
 **hatii**

Introduction to Digital Curation Projects at HATII: Digital Curation and Research Activities


HATII U of Glasgow Seminar
Wolfson Medical Building,
Seminar Room 3
26 September 2006



Library at Hadrian's Villa at Tivoli


Prof Seamus Ross,
Director HATII University of Glasgow

© HATII, University of Glasgow

 **hatii**

Humanities Advanced Technology and Information Institute (HATII)

- **HATII**
<http://www.hatii.arts.gla.ac.uk>
- **Digital Curation Centre (DCC)**
<http://www.dcc.ac.uk>
- **DELOS**
<http://www.delos.info>
<http://www.dpc.delos.info>
- **PLANETS**
<http://www.planets-project.eu/>
- **CASPAR**
<http://www.casparpreserves.eu/>
- **DPE**
<http://www.digitalpreservation.europe.eu>
- **AHDS Performing Arts**
<http://www.ahds.ac.uk/performingarts/index.htm>



George Service House, Humanities
Advanced Technology and Information
Institute (HATII)

© HATII, University of Glasgow



Acknowledgements

- To DCC colleagues at Glasgow and the DCC Partners
- To PLANETS Partners and especially Adam Farquhar of the British Library
- To CASPAR Partners and especially David Giaretta of CCLRC

© HATII, University of Glasgow



Digital Futures

- Digital materials are fragile as they are subject to loss because of such factors as hard and software obsolescence, media degradation, syntactical inoperability, semantic loss
- Guaranteeing long term access to them is a challenge.
- Introduction to some of HATII's work in the area of digital preservation
- The chance to meet some of the participants in its research and support programmes.

© HATII, University of Glasgow


 **hatii**

Objective of digital preservation

“retaining the ability to display, retrieve, manipulate, and use digital information in the face of constantly changing technology”



© HATII, University of Glasgow

 **hatii**

A Starting Point

- Study of Digital Archaeology funded by JISC and British Library
- ERPANET – EU co-funded activity in area of digital preservation
- NSF/DELOS Workgroup in Digital Preservation to establish a Digital Preservation Research Agenda – ‘Invest to Save’



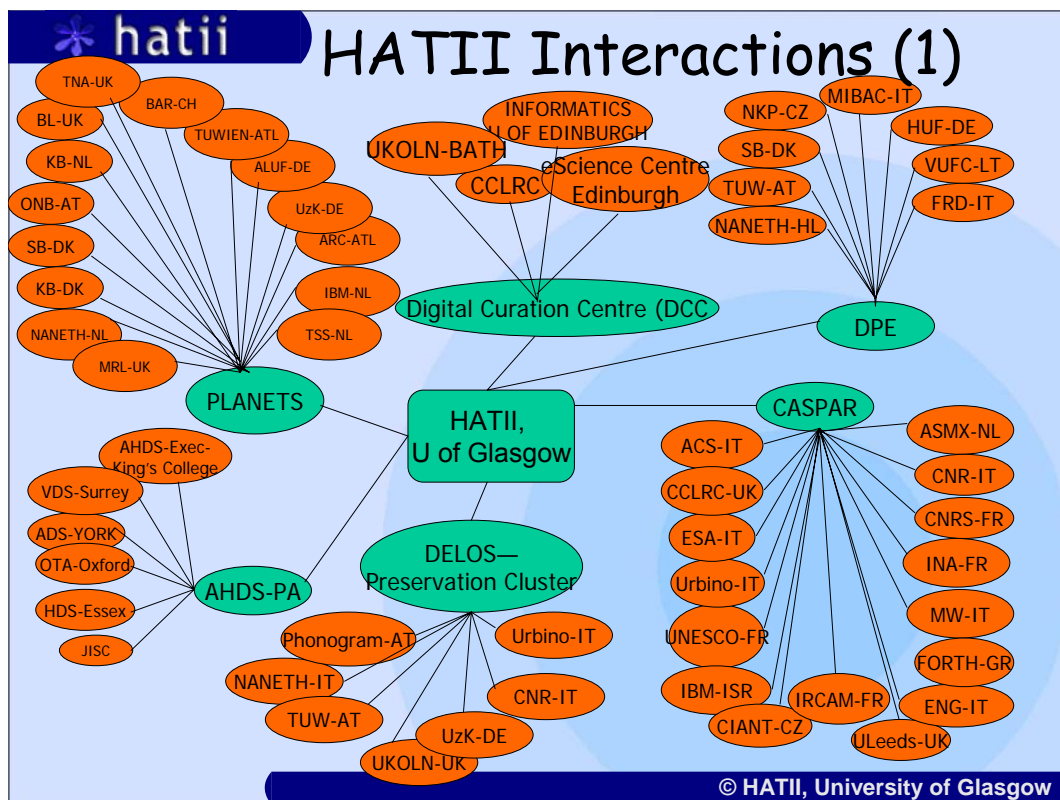
Historic Media on Display at the Launch of the UK Digital Curation Centre (DCC), Nov 04

