODS data grid system are coded as rules which explicitly control the conditions, under which particular operations (called micro services) are performed. An iRODS-based system is composed of heterogeneous, distributed storage and computing systems which are controlled by the evaluation of rules and execution of micro services.

European Union

The goal of the European Commission on Preservation and Access (ECPA) is to harmonise institutional and national efforts and conduct extensive surveys and studies on the state of digital preservation[39]. Furthermore it has promoted the development of search engines and common standards for the digitisation of content, and also on the use of multimedia digital archives.

Past European funded projects such as MINERVA (Plus) or the DIGICULT forum were mainly focused on the exchange of best practices, the establishment of benchmarks, testing for the interoperability of archival systems and promoting dialogue between all actors involved in digital preservation.

PRESTO and PRESTOSPACE were focused on the development of fast, high-quality digitisation methods for television archives. The MEDIA programme supports the formal indexing of digital film material for promoting digital technologies in the European audiovisual industry.

The eContentplus programme aims to reduce fragmentation in the market for digital content and improve access to and use of geographical information, cultural content and teaching material. The projects funded by eContentplus programme are on their website[40].

The SEPIA (II) project deals with the long-term preservation of photographic material, the exchange of experiences, public relations and semantic description methods. The Erpanet

---

[37] http://irods.sdsc.edu/

[38] http://www.sdsc.edu/srb/

[39] http://www.knaw.nl/ecpa/

[40] http://europa.eu.int/information_society/activities/econtentplus/projects/index_en.htm

project (Electronic Resource Preservation and Access Network) established a European platform for the exchange of information related to digital preservation.

UNESCO

The UNESCO Archives Portal provides documents such as the 'Charter on the Preservation of Digital Heritage' which has been ratified in 2003 by the general assembly of the UNESCO. 'The Guidelines for the Preservation of Information Digital Heritage' were published in the same year. The role of the UNESO with respect to the development and standardisation of international rules for digital preservation is currently under discussion.

UNESCO has launched the Memory of the World Programme to guard against collective amnesia calling for the preservation of the many valuable archive holdings and library collections all over the world to ensure their wide dissemination.

A.2   Selected Projects

Some project descriptions, with additions, have been taken from John McDonalds review on relevant international initiatives in digital information [38].

Danish State Archives: Electronic Records - strategies and requirements

A conversion strategy has been developed at the Danish State Archives for the long-term preservation of IT filing systems. The electronic archival materials which are submitted to the State Archives are required to be stored by ministries in a way that enables data conversion to formats suitable for continuous conversion without significant data loss[41].

National Archives of Norway: NOARK

Noark-4 is a specification of functional requirements for electronic recordkeeping systems used in public administration (in Norway). The specification lists requirements with regard to information content, data structure and functionality. In some cases there are requirements relating to users but this is mainly left to the individual system developers or vendors to decide. The specification does not contain requirements relating to the implementation of the data structure or how systems are to be designed[42].

Archives New Zealand

Archives New Zealand has forged its own strategy called, Continuum – Create and Maintain (currently under review) which has been designed to provide tools and services to government agencies to enable them to meet best practice record keeping standards. It assists agencies in developing their own records management programs to fulfil business and accountability requirements, and promotes good records management so that the most significant records of government are preserved for current and future generations. It is also designed to promote strong, cooperative and mutually beneficial partnerships between Archives New Zealand and government agencies[43].

In February, 2006, Archives New Zealand launched a new web site called Digital Record keeping which is designed to serve as a focal point for electronic records issues ranging from access to preservation. http://www.archives.govt.nz/continuum/digitalrecordkeeping.php

---

[41] www.sa.dk/sa/omarkiverne/english/earchives.htm

[42] http://www.riksarkivet.no/english/electronic.html

[43] http://www.archives.govt.nz/continuum

National Archives and Records Administration - Electronic Records Archives program

The multi-million dollar seven year Electronic Records Archives program of the National Archives and Records Administration (NARA) has just selected Lockheed-Martin as the company that will lead the development of a comprehensive persistent digital archives program. The project is expected to have a far ranging impact on digital archives development in other US archives and beyond. NARA is also leading other related initiatives in partnership with the San Diego Supercomputing Lab [first increment to be delivered in 2007]. Another interesting initiative is the development of RM Service Components (RMSC) that aims to help government agencies to integrate records management functionality at an early stage into business processes[44].

National Archives of the Netherlands, Digital Longevity (Digitale Duurzaamheid)

The Nationaal Archief, which can be seen as the 'national memory' of the Netherlands, manages both archives of central government and archives of private organisations or individuals deemed to be of national importance. As such it is responsible for appraising, managing and maintaining accessibility to government records through time. In connection to this, the Nationaal Archief advises government agencies on their records and archives management. The Digitale Duurzaamheid or Digital Longevity department, which is responsible for these activities, intends to be a centre of expertise for the management and preservation of digital information (records). It maintains a website with the latest relevant information on research, projects and events (www.digitaleduurzaamheid.nl). One of the main current activities is the development and building of a digital repository, that will enable the Nationaal Archief to ingest and preserve archival records through time.

Modular emulation project

Modular emulation project (2005-2007), a collaborative research project of The Dutch National Archives and the Dutch National Library (KB) on emulation[45].

