



A National Preservation Policy for the UK(?) -Thoughts and Challenges

International Conference on Preservation of
Digital Objects

15-16th Sept. 2005

Manjula Patel
UKOLN and DCC

Funded by: **JISC** 



Digital | Curation | Centre

Outline

- No overall national policy
 - Developments are incremental and organic
- Recent developments in the UK
 - National Preservation Office (1984)
 - JISC Digital Preservation Focus (200)
 - Digital Preservation Coalition (2001)
 - The National Archives (2003)
 - JISC Programmes (2002-2005)
 - Selected influential reports, white papers and surveys
- Digital Curation Centre

Sources of Funding: JISC

Joint Information Services Committee

- Strategic guidance, advice and opportunities for use of ICT to support teaching, learning, research and administration in FE and HE
- Operates through a committee system
- Funded by UK FE and HE councils
- Provides centralised and co-coordinated direction for the development of infrastructure and activities (5 year strategy)

Sources of Funding: RCUK

Research Councils UK

- Eight Research Councils
 - Arts & Humanities Research Council
 - Biotechnology & Biological Sciences Research Council
 - Council for the Central Laboratory of the Research Councils
 - Engineering & Physical Sciences Research Council
 - Economic & Social Research Council
 - Medical Research Council
 - Natural Environment Research Council
 - Particle Physics & Astronomy Research Council
- Fulfil the objectives set out by Government in the White Paper "*Realising our Potential*" (1993)
- Under control of the government's Department of Trade & Industry

4

National Preservation Office

- British Library Board established the NPO in 1984
 - Raise awareness of preservation issues in libraries
 - Provide a focal point for preservation management
- Independent focus for the preservation of and continuing access to cultural heritage materials
- Guided by an independent management committee, members drawn from various funding institutions
- Has become established as a major force in supporting the preservation of library, archive and museum materials in the UK and Ireland

JISC Digital Preservation Focus

JISC DPF was established in June 2000:

- Developing a **long-term retention strategy** for digital materials of relevance to HE/FE institutions in the UK
- Providing a UK focus for the development of **practices, policies and strategies** for the preservation of digital materials
- Generating support and **collaborative** funding from and promoting inter-working with appropriate agencies worldwide
- activities coordinated by Neil Beagrie

Digital Preservation Coalition

- DPC established in 2001 (following Warwick II Workshop, March 1999)
 - to foster joint action to address the urgent challenges of securing the preservation of digital resources in the UK
 - to work with others internationally to secure global digital memory and knowledge base
- Aims to develop a **UK digital preservation agenda within an international context**
- Coalition of 27 members including OCLC and RLG
- Selected DPC outputs
 - **Forums:** e.g. Preservation metadata, Cost models, Institutional repositories, e-Learning materials, web-archiving
 - **Reports:** e.g. Technology watch reports, surveys, annual reports
 - **DPC/PADI** quarterly current awareness digest
 - **DPC Guides, DPC Handbook**

The National Archives

The National Archives of England, Wales and the UK has one of the largest archival collections in the world, spanning 1000 years of British history, from the Domesday Book of 1086 to government papers recently released to the public

- Formed in April 2003 by bringing together the Public Record Office and the Historical Manuscripts Commission
- **Selection**: formal acquisition policy (only 5% of govt. records archived)
- **Preservation**: variety of media (e.g. medieval parchment to digital tapes); risk management framework
- **Access**: online catalogue of records
- **Advice**: management of current records and care of historical archival collections
- Development of PRONOM file-formats registry

E-Science Curation Report

- Commissioned by JISC (2002-2003), authored by Philip Lord and Alison MacDonald, Digital Archiving Consultancy Ltd.
- Study to examine current provision and future needs for curation of primary research data in UK, particularly within an eScience context
- Key findings
 - Confirmation of need for generic support in digital curation
 - Firm endorsement for a **Digital Curation Centre**
 - Identified specific areas of need as well as areas for research
 - Funding for repositories, archive services and related research is generally short-term in nature and has to compete with research projects
- Clarified definition of digital curation

E-Science Curation Report

Differentiation of digital curation, archiving and preservation

- **Curation:** The activity of managing and promoting the use of data from its point of creation, to ensure it is fit for contemporary purpose, and available for discovery and re-use. For dynamic datasets this may mean continuous enrichment or updating to keep it fit for purpose.
- **Archiving:** A curation activity which ensures that data is properly selected, stored, can be accessed and that its logical and physical integrity is maintained over time, including security and authenticity.
- **Preservation:** An archiving activity in which specific items of data are maintained over time so that they can still be accessed and understood through successive change and obsolescence of technologies.