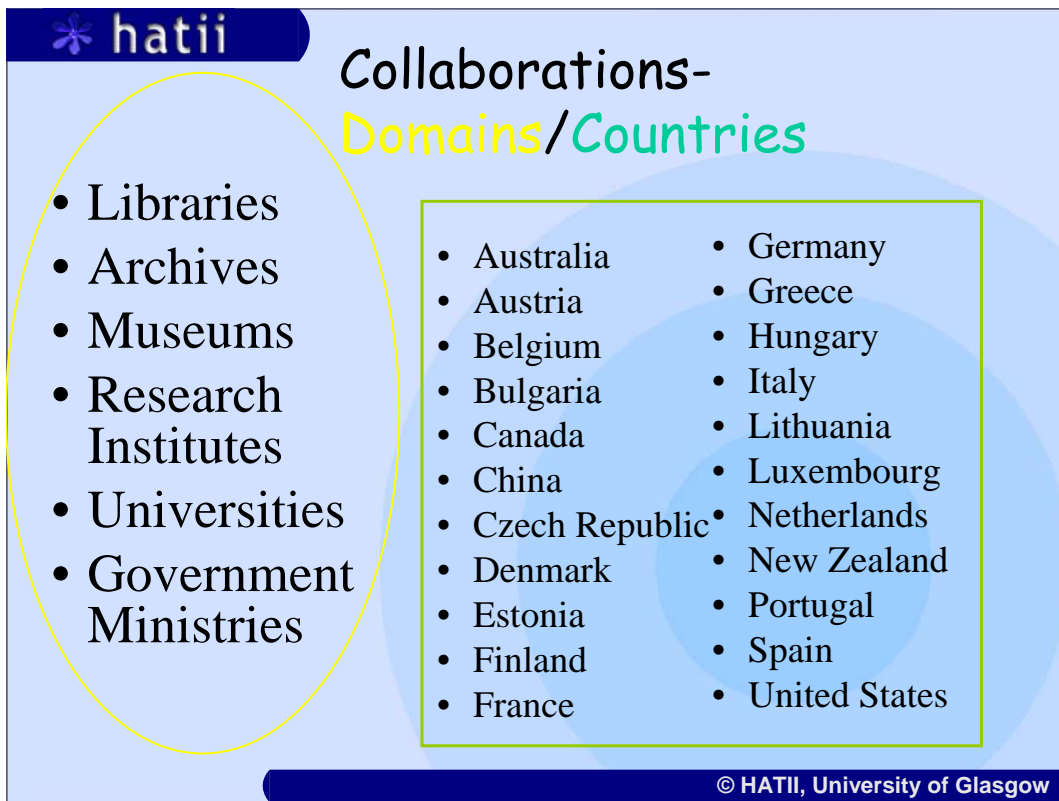
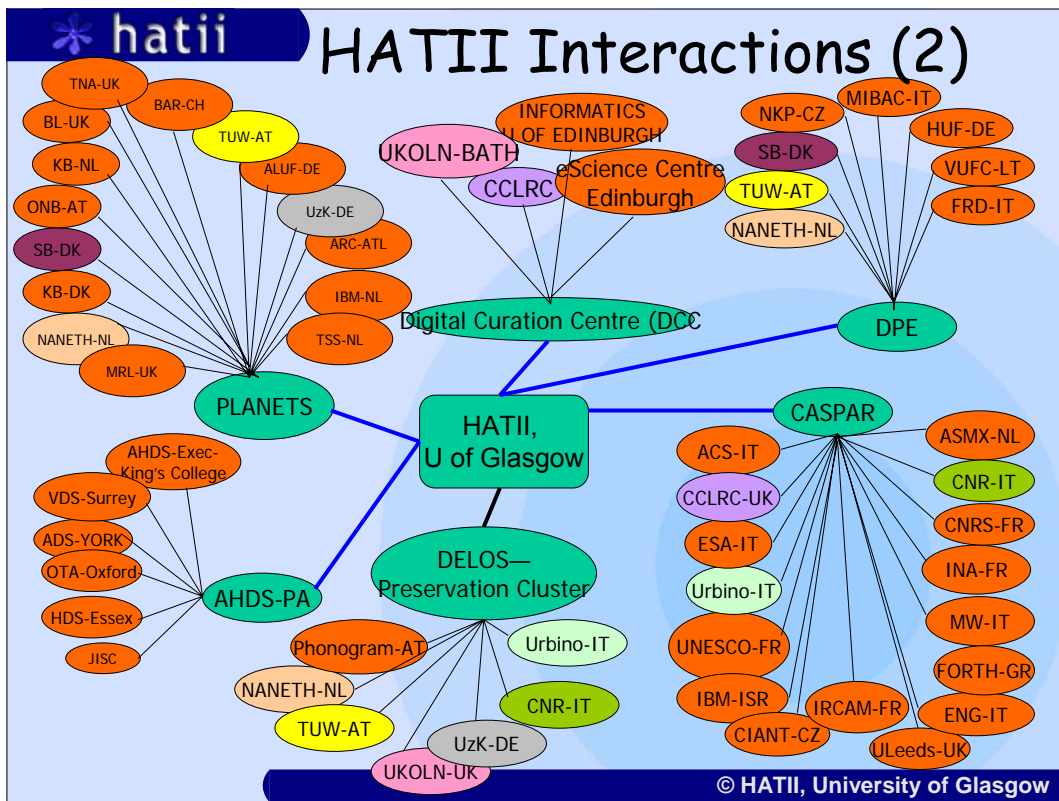
© HATII, University of Glasgow

hatii Leadership in Digital Preservation

- Digital Curation Centre (DCC)
- LOCKSS UK Pilot Programme
- Digital Preservation Cluster of the DELOS Network of Excellence in Digital Libraries
- CASPAR: Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval
- DigitalPreservationEurope (DPE)
- PLANETS: Preservation and Long-term Access through NETworked Services
- AHDS Performing Arts