DOMEA

In November 1999, the German Federal Government Co-ordination and Advisory Agency for IT in the Federal Administration (KBSt) published its concept of the Paperless Office (DOMEA concept). Since that time, it has become established as a quasi-standard for

---

[44] http://www.archives.gov/era/rmsc/rmsc-schedule.html

[45] http://www.digitaleduurzaamheid.nl/index.cfm?paginakeuze=286&categorie=6

electronic records management. Federal, Länder and local governments as well as product manufacturers have used the potential of this concept as a basic illustration of public administration's requirements for electronic records management[46].

eDavid project

City Archives of Antwerp: Expertisecentrum DAVID vzw is a centre of research and knowledge on digital archiving[47]. The research domain of eDAVID concerns:

- born digital documents, specifically those produced in business and government processes and requiring long term preservation

- digitised documents, notably photographs, audio and video material present in archives, museums and libraries, etc.

National Archives of Australia (NAA)

The **National Archives of Australia (NAA)[48]**, launched the e-permanence program in 2003 to bring a focus to its efforts to preserve both current and archival electronic records. It complements and is supported by the tools and guides that have been issued under, 'Designing and Implementing Record Keeping Systems' (DIRKS) initiative[49] which is designed to enhance the management of records under the control of Australian Commonwealth ministries. The National Archives has developed a preservation approach for archival digital records based on converting or 'normalising' digital records into archival data formats for long-term storage and access. The basic conceptual approach of the project is described in a National Archives Green Paper, "An Approach to the Preservation of Digital Records"[50]. The Archives' approach to the long-term preservation of digital records is also expected to be useful for preserving digital records in agency custody. In order to guide departments and agencies, the NAA released, "Digital Recordkeeping: Guidelines for Creating, Managing and Preserving Digital Records"[51]. This site sets out the National Archives' policy on the status and management of Australian Government websites and other online resources as Commonwealth records. It will assist agencies to establish internal

---

[46] http://www.kbst.bund.de/-,413/DOMEA-Konzept.htm

[47] http://www.expertisecentrumdavid.be/

[48] http://www.naa.gov.au/

[49] http://www.naa.gov.au/recordkeeping/dirks/summary.html

[50] www.naa.gov.au/record keeping/er/digital_preservation/summary.html

[51] http://www.naa.gov.au/recordkeeping/er/guidelines.html

mechanisms for creating, managing and retaining web-based records, for as long as those records have value[52].

Australasian Digital Recordkeeping Initiative (ADRI)

The **Australasian Digital Recordkeeping Initiative (ADRI)** is an undertaking of the Council of Australasian Archives and Records Authorities, the peak body of government archives and records institutions in Australia and New Zealand. The primary objective of ADRI is to pool the resources and expertise of national, state and territory public records institutions in Australia and New Zealand to find better ways to ensure that digital records are preserved and made accessible for the future[53].

Clever Metadata Project

The objective of the 'The Clever Metadata Project' of Monash University project is to develop a proof of concept prototype to demonstrate how standards-compliant metadata can be created once in particular application environments then used many times for multiple purposes across business applications and in different environments. The prototype will be implemented in a test-bed site to provide a model for best practice. Development of the prototype and its implementation model will also require the prototyping of model metatools, including a mini-metadata registry[54].

New York State Archives, Models for Action: Developing Practical Approaches to Electronic Records

The Models for Action project, begun in the late 1990's, seeks to find practical solutions to electronic recordkeeping problems in networked computing environments. The project incorporates principles from business process analysis, information systems development, electronic records management and archival preservation in order to find solutions that address recordkeeping at all stages of the records life cycle and, more importantly, within the context of the business process which the records support[55].

Minnesota State Archives, Trustworthy Information Systems Project

With the support of the National Historical Publications and Records Commission (NHPRC), the Information Policy Council (IPC) and in collaboration with select state agency partners

---

[52] http://www.naa.gov.au/recordkeeping/er/web_records/intro.html

[53] http://www.adri.gov.au/content.asp?cID=14

[54] http://www.sims.monash.edu.au/research/rcrg/research/crm/background.html

[55] http://www.ctg.albany.edu/publications/guides/digital_preservation_partnerships

and the Data Issues Group for Information Technology (DIG-IT), the State Archives Department of the Minnesota Historical Society developed and tested a set of criteria to establish the trustworthiness of information systems. The Trustworthy Information Systems (TIS) Project produced a Handbook, which details the TIS criteria set and a methodology to evaluate government information systems for trustworthiness[56].

State Archives of Michigan: Preserving the Electronic Records Stored in a Records Management Application (PERM)

The State Archives of Michigan and the San Diego Supercomputer Center (SDSC) wanted to develop and test a model for preserving the records in Michigan's Records Management Application (RMA) environment. While the RMA is capable of storing and providing access to electronic records, it cannot ensure that they remain accessible as software becomes obsolete. A final report describes the results of the project[57].

PrestoSpace

The PrestoSpace Integrated Project funded by the European Commission aims to provide technical solutions and integrated systems for the complete digital preservation of a variety of audio-visual collections[58]. The overall goal is to provide a range of audiovisual services for *preservation*, *restoration*, *storage, management, metadata, delivery and access* in cooperation with commercial partners.

One of the project's outcomes is the A/V Archive Digitisation and Storage Guide website[59], providing information and management tools on digital technology for the storage of film, video and audio content and associated metadata.

PLANETS

The Planets project brings together European National Libraries and Archives, leading research institutions, and technology companies to address the challenge of preserving access to digital cultural and scientific knowledge[60].

