

Project no. 034762

DigitalPreservationEurope (DPE)

Instrument: Coordination Action

Thematic Priority: IST-2005-2.5.10

Access to and preservation of cultural and scientific resources

**Report on Workshop on Practices and Challenges in Preservation
and Access for Scientific and Scholarly Digital Repositories**

Start Date of Project: 01 April 2006

Duration: 36 Months

Name of Organisation of Lead Contractor for this Deliverable:

Statsbiblioteket

Names of Partners Engaged in this Deliverable:

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)

Dissemination Level: **PU (Public)**

Legal Notices



The DPE Report on the Legal Framework on Repository Infrastructure Impacting on Cooperation across Member States is licensed under a Creative Commons Attribution - Non-Commercial - Share-Alike 2.0 License.

© in the collective work - DigitalPreservationEurope (which in the context of these notices shall mean one or more of the consortium consisting of HATII at the University of Glasgow, FernUniversität 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, and the staff and agents of these parties involved in the work of DigitalPreservationEurope), 2006.

© in the individual reports – the organisations as indicated in catalogue entry below unless otherwise stated.

DigitalPreservationEurope (DPE) confirms that the owners of copyright in the individual instalments have given permission for their work to be licensed under the Creative Commons license.

Catalogue Entry

Title	Report on Workshop on Practices and Challenges in Preservation and Access for Scientific and Scholarly Digital Repositories
Creator	Statsbiblioteket
Subject	Information Technology; Science; Technology--Philosophy; Computer Science; Digital Preservation; Digital Records; Science and the Humanities.
Description	This is a report on the workshop held in Berlin in November 2007
Publisher	HATII at the University of Glasgow, includes FernUniversität 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	SB
Contributor	
Contributor	
Contributor	
Contributor	
Date	
Type	Text
Format	Adobe Portable Document Format v.1.3 generated from Microsoft Word original
Resource Identifier	ISSN
Resource Identifier	ISBN
Resource Identifier	DOI
Language	English

Title	Report on Workshop on Practices and Challenges in Preservation and Access for Scientific and Scholarly Digital Repositories
Rights	© DigitalPreservationEurope Partners

Citation Guidelines

DigitalPreservationEurope, (June 2006), "J", Retrieved 07.02.2008 from <http://www.digitalpreservationeurope.eu/publications/reports/>

Document Version Control

Version	Date	Change Made (and if appropriate reason for change)	Initials of Commentator(s) or Author(s)
1.0	5/2/2008	Initial Release	CSR

Document Quality Control

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

Document Change Commentator or Author and Reviewer List

Author Initials	Name of Author/Reviewer/Responsible QA Person	Institution
CSR	Colin Rosenthal	Statsbiblioteket
EW	Emily Witham	HATII

Approved By with DPE (signature) (SRoss)	Date

Accepted by at European Commission (signature)	Date



1. EXECUTIVE SUMMARY

The Workshop on “Practices and Challenges in Preservation and Access for Scientific and Scholarly Digital Repositories” was held at the Humboldt University of Berlin on the 27th and 28th of November, 2007. 60 participants from 15 countries participated in sessions covering the major issues in repository organisation in Europe and beyond. Sessions covered Federated Repository Infrastructure, Selection Appraisal and Ingest, Operational Activity, Preservation Activity, and Quality Control and Self-audit. In addition, a number of participants remained in Berlin for a DPE/DCC Forum on Developing an International Framework for Audit and Participation the following day.



Table of Contents

1. EXECUTIVE SUMMARY	7
2. INTRODUCTION.....	9
3. THE WORKSHOP SESSIONS	10
3.1. SESSION 1: FEDERATED REPOSITORY INFRASTRUCTURES AND SOCIAL SUPPORT FRAMEWORKS	10
3.2. SESSION 2: SELECTION, APPRAISAL, INGEST AND DESCRIPTION	10
3.3. SESSION 3: OPERATIONAL ACTIVITY FOR DISCOVERY, ACCESS AND PRESERVATION	10
3.4. SESSION 4: PRESERVATION ACTIVITY	11
3.5. SESSION 5: QUALITY CONTROL, CERTIFICATION AND SELF-AUDIT.....	11
4. CONCLUSION	12



2. INTRODUCTION

The Workshop on “Practices and Challenges in Preservation and Access for Scientific and Scholarly Digital Repositories” was held at the Humboldt University of Berlin on the 27th and 28th of November, 2007. 60 participants from 15 countries participated in sessions covering the major issues in repository organisation in Europe and beyond. Sessions covered Federated Repository Infrastructure, Selection Appraisal and Ingest, Operational Activity, Preservation Activity, and Quality Control and Self-audit. In addition, a number of participants remained in Berlin for a DPE/DCC Forum on Developing an International Framework for Audit and Participation the following day. In keeping with DPE's coordinating role, the workshop was organised jointly with DRIVER, Nestor, and the DCC.

This document is a summary of the major points raised and discussed in the Workshop.