© HATII, University of Glasgow

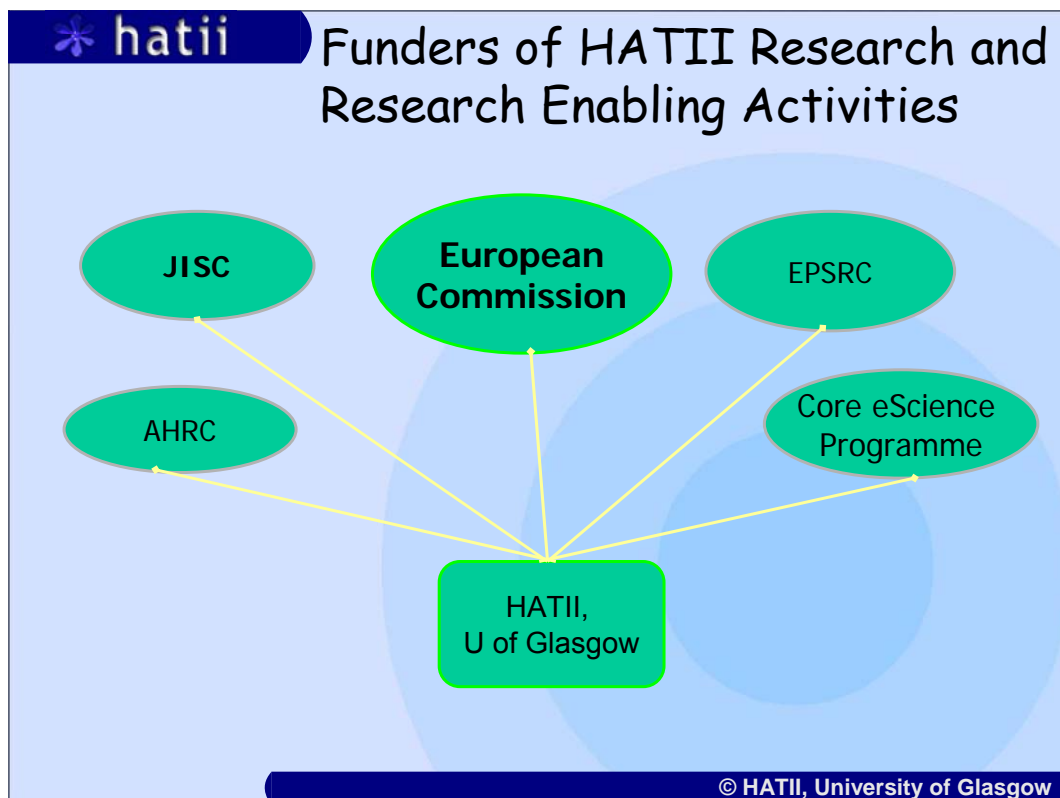




hatii Collaborations with over 100 institutions across 20 countries -- more than 50 of which involve partners engaged in jointly funded projects

- ADS-AHDS, University of York (UK)
- Advanced Computer Systems ACS S.p.A. (IT)
- AHDS-EXEC, King's College London (UK)
- Albert-Ludwigs-Universität Freiburg (DE)
- ARC Seibersdorf research GmbH (AT)
- Asemantic S. r. l. (IT)
- Centre National de la Recherche Scientifique (IT)
- Council for the Central Laboratory of the Research Councils (UK)
- Det Kongelige Bibliotek (NL)
- Engineering – Ingegneria Informatica (IT)
- European Space Agency, ESRIN (IT)
- FernUniversität Hagen (DE)
- Fondazione Rinascimento Digitale (IT)
- Foundation for Research and Technology – Hellas (GR)
- HDS-AHDS, University of Essex (UK)
- IBM Haifa Research Laboratory (IS)
- IBM Nederland N.V.(NL)
- Informatics University of Edinburgh (UK)
- Institut National de l'Audiovisuel (FR)
- International Centre for Art and New Technology in Prague (CZ)
- Istituto di Scienze e Tecnologie dell'Informazione – CNR (IT)
- Koninklijke Bibliotheek (NL)
- l'Institut de Recherche et Coordination Acoustique/Musique (FR)
- Metaware (IT)
- Microsoft Research Ltd (UK)
- Ministero Per I Beni E Le Attività Culturali (IT)
- Národní knihovna Ceske republiky (CZ)
- Nationaal Archief (NL)
- National and Capodistrian University of Athens (GR)
- Österreichische Nationalbibliothek (AT)
- OTA-AHDS, University of Oxford (UK)
- Phonogrammarchiv der Österreichischen Akademie der Wissenschaften (AT)
- Schweizerisches Bundesarchiv (CH)
- Statsbiblioteket (DK)
- Technical University of Crete (GR)
- Technische Universität Wien (AT)
- Tessella Support Services plc (UK)
- The British Library (UK)
- The National Archives of England, Wales and the United Kingdom (UK)
- UKOLN, University of Bath (UK)
- UNESCO (FR)
- Università degli Studi di Firenze (IT)
- Università degli Studi di Roma "La Sapienza" (IT)
- Università degli studi di Urbino, Italia (IT)
- Universität Duisburg-Essen (DE)
- Universität zu Köln (DE)
- University of Göttingen (DE)
- University of Leeds (UK)
- VDS-AHDS, University of Surrey (UK)
- Vilnius University Faculty of Communication (LT)

© HATII, University of Glasgow





Reasons for Research Gap



- there has been a lack of appreciation of the research challenges posed by digital preservation,
- a lack of a sense of urgency, the lack of proven business cases which might have encouraged the development of this as a research or technology sector,
- the fact that in the past the research agenda has been driven by information professionals working in memory institutions or corporate records management teams,
- the limited funding for this kind of research, and,
- of course, the speed of technological development.



Background



- Working Group on Digital Preservation and Archiving
 - Joint EU and US Workgroup
 - Sponsored by DELOS -- Europe's Digital Library Network of Excellence) & National Science Foundation (NSF) in the US
 - *Create a Research Agenda for Digital Preservation that will enable research and create new market opportunities for the information society.*
 - *Published as Invest to Save and some bits have been taken up by the Digital Preservation Cluster of DELOS*



DELOS: The Main Objective

To define and conduct **a joint program of activities** in order to integrate and coordinate the on-going research activities of the major European research teams in the field of digital libraries for the purpose of developing the next generation digital library Technologies – <http://www.delos.info>



Foci in DELOS

- Digital Library Architecture (WP1)
- Information Access and Personalization (WP2)
- Audio/Visual and Non-Traditional Objects (WP3)
- User Interfaces and Visualization (WP4)
- Knowledge Extraction and Semantic Interoperability (WP5)
- **Preservation (WP6) (<http://www.dpc.delos.info>)**
- Evaluation (WP7)





DELOS DP Cluster Objectives

- Catalyse research and funding environment to promote delivery of the DELOS/NSF research agenda for *Digital Preservation and Archiving*;
- To examine core issues that will deliver essential methods, and tools to enable the construction of preservation functionality within digital library activities and deliverables are created;
- Establish testbed frameworks and validation metrics;
- Relate the digital preservation research agenda more directly to the development of exploitable product opportunities and to develop links with the industrial sectors;
- Make progress on the automation (or at least semi-automation) of the processes of ingesting material into preservation environments
- Promote the adoption of preservation technologies in digital library development designs; and,
- To raise the profile of digital preservation issues within the Digital Library Community.



Digital Preservation Testbed Forum -- Achievements

- Primary Objective
 - Establish a framework of a digital preservation testbed environment.
- Achievements
 - Defined on paper a framework for a digital preservation testbed environment and produce metrics for testing and validating digital preservation strategies supporting
 - Design Digital Preservation Testbed Environment for conducting repeatable and measurable experiments for identifying appropriate preservation strategies and methods,
 - Ensuring consistency in the framing of research questions,
 - Delineating the experiment process, describes the Testbed environment, and presents the products of the research
 - Will provide an add-on to OAIS which lacks preservation system





File Formats, Classification, and Typology -- Achievements

- Primary Objective
 - Contribute to the development of file format registries and the mechanisms for their use and define the relationship between file format types and preservation methods
- Achievements
 - Fundamental file format criteria identified and discussed
 - Analysis of existing literature showed inconsistencies
 - Digital preservation strategies for the file format management are presented
 - Defined the concept of transformability as
 - file format categorization and the identification of the existent classification criteria. Defining the relationship between file format types and preservation methods to be done in WP6-T3 to do by feeding into WP6-T4
 - Defined further research that is needed but the research questions are outside the funding scope.