Co-Financed by the European Commission, Planets aims out to develop services for *preservation planning*, enabling organisations to define, evaluate, and execute preservation

---

[56] http://www.mnhs.org/preserve/records/tis/tis.html

[57] http://www.sdsc.edu/PERM/

[58] http://www.prestospace.org

[59] http://prestospace-sam.ssl.co.uk/

[60] http://www.planets-project.eu

plans. It develops methodologies, tools and services for the *characterisation of digital objects* and for the support of *preservation actions*. It further aims to establish a *preservation testbed* to provide a consistent and coherent evidence base for the objective evaluation of different preservation protocols, tools and services and for the validation of preservations plans. The implementation of an interoperability framework seamlessly integrates the sub-projects within a distributed service network.

CASPAR

The CASPAR project (Cultural, Artistic, and Scientific Knowledge for Preservation, Access and Retrieval)[61] is co-financed by the European Commission. It aims to implement, extend, and validate the OAIS reference model, to enhance techniques for capturing representation information and other preservation-related information for content objects, and to design virtualisation services for abstracting from underlying computing and storage environment.

Furthermore, CASPAR integrates services for digital rights management, authentication, and accreditation, conducts research into more sophisticated access to and use of preserved digital resources such as intuitive query and browsing mechanisms, and develops case studies for validating its approach across different user communities.

nestor II

The application for nestor II, a joint project of the Deutsche Nationalbibliothek, SUB Göttingen, BSB München, HUB Berlin, Bundesarchiv, Institut für Museumskunde and FernUniversität in Hagen, has recently been approved by the German Federal Ministry of Education and Research[62]. A key content-related priority for the project will be to interlink the fields of e-Science, grid-computing and the long-term preservation of digital resources.

A major task in the second phase of the project is to create the necessary organisational, political and financial conditions for the competence network as a means of anchoring the entire subject area and associated activities on a permanent basis in Germany. This task applies equally to the preservation of the written (carrier-bound) cultural heritage in Germany, and accordingly a joint initiative between nestor and the "Allianz zur Erhaltung des schriftlichen Kulturgutes" (Alliance for the Preservation of Written Cultural Heritage) is also being prepared.

---

[61] http://www.casparpreserves.eu

[62] http://www.langzeitarchivierung.de/index.php?newlang=eng

kopal

kopal – Cooperative Development of a Long-Term Digital Information Archive[63] is a national project sponsored by the German Federal Ministry of Education Research. It aims to develop a technological and organisational solution to ensure the long-term availability of electronic publications. Thereby, the transparent integration into existing library systems and the reusability through memory institutions play a critical role.

In the implementation of the system, international standards for long-term archiving and metadata will be adopted. In this way, both sustainability and the ability to further develop the system are guaranteed. Operational prototypes deployed at the Deutsche Nationalbibliothek and the SUB Göttingen are currently under evaluation.

DELOS Preservation Cluster

The preservation cluster of the DELOS Network of Excellence in Digital Libraries[64] provide the methodological framework and theory for ensuring that digital libraries research addresses preservation issues and digital libraries incorporate preservation elements in their designs.

The major objectives of the Preservation Cluster[65] are to catalyse the research and funding environment to enable of delivery of the DELOS/NSF research agenda for Digital Preservation and Archiving, to lay the foundation for testbeds and necessary metrics and tools for assessing preservation strategies, to raise the profile of digital preservation issues within the Digital Library Community, to collaborate with other international bodies to ensure consistencies of digital repository standards, to ensure access to file format information, to establish a relationship between a typology of file formats and preservation strategies, to enable the definition of attributes and functionalities that need to be represented, and ensure that system development methodologies reflect preservation analysis and design issues.

Other

The "Digital Repository Infrastructure Vision for European Research" (DRIVER) project[66] will put a test-bed in place across Europe to assist the development of a knowledge infrastructure for the European Research Area.

---

[63] http://kopal.langzeitarchivierung.de/index.php.en

[64] http://www.delos.info/

[65] http://www.dpc.delos.info/

[66] http://www.driver-repository.eu/

The BRICKS Project – Building Resources for Integrated Cultural Knowledge Services – researches and implements advanced open source software solutions for the sharing and the exploitation of digital cultural resources[67].

EPOCH is a network of around a hundred European cultural institutions joining their efforts to improve the quality and effectiveness of the use of Information and Communication Technology for Cultural Heritage[68].

The ERPANET project is covered in Appendix B.10.

---

[67] http://www.brickscommunity.org/

[68] http://www.epoch-net.org

A.3   Standardisation

ISO

The ISO records management standard (ISO 15489:2001), offers a model framework for the management of records in all forms including digital[69]. Recently a standard for the metadata of records has been published (ISO-23081-1:2006) that explains the principles for records management metadata[70]. Currently, work is underway in TC46/SC11 to address functional requirements for long term preservation of digital records and for work process analysis in support of records management.

In late 2005 a subset of PDF was transformed into an international ISO standard PDF/A (ISO 19005-1:2005)[71]. These, and standards such as those developed in individual sectors (e.g. the international pharmaceutical industry) or for distinct media (e.g. MPG3, JPEG, etc.) are being built upon the growing suite of standards responding to the global pressure for greater inter-connectivity across both space and time.

An example of the work of ISO can be found in Joint Technical Committee 1/SC32/WG2 which is currently responsible for developing and maintaining standards that facilitate specifications and management of metadata. The goal is to enhance the understanding and sharing of data, information and processes to support interoperability, electronic commerce and component-based development[72]. Organisation for the Advancement of Structured Information Standards (OASIS) is a not-for-profit, international consortium that drives the development, convergence, and adoption of e-business standards. The consortium produces more Web services standards than any other organisation along with standards for security, e-business, and standardisation efforts in the public sector and for application-specific markets. Founded in 1993, OASIS has more than 5,000 participants representing over 600 organisations and individual members in 100 countries[73].