3. THE WORKSHOP SESSIONS

3.1. SESSION 1: FEDERATED REPOSITORY INFRASTRUCTURES AND SOCIAL SUPPORT FRAMEWORKS

This session explored the ways in which repositories can work together to produce a whole greater than the sum of its parts. In the first part of the session, dealing with technical aspects, speakers introduced the DRIVER and CASPAR projects. The DRIVER project focuses on developing and implementing an architecture to allow repositories to collaborate in discovery and access. The DRIVER open architecture also allows interesting possibilities for future collaborators to develop services offering elements of preservation – for example validation, replication, or migration services. Amongst DRIVER's outputs is a set of guidelines for exposing textual resources via OAI-PMH. The adoption of such a standard would have many benefits in enabling the development of new federated services for discovery, access and preservation.

The CASPAR project is more directly focused on digital preservation, and specifically on developing an architectural model for the OAIS. CASPAR lays special emphasis on the domain-specific information embodied in representation metadata.

The second part of the session dealt with social networks. A key theme which emerged is the institutionalisation/professionalisation of repository work. In the current situation, the people with actual responsibility for repository management have a wide range of backgrounds, job-titles, and job-descriptions. A future focus for the repository community will be to standardise the functions of repository professionals so as to enable proper accreditation and career paths in the repository field.

A further theme to emerge was the importance of further study of the needs and practices of actual repositories. The term “repository” represents a very diverse range of organisations managing digital material, not all of which belong to familiar categories (such as “library” or “institutional repository”) or implement standard architectural models such as OAIS. The digital preservation community needs to define how to position itself with respect to this broad category of repositories and the first step to this will be better analysis and categorisation of existing repositories.

3.2. SESSION 2: SELECTION, APPRAISAL, INGEST AND DESCRIPTION

The process of preservation and conservation of digital material begins with the decision of which material to conserve and the process of ingesting it into a repository. This session dealt with aspects of these problems. In general, all repositories engage in some form of appraisal and an important theme was the need to document the criteria which underlie the appraisal process. The discussion of appraisal led also to some lively discussion of DPE's Position Paper on the subject.

A precondition for future preservation is the generation of appropriate metadata during ingest and this session also included a presentation of the PREMIS standard for preservation metadata.

3.3. SESSION 3: OPERATIONAL ACTIVITY FOR DISCOVERY, ACCESS AND PRESERVATION

In this session we were introduced to a number of working repositories and federated systems which provided excellent examples of current trends in the field. Kopal is a German project to build an archival system capable of offering long-term preservation for a wide variety of object types. The project's software is Open Source, enabling other institutions to benefit from the effort expended on Kopal. The workshop also heard about the Koninklijke Bibliotheek (KB) of the Netherlands e-Depot digital archiving environment. Like Kopal, e-Depot is based on IBM's DIAS system. This is a highly scalable architecture, capable of dealing with the very large number of objects which are ingested into



the Dutch system from Elsevier. The Workshop also heard about KB's research on preservation, and specifically the DIOSCURI system – a modular open-source emulator, written in Java.

The BASE project is an aggregated search engine. Data is harvested from many (currently 516) repositories via OAI-PMH and made available for searching via a single interface. It is an excellent example of how the adoption of interoperability standards (in this case OAI-PMH) allows the development of a service layer which can add value to the material in a given repository by linking it to material in other repositories.

3.4. SESSION 4: PRESERVATION ACTIVITY

Directly following on from the previous session, this session focused more specifically on the techniques and strategies for dealing with preservation in operational environments. The PLANETS project is developing the concept of Preservation-Planning Trees. These are decision trees which will enable repository holders to take appropriate preservation actions for their holdings. The workshop also heard about LOCKSS which provides a sophisticated and award-winning system for digital preservation which is nevertheless cheap to setup and run. In addition, the workshop heard about the Danish PINDAR project which aims to provide a preservation service to the national repository infrastructure, with functional preservation of the material built-in from the start.

3.5. SESSION 5: QUALITY CONTROL, CERTIFICATION AND SELF-AUDIT

A central issue in preservation activity is how to ensure both the quality of the material and metadata ingested into a repository and the fitness of the repository itself to store the material. This session addressed both the principles and practices behind the concept of a “Trusted Repository”. An important development in the field has been the adoption of a core set of ten basic requirements, agreed on by several of the most active participants in this field. The workshop then heard about the DINI certification program for German institutional repositories. Extending certification programs to cover scientific data repositories is a great challenge because of the diversity of such repositories and the amount of domain-specific knowledge embedded in their structure. The workshop heard about the analysis by the TRAC group of eight such repositories representing widely divergent fields and technical requirements. An important point was that repositories, in general, seem to have a positive attitude to being audited, with many hoping to receive constructive feedback to enable them to improve their long-term viability. The workshop was then introduced to the DRAMBORA toolkit for repository self-audit which takes a risk-analysis approach which is complementary to the checklist approach adopted by other related projects.



4. CONCLUSION

In addition to the many interesting presentations at the workshop, there were many lively and heated, though good-natured, discussions. It was clearly felt that the program had avoided the error of being overfilled with prepared talks to the detriment of open discussion. The basic principles, and indeed definition of, “Digital Preservation” came to the fore several times. The concept of “repository” was also heavily discussed, and the meeting was reminded several times that there exist many important collections of data which are not managed in any standards-compliant way, but which nevertheless ought perhaps to be of concern to the digital-preservation community.

This workshop was organised as a follow-up to the DCC workshop in 2005. From the interesting presentations and lively discussions it is clear that this is a field whose importance continues to grow.



Project information

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

Project Officer: Ariane Labat

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