Documentation of Functionality and Behaviour Metrics -- Achievements

- Primary Objective
 - Define framework for documenting behaviour and functionality through building the attributes of functionality and behaviour described and layout mechanisms for representing them.
 - Investigate the viability of automating the process of functionality and behaviour verification
- Achievements
 - A workflow for the elicitation of these requirements was designed, resulting in a tree of objectives.
 - First approach to make preservation approaches comparable and measurable
 - the application of preservation actions evaluated and different preservation strategies examined
 - a model was put into first field trials with partners of the preservation cluster, particularly focusing on the audio collection of the OEAW-Phonogrammarchiv.
 - Supports transparent and informed decision making





Enabling the Integration of Digital Preservation Architectures -- Achievements

- Primary Objective
 - Develop the requirements for a preservation functionality modelling tool.
 - Develop a models of preservation functionality.
- Achievements
 - a comparative evaluation of diverse design methodologies (completed)
 - Examination as to how preservation aspects can be integrated into design methodology (completed)
 - The effort is currently directed, therefore, to an attempt to identify as many possible spots within the existing UML language constructs to which "preservation extensions", based primarily upon the state diagrams of UML, to which preservation aspects can easily be attached.



DELOS DPC 2006-7

- we will integrate the current extraction tool with Utility Analysis Tools;
- we will promote the extrapolation of our work to ensure that other tools are developed; and,
- we will aim to proceed with automating the process of appraisal (and re-appraisal).
- Construction of a corpus to support research in appraisal and automated metadata extraction
- 3rd DELOS International Summer School in Digital Preservation in Digital Libraries (San Miniato)





Mission of the Digital Curation Centre (DCC)

“The over-riding purpose of the DCC is to support and promote continuing improvement in the quality of data curation, and of associated digital preservation”

Note the emphasis on ‘DATA’



Digital curation?



Dynamic
Long-term



Static



Digital curation

In use now (and the future)

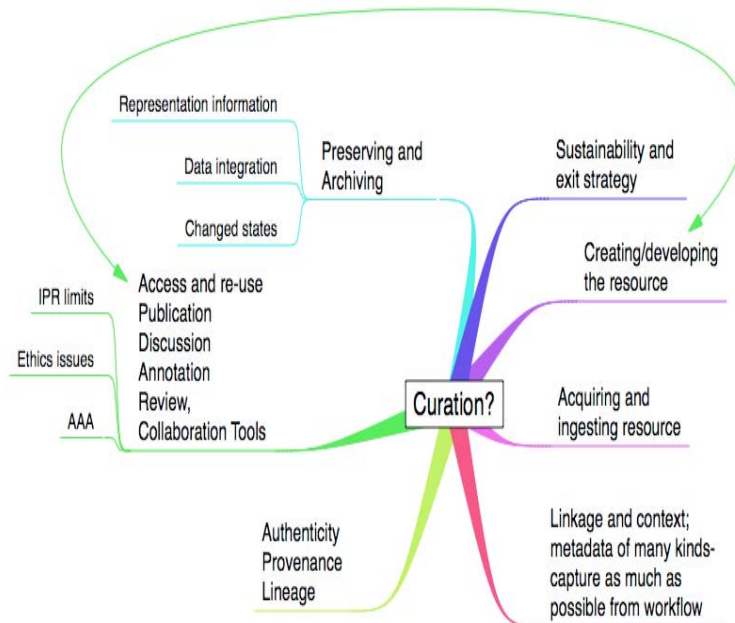
For later use

Dynamic
Long-term

Static



“maintaining and adding value to a trusted body of digital information for current and future use”





DCC Objectives

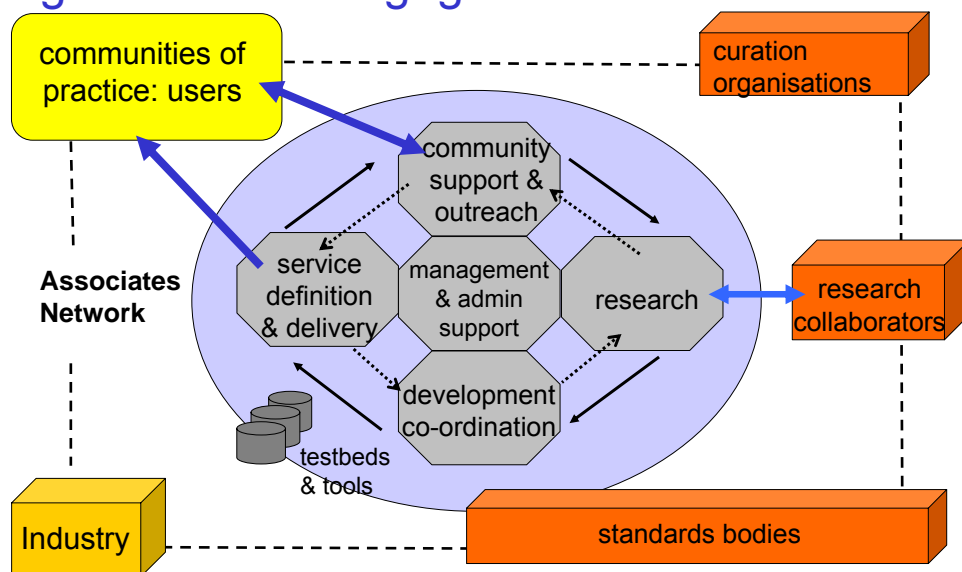
- Vibrant research programme
 - addressing the wider issues of digital curation
- Collaborative Associates Network of Data Organisations
 - strong links across existing community of practice
 - engagement with curators (individuals & organisations)
- Development & services
 - to evaluate tools, methods, standards and policies
 - a repository of tools and technical information
- ‘virtuous circle’
 - expertise, experience & requirement feed into the DCC research programme

:: DCC—An Introduction ::

27



Organisation to Engage & Collaborate



:: DCC—An Introduction ::

28



D | C | C

a centre of expertise in data curation and preservation

DCC Research Achievements

- Bath: repository interactions
- CCLRC: metadata curation
- Edinburgh Database Group: Annotation, archiving, citation, lineage, provenance, publishing
- Glasgow: metadata extraction and genre classification and Audit and Certification methodologies

:: DCC—An Introduction ::

29



D | C | C

a centre of expertise in data curation and preservation

Why these research topics?