---

[69]http://www.iso.ch/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=31908&ICS1=1&ICS2=140&ICS3=20

[70]http://www.iso.ch/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=40832&ICS1=1&ICS2=140&ICS3=20

[71]http://www.iso.ch/iso/en/CombinedQueryResult.CombinedQueryResult?queryString=PDF%2FA

[72] http://metadata-standards.org/metadata-stds/

[73] http://www.oasis-open.org/home/index.php

Dublin Core

The Dublin Core Metadata Initiative (DCMI) is an organisation dedicated to promoting the widespread adoption of interoperable metadata standards and developing specialised metadata vocabularies for describing resources that enable more intelligent information discovery systems. The Dublin Core Metadata Initiative provides simple standards to facilitate the finding, sharing and management of information[74].

The DCMI abstract model provides a reference model against which particular Dublin Core encoding guidelines can be compared[75]. It describes *resources*, *descriptions*, *properties*/*values* and their corresponding relations.

Dublin Core application profiles specify which metadata terms an organisation, information provider, or user community uses in its metadata. It identifies the terms used to describe a resource and optionally provides additional information about term usage. Existing application profiles, for example, include collections[76] and e-prints[77].

Semantic Web

The Simple Knowledge Organisation System (SKOS) [78] provides a model for expressing the basic structure and content of *concept schemes*. A concept scheme is defined as a set of concepts, optionally including statements about semantic relationships between those concepts. Thesauri, classification schemes, taxonomies and others types of controlled vocabulary are examples of concept schemes, which might also be embedded in glossaries and terminologies. The SKOS core vocabulary is represented by a set of RDF[79] properties and RDF(S) [80] classes.

RDFa is a syntax that uses a set of elements and attributes to embed RDF metadata expressions in XHTML[81], strictly separating between content and metadata[82]. Applications scenarios for RDFa include direct import of events from webpages into the user's calendar or the automatic detection of licensing information associated to documents.

---

[74] http://dublincore.org

[75] http://dublincore.org/documents/abstract-model/

[76] http://dublincore.org/groups/collections/collection-application-profile/

[77] http://www.ukoln.ac.uk/repositories/digirep/index/Eprints_Application_Profile

[78] http://www.w3.org/TR/swbp-skos-core-guide/

[79] http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/

[80] http://www.w3.org/TR/2004/REC-rdf-schema-20040210/

[81] http://www.w3.org/TR/xhtml2

[82] http://www.w3.org/TR/xhtml-rdfa-primer/

In 2006, the W3C also released "Best Practice Recipes for Publishing RDF Vocabularies"[83], providing instructions to creators and maintainers of RDF(S) and OWL[84] vocabularies or ontologies on how to publish them on a web server.

PREMIS

PREMIS (PREservation Metadata: Implementation Strategies) is a metadata model developed by an international group of metadata experts. The "Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group"[85] was released in May 2005. It is a comprehensive guideline for implementing preservation metadata in digital archiving systems and defines a set of core metadata for preservation. A maintenance activity has been set up to manage the further development of the standard[86].

LMER

"Long-term preservation Metadata for Electronic Resources" (LMER)[87] is a metadata model which was developed at the German National Library and which is based on the "Preservation Metadata: Metadata Implementation Schema" of the National Library of New Zealand. It is implemented in the kopal solution of the IBM DIAS-Core. LMER was designed as an exchange format, not as a general data model for long-term preservation metadata. It is based on a modular design.

UOF

The "Universal Object Format" (UOF)[88] is based on LMER and the Metadata Encoding and Transmission Standard (METS). It is again a development of kopal. It is a format for the archiving and exchange of digital objects along with metadata.

Persistent Identifiers

The persistent identification of digital information plays a vital role in enabling its long-term accessibility and re-use[89]. A unique persistent identifier preserves access to a resource regardless of its location via corresponding resolver services.

---

[83] http://www.w3.org/TR/swbp-vocab-pub/

[84] http://www.w3.org/TR/owl-ref/

[85] http://www.oclc.org/research/projects/pmwg/premis-final.pdf

[86] http://www.loc.gov/standards/premis/index.html

[87] http://www.ddb.de/standards/pdf/lmer12_e.pdf

[88] http://kopal.langzeitarchivierung.de/downloads/kopal_Universal_Object_Format.pdf

[89] http://www.dcc.ac.uk/resource/briefing-papers/persistent-identifiers/

The **Uniform Resource Name** (URN) is a standard, persistent and unique identifier for digital resources on the internet. URNs include a namespace identifier code (NID) and a namespace specific string (NSS). The NID is used to interpret the NSS, which is the local name identifying the individual document[90]. The EPICUR project[91] presents a framework for cooperative URN management at the German national library (cf. [44]), covering URN assignment and administration, resolving services, and standardization within bibliographic information services building upon experiences of CARMEN-AP4 (described in [45]).

The **Digital Object Identifier** (DOI) system was initiated by the Association of American Publishers. The DOI system allows the allocation of a unique digital identifier to commercial digital publications**Error! Bookmark not defined.**[92]. DOI has been investigated for the DFG funded project CODATA (see [46]) and is preferred in the project because of CrossRef[93], existing centrally managed global services, and because DOI is already internationally established for a longer time than URN.