- they involve fundamental problems
- they will generate highly visible research
- most have components that can be almost immediately carried into development and testing
- of interest to partners and to associates and clients.

:: DCC—An Introduction ::

30



D | C | C

a centre of expertise in data curation and preservation

DCC Outreach Objectives

- **Publicity, promotion and dissemination**
 - Raise awareness, promote and disseminate DCC activities.
- **Web portal**
 - Provide support for the community/individuals and provide access to services and resources
- **Training and professional development**
 - Increase the skills and knowledge base of practitioners within different communities of practice.
- **User Requirements Analysis**
 - Gather information about the user / stakeholder community.
- **2nd Annual DCC Conference, Glasgow 20-22 November 06**

:: DCC—An Introduction ::

31



D | C | C

a centre of expertise in data curation and preservation

Services Objectives

- **Dissemination of research & development**
 - Transform and make viable results of research
- **Standards Watch, tools, and repositories**
 - Ensure access to standards, tools and repositories of Representation Information
- **Advisory Services**
 - Provide guidance on curation and preservation
 - Curation Manual, Briefing Papers
- **Audit and Certification**
 - Deliver Audit and Certification Services to the UK Community
- **Developmental Workshops & CPD**

:: DCC—An Introduction ::

32



Development Deliverables

- Registry/Repository development
 - Review registry data model of related projects and agree common model as far as possible
 - Pilot Registry for Data Formats
 - Design of Representation Information Repository
- Representation Information development
 - Data Description
 - Software
 - Time Dependence
 - Process
- Testing and Certification processes
 - Standards
 - Testbeds
 - Certification

:: DCC—An Introduction ::

33



DCC
Digital Curation Centre

home | contact us | help | search | sitemap | rss

About the DCC

- Training & Events
- Resource Centre
- Tools & Standards
- Research & Development

Associates

- Network
- Discussion Forum

Adding Information

- Helpdesk
- FAQs

International

- Journal of Digital Curation

Welcome [Smaller Text](#) | [Larger Text](#)

Welcome to the **Digital Curation Centre**. Scientists and researchers across the UK generate increasingly vast amounts of digital data, with further investment in digitisation and purchase of digital content and information. The scientific record and the documentary heritage created in digital form are at risk, by technology obsolescence and by the fragility of digital media.

Working with other practitioners, the **Digital Curation Centre** will support UK institutions to store, manage and preserve these data to ensure their enhancement and their continuing long-term use. The purpose of our Centre is to provide a national focus for research into curation issues and to promote expertise and good practice, both national and international, for the management of all research outputs in digital format.

News Key: DCC news External news

Collection and Preservation of At-Risk Digital Geospatial Data
Steve Morris of North Carolina State University (NCSU) Libraries presented this North Carolina NDIIPP Project - conducted with the U.S. Library of Congress in conjunction with the fairly new National Digital Information Infrastructure and Preservation Program (NDIIPP) - at the joint Digital Curation Centre (DCC)/Database Seminar held on 31 May 2005 at the University of Edinburgh. (09/06/05)
[View or download PowerPoint slides \(3.5MB\)](#)

Calendar

Keep up to date with Digital Curation events using the DCC calendar. You can also [add an Event to the calendar](#).

« June 2005 »

S	M	T	W	T	F	S
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30		

[View weekly calendar](#)

- [View complete list of DCC events](#)
- [View complete list of external events](#)
- [Take the Digital Curation Questionnaire](#)

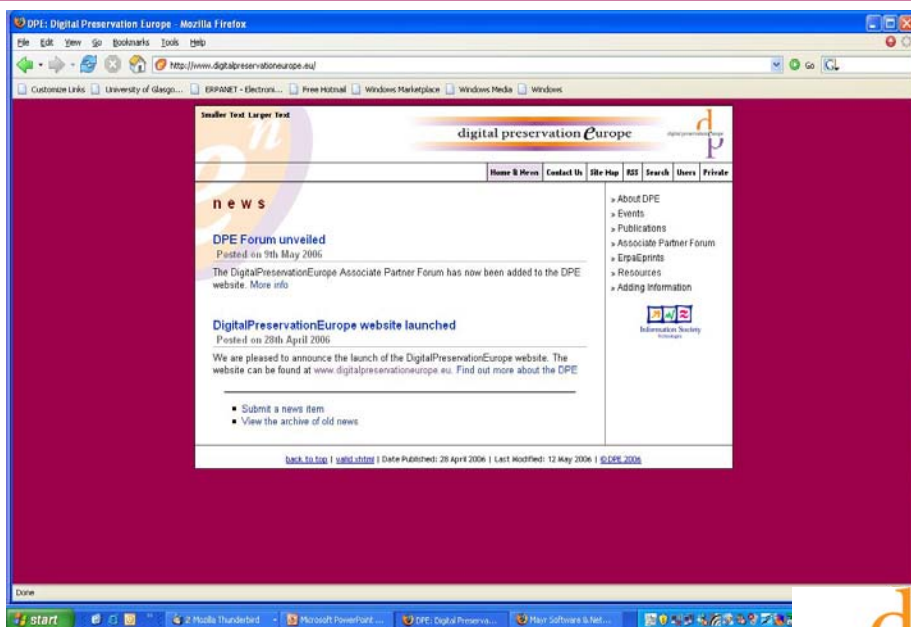
www.dcc.ac.uk

What is *DPE*?

- *DigitalPreservationEurope (DPE)* fosters collaboration & synergies between existing national initiatives across the European Research Area.
- *DPE* addresses the need to improve coordination, cooperation, and consistency in current activities to secure the longevity digital materials.
- DPE builds on ERPANET (2001-2004)



DigitalPreservationEurope (DPE): Introduction



DigitalPreservationEurope (DPE): Introduction



The *DPE* Vision

- 36 month project from 1 April 06
- To create a coherent platform for proactive cooperation, collaboration, exchange and dissemination of research results and experience in the preservation of digital objects.
- To increase prevalence of preservation services and their viability and accountability. Agenda
- Improve awareness, skills, and available resources.



DigitalPreservationEurope (DPE): Introduction



Core Activities

- **Advocacy and Outreach**
 - Briefing Papers and Fact Sheets
 - FP6 Preservation Showcase Annual Conference (Glasgow, 23 November 06)
 - Awareness Workshops and Presentations
 - European Digital Preservation Challenge
- **Training & Continuing Professional Development**
 - Coordination Framework for Training and Education based on emerging research results
 - Co-ordinate and Deliver DPE Research Driven Training Programme



DigitalPreservationEurope (DPE): Introduction



Deploying and Enabling Research

- Coordination of EU Repository Activities
 - Inventory of repository activities and federation support
 - Facilitate the collection of repository experiences
 - Repository Planning Tool
- Audit and Certification Services
 - Create Self-Audit Tool
 - Piloting of the DPE Audit and Certification Scheme
- Research and Practitioner Integration
 - Pilot Research and Industrial Exchange Programme
 - Creation of a Research Database



DigitalPreservationEurope (DPE): Introduction



Research Roadmap

- *Preservation Research Roadmap*
 - Fifteen years of discussion has not resulted in a concise and well-developed strategy
 - Builds on NSF/DELOS Workgroup Report – *Invest to Save*
 - DPE aims to deliver this Research Roadmap – it should impact on research funded under FP7
- *European Competence Centres*



DigitalPreservationEurope (DPE): Introduction





PLANETS Stakeholders

- ❑ The EU Co-Funded Planets Project engages partners motivated to find practical and deployable solutions to the obstacles to digital longevity
- ❑ ...leading Archives and Libraries
 - ❖ National Archives of Great Britain, The Netherlands, Switzerland
 - ❖ National Libraries of Austria, Denmark, Great Britain, The Netherlands
- ❑ ... leading research universities
 - ❖ University of Cologne, University of Freiberg, HATII at University of Glasgow, Technical University of Vienna
- ❑ ... leading technology suppliers
 - ❖ Austrian Research Center (ARC), Tessella Plc
 - ❖ IBM
 - ❖ Microsoft



Key Outcomes from PLANETS

- ❑ Increase Europe's ability to ensure long-term access to its cultural and scientific heritage
- ❑ Improve decision-making about long term preservation
- ❑ Ensure long-term access to valued digital content
- ❑ Control the costs of preservation actions through increased automation, scalable infrastructure
- ❑ Promote the adoption across the user community of preservation services and tools
- ❑ Establish market place for preservation services and tools

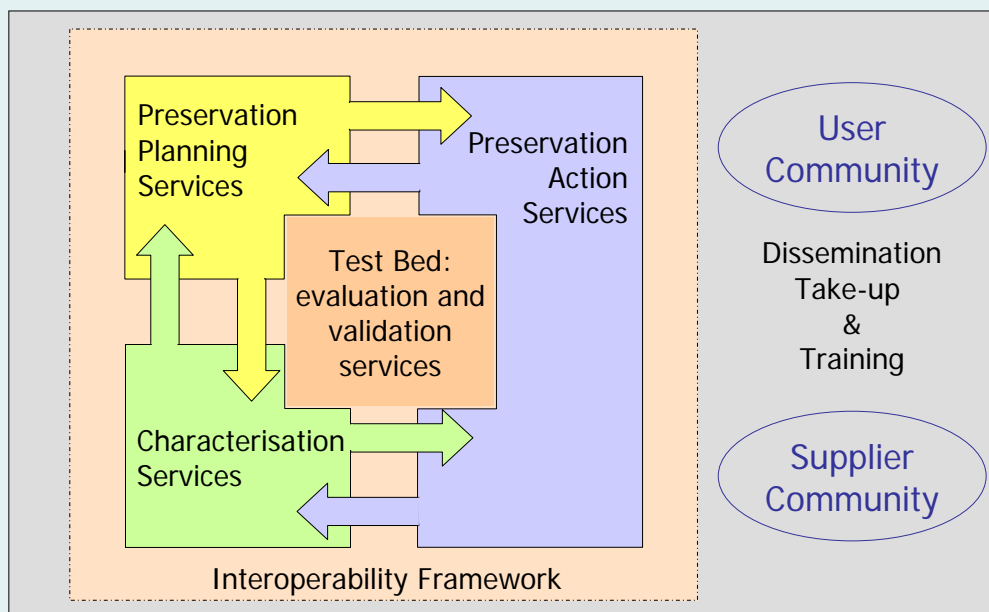


Modularisation of the Challenges

- PLANETS, in 48 months from 1 June 06, will develop:
 - ❖ Planning services that empower organisations to define, evaluate, and execute preservation plans
 - ❖ Methodologies, tools and services for Characterisation of digital objects
 - ❖ Innovative solutions for Preservation Actions
 - ❖ An Interoperability Framework to seamlessly integrate tools and services in a distributed service network
 - ❖ A Testbed to provide a consistent and coherent evidence-base for the objective evaluation of different protocols, tools, services and complete preservation plans.



Project Architecture Reflects Problem Structure



Preservation Planning, Characterisation, and Action

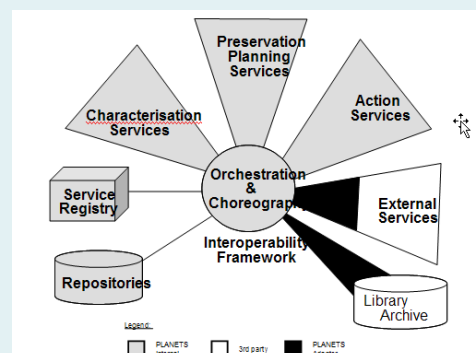


- ❑ **Preservation Planning**
 - ❖ Builds on DELOS' Digital Preservation Cluster's preservation planning outputs
 - ❖ Input:
 - ❖ Feedback
 - ❖ Execution
 - ❖ Validation
- ❑ **Characterisation**
 - ❖ Enhance tools for file-format identification
 - ❖ Define a characterisation language for file-formats
 - ❖ Define an extraction language
 - ❖ Define a pluggable interpreter
 - ❖ Leverage understanding to recommend improvements in file format design
 - ❖ Establish methods for evaluating loss due to preservation actions
- ❑ **Preservation Action**
 - ❖ Provide pluggable wrap third-party transformation tools
 - ❖ Address preservation of relational databases
 - ❖ Address need to preserve Microsoft Office content
 - ❖ Modular emulation of the full hardware/software environment
 - ❖ Layered durable approach to emulation