Persistent identifiers have also been discussed in the following conference and workshops:

- The Digital Library and its Services[94]

- Persistent Identifiers (Erpatraining)[95]

- DCC Workshop on Persistent Identifiers[96]

- EPICUR-Workshop "Persistent Identifier"[97]

- Nestor Workshop - Standardisierungsbedarf im Kontext von Langzeitarchivierung und Digitalisierung[98]

Multimedia

---

[90] http://www.nla.gov.au/padi/topics/36.html

[91] http://www.persistent-identifier.de/

[92] http://www.doi.org/handbook_2000/index.html

[93] http://www.crossref.org/

[94] http://www.ukoln.ac.uk/events/bl-jisc-conference-2006/

[95] http://www.erpanet.org/events/2004/cork/index.php

[96] http://www.dcc.ac.uk/piworkshop.html

[97] http://www.persistent-identifier.de/?link=340

[98]http://nestor.sub.uni-goettingen.de/aktuell/suche.php?suchwort=din&order=date+DESC&zahl=10&start=0

Broadcasting archives contain an inventory of selected audiovisual recordings, held for preservation and as a resource for broadcasting and production purposes (cf. [39]). The **Broadcast Metadata Exchange Format** (BMF)[99] supports the full broadcast and production lifecycle, serving as exchange format between components of a file-based production platform. Its class model specifies data structures, which correspond to BMF conceptual model.

The following definitions have been adopted from [40]: The **Unique Material Identifier** (UMID) acts as a lookup ID number for metadata, which does not have to be kept within the data stream. The **MXF** (Material eXchange) file and stream format combines data and metadata within a single file, which might also be streamed. The **Advanced Authoring Format** (AAF) describes work in progress, allowing interlinking data streams with metadata. The metadata representation at each production step might vary, depending on the information to be captured and described.

---

[99] http://www.irt.de/IRT/home/indexbmf.htm

A.4 Links to Other Disciplines

Computer Science and Engineering

Due to its versatile nature digital preservation has links to several disciplines within computer science. The areas showing a strong potential for digital preservation that we identified are:

Information Extraction deals with the analysis of digital objects on a semantic level. A given object's structural elements have to be defined. Based on the object's subparts, important information is extracted and used for more sophisticated analysis. Of course, which parts are to be extracted strongly depends on the type of object. For text documents techniques like part-of-speech tagging or thesauri (or more elaborate ontologies) can be used to extract concepts. Techniques developed in information extraction are very promising in the area of object characterisation. Especially for the use of dynamically growing archives such as internet archives, which contain a-priori unknown content, the process of extracting text and acquiring metadata from documents is a key issue. Automatic metadata acquisition will be very relevant to digital preservation. Research should be directed towards automatic extraction of domain-dependent information from documents. The extraction of semantic concepts should also be a priority.

Information Retrieval deals with searching for documents, their metadata, and information within those documents. A typical IR task is to find similar documents to one example document. It can be used for access and retrieval of digital objects. Information Retrieval shows great potential for searching in digital libraries in general and possibly identifying similar objects to for adopting preservation strategies in digital preservation. Every digital repository, and all systems designed for digital preservation, need efficient ways to access their contents. This is particularly important once those repositories reach a certain size, for example the difference between an internet search engine and a search engine for Internet archives.

The complexity of the digital preservation domain poses new challenges for the Information Retrieval community. Firstly, the huge variety of formats and metadata models extend the use of classic information retrieval. Secondly, the presence of obsolete file formats represents another challenge. Specifically, files which can be accessed in a digital library, but cannot be indexed by its retrieval component will have to be dealt with. Important tasks will be: to scale information retrieval techniques to Internet archives, to establish how duplicate detection can work over time, and to establish is classic information retrieval techniques can be adapted to the more complex search challenges of the future.

In an archive containing enormous amounts of data, where parts of the objects are at risk, Information visualisation together with machine learning could greatly help in providing overview of high-risk areas and judging which areas are affected. Information Retrieval which allows one to find objects with similar characteristics will be very important where one obsolete object is found and there is a need to identify which other objects might be at risk.

Information visualisation is used to display complex data and their relations and provide an abstract view on that data. Note that visualisation differs from display as it is not a means of rendering an object, but of showing how objects relate to each other. When applied to digital preservation, information visualisation could greatly improve the quality of access and retrieval in digital libraries. Explorative data analysis through visualisation is often of great aid to discovery. Accessing digital objects is an important indicator of the quality of a certain preservation strategy. Important research questions will establish: if appropriate access be guaranteed, whether new formats can be easily displayed over a variety of information channels, and whether different instances of the same object and their relations be visualised. Visualisation of migrated objects and their differences to the original object would greatly support the evaluation of preservation strategies, however this is not yet developed.

Machine learning tries to automatically find structures and natural clusters, for example similar instances within data, by developing algorithms and techniques to enable computers to abstract and thus 'learn'. A possible application of machine learning techniques would be in the area of automatic metadata extraction and assignment. Classic text or document categorisation may prove very important when dealing with larger numbers of elements. The classification or categorisation community can offer many tools and much expertise, but still many specific issues and challenges arising in the context of large-scale repositories and digital preservation are yet unsolved. In the context of digital preservation, important research areas will be finding feature sets for complex objects, as well as automatic metadata assignment and its varieties. If an object X is identified as 'risky', an important task will be to a) identify similar objects and b) adopt the preservation/migration strategy of X to fit its nearest neighbours.

eScience/Grid-Computing - Grid computing deals with the use of large amounts of data and huge processing power distributed across a large number of elements in the grid. This field may become very important once digital preservation tools and techniques reach a certain level of maturity and become used on a large-scale basis. There is also a strong correlation with cyberinfrastructure.

Cyberinfrastructure denotes the building of a new or better utilisation of existing infrastructure concentrating on new applications which interoperate across institutions as well as disciplines. Data and knowledge exchange are key issues here, illustrating how the cyberinfrastructure challenge tries to integrate several other research areas like information extraction and exchange or semantic interoperability.

Knowledge engineering offers promising ways to store and access data as well as approaches to collaboration and data exchange that can be beneficial for digital preservation. Interoperability and exchange of knowledge between disciplines and organisations is the logical next step to achieve a more comprehensive view of knowledge and data management. Results from the underlying information extraction elements are essential. Ontologies are commonly used to describe information that is about to be exchanged, and will prove important as soon as data exchange issues in digital preservation arise on a large scale.

Decision Support Systems - A range of tools exists today to support the variety of preservation strategies such as migration or emulation. Yet, different preservation requirements across institutions and settings make the decision on which solution to implement very difficult. The application of decision support tools to aid in preservation planning thus forms one of the core current research areas. Finding candidates among existing planning strategies is the main aspect of preservation planning that might be improved. Moreover, case based techniques may be feasible for identifying similar strategies.

Semantic Technologies include software standards and methodologies that aimed to provide more explicit meaning. Semantic technologies are developed based on the research results from domains such as cognition, networking, content engineering and web services. Digital preservation benefits from ongoing standardisation activities, resulting in open, documented formats, protocols and representation schemes. However semantics-enabled repositories might also offer improved recall or precision for concept-based document access. By capturing information about context and content in machine processable form, semantic technologies add value to digital preservation[100].

Cognition

Following the projections for computer science presented in [25], results and insights from cognition and related disciplines might also be explored to yield new directions and prospects for research in digital preservation.

---

[100] http://www.archives.gov/era/pdf/it-conference-ludaescher.pdf

Access - Units within a neural net are activated by receiving input patterns and pass output to other connected units. If a neural unit corresponds to attributes of data, this data can be accessed by each connected unit. Hence, data cannot only be recalled by the indexing system, but by any of the attributes [25]. Such a content-addressable memory allows for flexible, robust recall and retrieval.

Memory - To meet future challenges for the construction and indexing of large-scale multimedia memories, the mechanisms employed in the Human memory are directing more and more attention, especially the relationship between working memory and long-term memory [25]. Which particular storage or recall mechanisms are employed in a given context is defined by the given task at hand and the type of memory and the conceptualization of the environment.

In Human psychology, forgetting is not just a failure of memory, but an essential mechanism for clearing out "irrelevant" facts and preserving the efficiency of recall [25]. Problems here are the identification of what to forget, how to deal with related knowledge and the actual implementation of "forgetting".

Summary of Key Research Agendas

A.1 Understanding Electronic Incunabula: A Framework for Research on Electronic Records

The article "Understanding Electronic Incunabula: A Framework for Research on Electronic Records" [9] by Margret Hedstrom was published in 1991. The research for this report was partly supported through the research Fellowship Programme in Modern Archives at the Bentley Historical Library and funded by the National Endowment for the Humanities and the Mellon Foundation.

It addresses the archivists' need for a research framework on electronic records within a historical, social, and cultural context.

Research in digital preservation is approached from the perspective of changes in the archival profession, ignoring other domains. It describes five research objectives, claiming context to be the essential concept to be researched. New forms of material, intervention at critical points regarding the development and introduction of new technologies, changes to archival practice and requirements for management and preservation represent the other issues raised, which are mostly related to the organisational environment.

A.2 Preserving Digital Information

The report "Preserving Digital Information" [17] of the Task Force on Archiving of Digital Information co-chaired by John Garrett and Donald Waters was commissioned by The Commission on Preservation and Access and the Research Libraries Group.

The purpose of the task force was to investigate the means of ensuring "continued access indefinitely into the future of records stored in digital electronic form". It was charged to frame the key problems (organisational, technological, legal, economic) in the context of media deterioration and technological obsolescence, to define the critical issues inhibiting resolutions, recommend actions and the consideration of alternatives to technology refreshing.

The report concludes with a set of recommendations to the funding institutions for specific actions, including the provision of coordination services, securing of funding, fostering practical experiments, engagement in national policy efforts, development of effective fail-safe mechanisms, interdisciplinary discussions, certification of repositories, international collaboration and the commissioning of follow-up studies.

A.3 An Investigation into the Digital Preservation Needs of Universities and Research Funders

The study "An Investigation into the Digital Preservation needs of Universities and Research Funders" [15] by Denise Lievesley and Simon Jones was part of a series of preservation studies which were managed by the British Library Research and Innovation Centre (BLIRC) and published in 1998. It was commissioned by the Digital Archival Working Group and carried out by the UK Data Archive.

Motivated by the need to develop strategies for the selection and preservation of digital material, its purpose was to explore the extent, by which preservation policies for electronic materials are implemented by universities and other institutions with public funding.

The research was carried out by postal questionnaires, telephone interviews or personal visits with representatives responsible for preservation policies from Universities and funding agencies. The themes explored covered the supported types of electronic material, their preservation, policies and practices, selection, legal issues and roles and responsibilities. Its main recommendations are the development of national guidelines for digital preservation, standards for the assessment of repositories and national policies on research data.