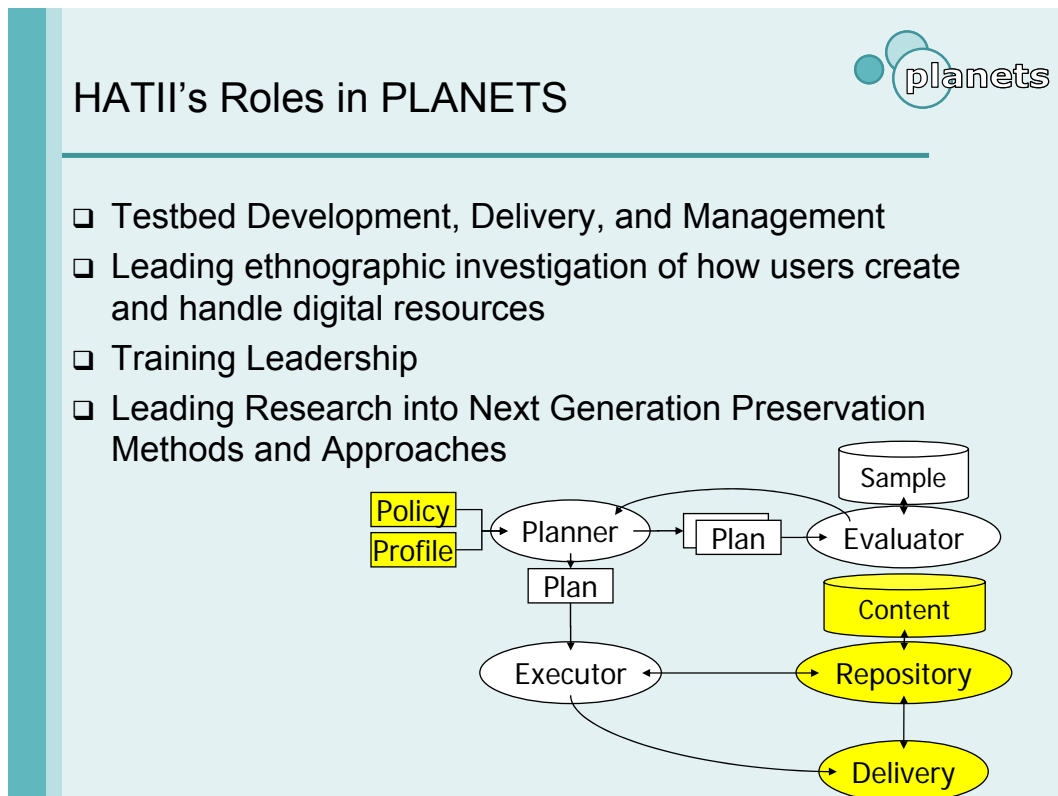
Interoperability Framework and Testbed



- ❑ Planets provides an interoperability framework



- ❑ **PLANETS testbed**
 - ❖ Provides a foundation for objective evaluation of preservation actions, characterisation services, and preservation plans
 - ❖ Provides a controlled research infrastructure
 - ❖ Enable partners to conduct experiments, load content objects, execute tests, evaluate the results, and compare the outcome to similar tests



WHY CASPAR

- Need to preserve information & knowledge – not just “the bits”
- Need to manage knowledge to keep archives alive through time
- Preservation is a process, not a one-shot event
 - transforming content (migration, emulation, etc.) to adapt it to new constraints of rendition and playability
 - enriching content to preserve its intelligibility and (re)usability (not just rendering)
- OAIS provides a general framework:
 - current implementations deal more with format than the interpretation of data
 - CASPAR proposes dealing with content interpretation

SIP = Submission Information Package
AIP = Archival Information Package
DIP = Dissemination Information Package