A.4 The State of Digital Preservation - An International Perspective

The conference proceedings "The State of Digital Preservation: An International Perspective" does not strictly represent a research agenda, but a collection of individual articles covering various aspects of digital preservation research. The list of authors comprises Meg Bellinger, Laura Campbell, Margret Hedstrom, Deanna Marcum, Kenneth Thibodeau, Donald Waters, Titia van der Werf and Colin Webb. The corresponding symposium, which was supported by a grant from Documentation Abstracts Inc. (DAI) and presented by the Council on Library and Information Resources, was held in late April, 2002 and its proceedings were published in July 2002.

Six years after [17], the symposium aimed to assess the current state of digital preservation, identifying the barriers impeding progress and figure out ways to overcome them by looking at the most interesting developments in recent years and an analysis of the challenges ahead.

A.5 It's About Time: Research Challenges in Digital Archiving and Long-term Preservation

"It's About Time: Research Challenges in Digital Archiving and Long-term Preservation" [3] is the final report of the workshop, which was organised by Margret Hedstrom, Sharon

Dawes, Carl Fleischhauer, James Gray, Clifford Lynch, Victor McCrary, Reagan Moore, Kenneth Thibodeau and Donald Waters and took place in April 2002. It was sponsored by the National Science Foundation Digital Government Program and Digital Libraries Program and the Library of Congress National Digital Information Infrastructure and Preservation Program and published in August 2003.

The objective of the workshop was to examine the prospects for advancing computer and information technology research through a research programme that addresses the unique challenges of long-term digital preservation.

The resulting digital archiving and long-term preservation research agenda recommends a substantial research program to enhance the state of knowledge and practice for long-term preservation of digital information. The agenda is organised around for main themes: i) technical infrastructures for archival repositories, ii) attributes of archival collections, iii) digital archiving tools and technologies, and iv) organisational, economic and policy issues. Research scenarios describing the types of projects required for the implementation of the research agenda in a 10 year time frame suggest theory-building, exploratory research, simulations and experiments, observations and testbeds.

A.6 Invest To Save: Report and Recommendations of the NSF-DELOS Working Group on Digital Archiving and Preservation

The report "Invest To Save: Report and Recommendations of the NSF-DELOS Working Group on Digital Archiving and Preservation" [4] was prepared for the NSF Digital Library Initiative and the DELOS Network of Excellence for Digital Libraries and published in 2003,. The transcontinental working group was chaired by Seamus Ross and Margret Hedstrom, other participants were Kevin Ashley, Birte Christensen-Dalsgaard, Wendy Duff, Henry Gladney, Claude Huc, Anne R. Kenney, Reagan Moore and Erich Neuhold.

The objective was to identify critical problems and issues to be solved for improving the society's ability to preserve valuable digital information for future reuse. The report sets research priorities and discusses corresponding strategies and methods. It furthermore identified potential synergies for collaborative research between European and North American researchers.

The research agenda developed by the working group is structured into three main areas: i) preservation strategies: emerging research domains, ii) re-engineering preservation processes and iii) preservation systems and technology. Additional issues which were deliberately not part of the research agenda comprise policy, organisational, educational and other activities.

The working group agreed self-contextualising objects, metadata and the evolution of ontologies and mechanisms for preservation of complex and dynamic objects to be the research areas likely to have the greatest impact in the future.

A.7 e-Science Curation Report

The "e-Science Curation Report" [16] by Philip Lord and Alison McDonald of The Digital Archiving Consultancy Limited was funded by The JISC Committee for the Support of Research (JCSR) and published in 2003.

The goal was to examine the current provision and future needs of primary research data in the UK, especially within the e-Science context.

This was done by carrying out desk-based research into literature, the administration of a series of questionnaires researchers, IT service providers and policy makers and by conducting face-to-face interviews with experts and researchers. Their findings at strategic and policy level resulted in a list of 10 strategic recommendations and corresponding responsibility assignments and 19 tactical recommendations.

A.8 Revolutionizing Science and Engineering Through Cyberinfrastructure

"Revolutionizing Science and Engineering Through Cyberinfrastructure" [18] is a report of the NSF Blue-Ribbon Advisory Panel on Cyberinfrastructure which was chaired by Daniel E. Atkins and published in January 2003. Further members of the board were Kelvin K. Droegemeier, Stuart I. Feldman, Hectore Garcia-Molina, Michael L. Klein, David G. Messerschmitt, Paul Messina, Jeremiah P. Ostriker and Margaret H. Wright.

Cyberinfrastructure (hardware, software, services, personnel, organisations) refers to infrastructure based upon distributed computer, information and communication technology in the knowledge economy. The Blue Ribbon Panel was convened to inventory and explore opportunities for creating new research environments based upon cyberinfrastructure. The objective was to make strategic recommendations on the nature and forms of programs that NSF should take in response.

While the report is not specifically on digital preservation, cyberinfrastructure enables one to tailor efficient and effective application-specific, but interoperable knowledge environments for research and education. In this context, it has to be ensured "that data and software acquired at great expense are preserved for future generations and easily available to all". Hence, the NSF is advised to institute a program for the collection of exponentially growing data, its curation, management and archival for long-term access.

A.9   Digicult Roadmap

The roadmap report "The Future Digital Heritage Space – An Expedition Report" [6] was edited by Guntram Geser and John Pereira and published as DigiCULT Thematic Issue 7 in 2004. The DigiCULT forum project was funded within the IST priority of the European Union's Fifth Framework Programme for Research and Technological Development.