CASPAR Objectives

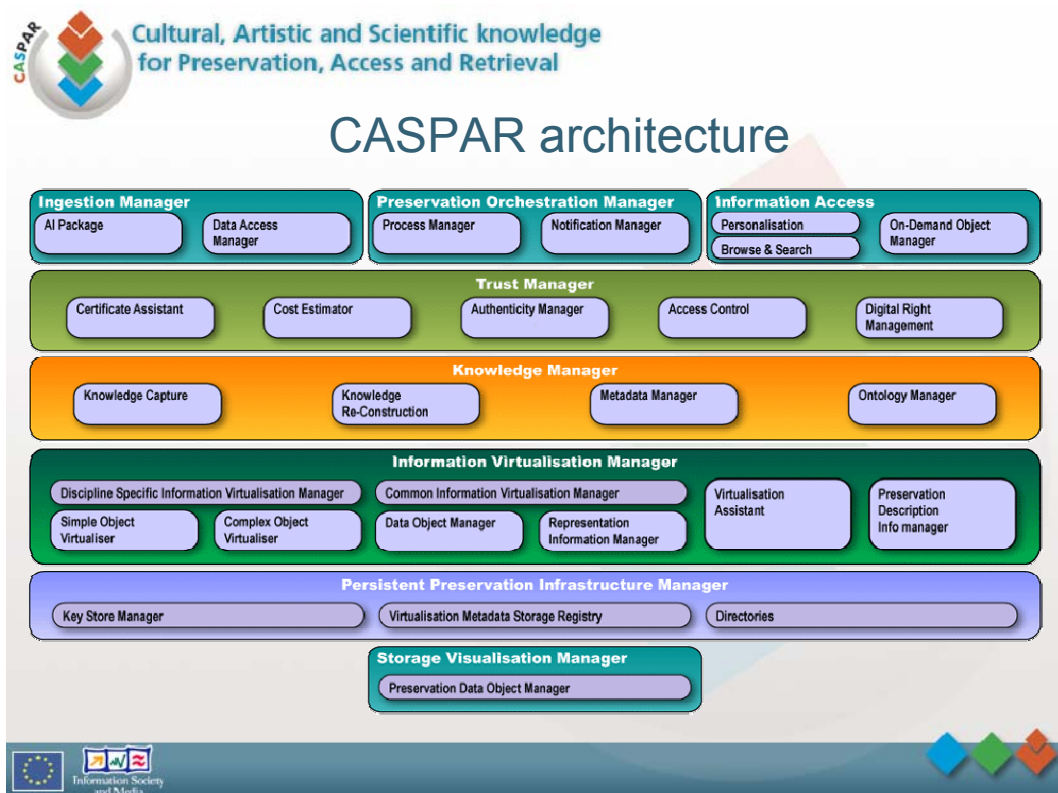
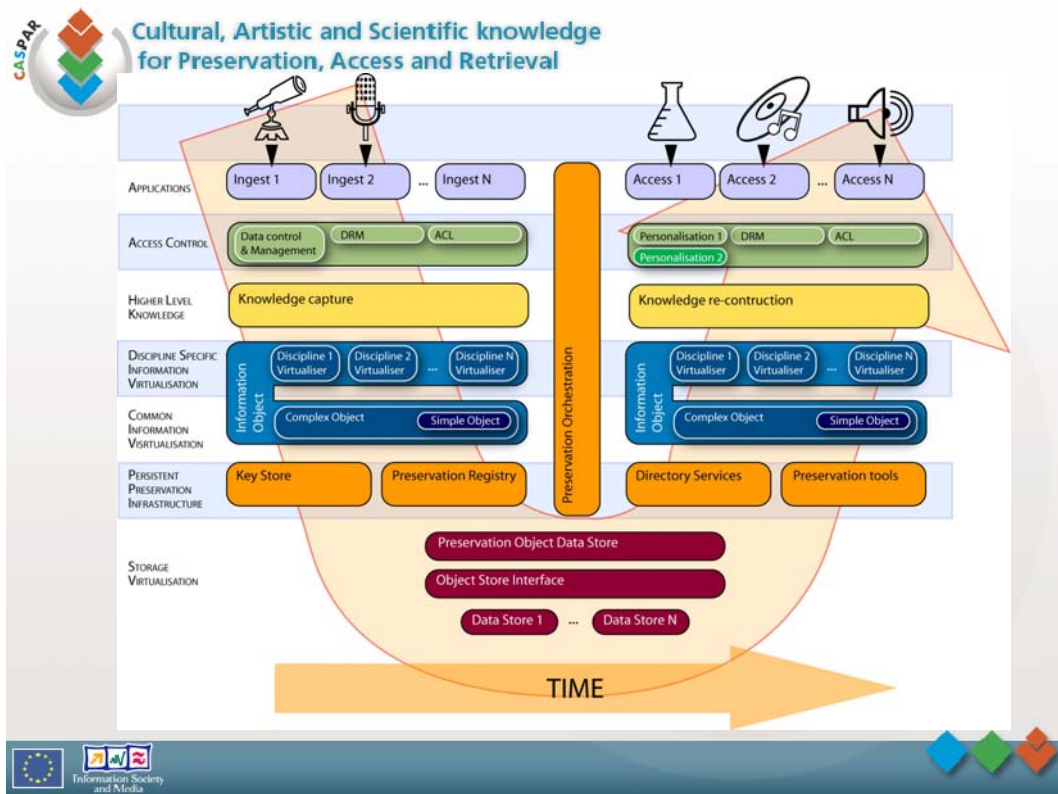
- **CASPAR methodology:** to lay the foundation for all future preservation activities
- **CASPAR components :** to create components to use in all preservation activities
- **CASPAR framework :** to create the long-term autonomous system to support all the preservation activities
- **CASPAR testbeds:** to demonstrate the validity of the CASPAR framework with heterogeneous data and a variety of innovative applications



CASPAR expected results

- **CASPAR approach and framework** to support the “end-to-end” lifecycle for digital resources
 - Infrastructure
 - Tools
 - Techniques
- **Testbeds:** science, culture, artistic to identify and test common infrastructure
 - Supported by discipline specific access
 - Embedded in long-lived institutions
- **Interfaces (long-lived!)**
- **Implementations (short-lived?)**







HOW: Knowledge at the heart of preservation

- Knowledge driven approach
- Knowledge management to support long-term preservation of concepts/information including:
 - Single, complex, on demand, interactive objects
 - DRM
 - Authenticity
 - Access
 - Storage
 - Designated Community – descriptions
 - Knowledge base definition
 - ontologies



CASPAR Testbeds

- Three testbeds
 - Cultural: UNESCO
 - Performing Arts: INA , IRCAM
 - Scientific: ESA (with CCLRC)
- Complex, multi-source, multifaceted data
- Common preservation & evaluation & validation
- Specific requirements on preservation (technical, delivery, legal)
- Specific research issues: as matter of facts, they represents three focused research streams
- Specific user communities
- Identifying and confirming common infrastructure elements





Cultural, Artistic and Scientific knowledge
for Preservation, Access and Retrieval

HATII's Role in CASPAR

- CASPAR methodology and models
- OAIS-based components –dealing with Representation Information
- OAIS-based access
- Construction of Generic Testbed Elements



Cultural, Artistic and Scientific knowledge
for Preservation, Access and Retrieval

www.casparpreserves.eu



* hatii Preservation/Curation Research


- HATII plays a core role in current leading research and preservation enabling activities across Europe
- HATII collaborates with teams based in Europe, North America, Asia, and Australia/New Zealand
- HATII has a world-class team working in preservation and aims to continue to build this

© HATII, University of Glasgow


* hatii Core Themes

- Research Agenda Shaping
- Research Vision & Leadership
- Cutting-edge research
- Training and education
- Awareness raising
- Capacity enabling
- Testbed design, development and deployment
- Move preservation research so that it is experimentally led
- Comparison of methodologies: DELOS DPC, CASPAR, PLANETS
- Developing new approaches to curation and preservation

© HATII, University of Glasgow

 **Humanities Advanced Technology and Information Institute (HATII)**

- **HATII**
<http://www.hatii.arts.gla.ac.uk>
- **Digital Curation Centre (DCC)**
<http://www.dcc.ac.uk>
- **DELOS**
<http://www.delos.info>
<http://www.dpc.delos.info>
- **PLANETS**
<http://www.planets-project.eu/>
- **CASPAR**
<http://www.casparpreserves.eu/>
- **DPE (DigitalPreservationEurope)**
<http://www.digitalpreservation.europe.eu>
- **AHDS Performing Arts**
<http://www.ahds.ac.uk/performingarts/index.htm>



© HATII UofGlasgow, 2001

George Service House, Humanities Advanced Technology and Information Institute (HATII)

© HATII, University of Glasgow