Intended as a navigation tool for boards and directors of heritage organisations and research centres, IT project managers and curators of digital collections its objective was to project an overview information and communication technologies, systems and applications that may be achieved over the next decade.

An online consultation forum was established around 6 themes, with feedback and contributions gathered from 62 experts and professionals. The final report is structured around these themes, providing individual roadmaps for intelligent heritage, contextual cultural information, natural and enjoyable interaction, create/re-create 3D/VR/AR, large scale and distributed information systems and persistent and perpetual access.

A.10 Electronic Resource Preservation and Access Network

The European Commission funded ERPANET project[101] established a consortium for making viable and visible information, best practice and skills development in the area of digital preservation of cultural heritage and scientific objects. ERPANET ran from 2001 – 2004 and had The Humanities Advanced Technology and Information Institute (HATII), Nationaal Archief van Nederland, Istituto di Studi per la Tutela dei Beni Archivistici e Librari and Schweizerisches Bundesarchiv as project partners.

While not a research agenda, ERPANET's main objective was to exchange knowledge on state-of-the-art developments in digital preservation and the transfer of expertise among individuals and institutions, making tools, knowledge and experience accessible to the community.

Throughout its duration, ERPANET released a suite of guidance tools on ingest strategies, costing orientation, selecting technologies, digital preservation policy and risk management. The erpaCharter comprises principles to enable better risk management, a framework for benchmarking digital access and preservation practice, a platform for sharing knowledge, a mechanism for measuring competitive advantage, a key to unlock the value of intellectual capital, synergy and avoid redundancy, and communication and collaboration between the

---

[101] http://www.erpanet.org

different players in the Digital Object Chain. Additionally, the ERPANET website maintains directories for assessments, e-prints and information resources.

A.11 Digital Curation and Preservation: Defining the research agenda for the next decade

"Digital Curation and Preservation: Defining the research agenda for the next decade" [1] is the report of the Warwick Workshop, which took place in November 2005. It was sponsored by the Joint Information Systems Committee(JISC), the Council for the Central Laboratory of the Research Councils (CCLRCs), the Digital Curation Coalition (DCC) and the British Library and coordinated by David Giaretta and Heather Weaver..

The objective was to map out the current state of play and future agenda, complementing the Task Force on Permanent Access to the Records of Science's research agenda, extending and clarifying a number of research issues.

The workshop focused three main strands, which were discussed in sessions: i) curation services and technologies, ii) drivers and barriers (policy issues), and iii) data life cycle management (process issues). The task for each session was to define scope and topic, analyse the current national and international state of play, develop a vision for the next 5 to 10 years, and specify the means to achieve the vision. Common research issues identified across the three sessions are discovery and location, trust, cost, and automation and virtualisation.

A.12 Digital Repositories Roadmap - Looking Forward

The report "Digital Repositories Roadmap: looking forward" [29] by Rachel Heery and Andy Powell was funded by the JISC as part of the Digital Repositories Programme and published in 2006.

Focusing on digital repositories for research outputs and learning material in the UK and their role in the information landscape, the purpose of the roadmap was to assess the current situation and specify a vision for 2010, including the route to getting there. Its principal audience comprises the JISC Executive, the Repositories, Preservation and Asset Management Advisory Group and relevant JISC committees.

The report assumes that in 2010, open access is established, with a high percentage of scholarly output made freely available for sharing and re-use. Furthermore, geospatial information will be better integrated with other data. The report was accompanied by an email questionnaire. The authors suggest focusing research activities in the areas of policy

and legal issues, culture and working practices, in addition to developments of the technical infrastructure in the UK.

A.13 E-Infrastructure Strategy for Research: Final Report from the OSI Preservation and Curation Working Group

The "E-Infrastructure Strategy for Research" [43] final report commissioned by the British Office for Science and Innovation (OSI) steering group on e-infrastructure research was written by Neil Beagrie and published in January 2007.

As part of the UK "Science and Innovation Investment Framework 2004-014", six working groups have been formed to establish a high-level roadmap on e-infrastructure. The task of one working group was to produce a report on preservation and curation.

Within a 10 year horizon (2004-2014), the report assumes that the transition to a digital knowledge economy will largely have been completed. Focussing primarily on the national public sector, particularly government and academic research, its key recommendations are in the areas of i) policy development, ii) digital preservation and curation research and development programme, and iii) persistent national information infrastructure development programme.

Project information

Project acronym:	**DPE**
Project full title:	**DigitalPreservationEurope**
Proposal/Contract no.:	**IST-2006-034762**

Project Officer: Manuela Speiser

Address:	INFSO-E3 Information Society and Media Directorate General Content - Learning and Cultural Heritage Postal mail: Bâtiment Jean Monnet (EUFO 1167) Rue Alcide De Gasperi / L-2920 Luxembourg Office address: EUROFORUM Building - EUFO 1167 10, rue Robert Stumper / L-2557 Gasperich / Luxembourg
Phone:	+352 4301 33632
Fax:	+352 4301 33190
Mobile:	
E-mail:	manuela.speiser@ec.europa.eu

Project Co-ordinator: Prof Seamus Ross

Address:	HATII, University of Glasgow
Phone:	+44 141 330 3635
Fax:	+ 44 141 330 2793
Mobile:	+44 797 435 7006
E-mail:	s.ross@hatii.arts.gla.ac.uk