House of Commons Select Committee on Science and Technology Report

- “Scientific Publications: Free for All?”, July 2004
- Endorsement of **institutional repositories** as a means of making research articles more freely and widely available
- Recommended creation of a network of institutional repositories and the **mandate of all publicly funded researchers to archive a copy of all their articles in such repositories**
- Recommendation rejected by British government, but supported by RCUK!
- June 2005 -major JISC programme to support the growth of digital repositories in the UK (£4m) and explore the many cultural, technical and management aspects of creating and managing institutional and other repositories

Freedom of Information Act

- Enables access to information held by public bodies:
 - Publication Schemes
Every public authority must make some information available as a matter of routine through a publication scheme -a public commitment to make certain information available and a guide to how that information can be obtained
 - General right of access
Any person has the right to make a request for information held by a public authority. The authority must usually respond to this request within 20 working days. This right came into force on the 1 January 2005.
- Places an obligation on public authorities to preserve and maintain access to records

DPC/MLA Report

- “Digital Knowledge for all, but what about for ever?”, June 2005
- Survey to inform a **national digitisation strategy**
 - Covered North East and West Midlands
 - How well prepared to deal with problems of keeping digital material in the long-term
 - Significant commitment to digitisation (over 80 projects), 90% projects externally funded and therefore took no account of need to provide long-term support
- Key findings
 - Need for help and advice on a range of DP issues
 - Awareness raising of whole issue of DP is needed
 - **Long-term management of digital material needs to be more firmly embedded in corporate thinking and planning**

JISC Preservation Programmes

- Based on “A Continuing Access and Digital Preservation Strategy for the JISC 2002-2005”, Oct. 2002
 - Major outcome: Digital Curation Centre
- Main related programmes
 - Supporting Digital Preservation and Asset Management in Institutions
 - Digital Preservation and Records Management
 - Supporting Institutional Records Management
 - Digital Repositories

JISC Preservation Programmes

Digital Preservation and Records Management (June 2000–May 2006)

Projects funded:

- Archiving E-Publications
- Data Curation for e-science in the UK
- Digital Curation Centre
- Long-Term Retention and Reuse of E-Learning Objects and Materials
- Requirements and Feasibility Study on Preservation of e-prints
- Revision of the Study of the Records Lifecycle
- The File Format Representation and Rendering Project
- Web-archiving: a feasibility study for JISC and the Wellcome Trust

JISC Preservation Programmes

Supporting Digital Preservation and Asset Management in Institutions
(Oct. 2004 – Sept. 2006)

- Institutional management support
- Digital preservation assessment
- Institutional repository infrastructure development

A selection of projects funded:

- DAAT: Digital Asset Assessment Tool
- eSPIDA: An effective Strategic model for the Preservation and disposal of Institutional Digital Assets
- LIFE (Lifecycle Information for E-literature)
- PRESERV (PReservation Eprint SERVices)
- SHERPA Digital Preservation: Creating a Persistent Preservation Environment for Institutional Repositories

Digital Curation Centre

- Jointly funded under JISC's Digital Preservation and Records Management Programme and the e-Science Core Programme
- Duration: March 2004 – March 2007
- Consortium comprising the Universities of Edinburgh and Glasgow, which together host the National e-Science Centre, UKOLN at the University of Bath, and CCLRC, which manages the Rutherford Appleton and Daresbury Laboratories
- Mission: to support and promote continuing improvement in the quality of digital curation and preservation

RLG DigiNews, Vol.9, No. 4 lists DCC as one of its "Ten Promising Digital Preservation Initiatives"

DCC: Organisation

Management & Co-ordination

- Director Chris Rusbridge (University of Edinburgh)

Community Support & Outreach

- Dr Liz Lyon (UKOLN, University of Bath)

Service Definition & Delivery

- Prof. Seamus Ross (HATII [ERPANET], University of Glasgow)

Development

- Dr David Giaretta (Astronomical Software & Services, CCLRC)

Research

- Prof. Peter Buneman (Informatics, University of Edinburgh)

Matrix Management

- teams drawn from across consortium partners

DCC: What is Digital Curation?

Digital Curation

Digital curation, broadly interpreted, is about maintaining and adding value to a trusted body of digital information for current and future use. The digital archiving and preservation community now looks beyond the preservation, cataloguing and cross referencing of static digital objects such as documents. The scientific community has data characterised by structure, volatility and scale. These require us to extend our notions of curation. We must also investigate the principles that underlie appraisal, and lessons learnt about the economics of preservation.

DCC: Aims and Objectives

- Promote an understanding of the need for digital curation amongst disciplines
- Encourage inter-disciplinary sharing of knowledge
 - Establish strong links across existing communities of practice
 - Engage with curators, both individuals and organisations
- Provide services to facilitate digital curation
 - Evaluate tools, methods, standards and policies
 - Provide a repository of tools and technical information
 - Disseminate practical advice and guidance
- Develop long-term research and technologies to support digital curation

DCC: Training and Events

- Workshops
 - DCC/ERPANET & Wellcome Trust Workshop on **Preservation of Medical Databases**: 13-14th Oct. 2005
 - DCC / DPC Workshop on **Cost Models**, 26th July 2005
 - DCC Workshop on **Long-term Curation within Digital Repositories**, 6th July 2005
 - DCC Workshop on **Persistent Identifiers**, 30 June - 01 July 2005
- Site visits
 - A series of regional site visits to gather and exchange information about current practice: Bath, Glasgow, London
- Sponsorship of Conferences and Workshops
 - First DCC Conference, 29-30th Sept. 2005, Bath, UK
- Various talks, presentations and seminars

DCC: Resource Centre

Aimed at data specialists, records managers, librarians, archivists, researchers (as data creators), and policy-makers

- **Curation Manual**: community-driven, peer-reviewed, detailed and practical information, 20 chapters commissioned
- **Briefing Papers**: concise overviews designed to help improve basic understanding of key curation issues and provide a high-level introduction to a given topic
- **Interviews**: carried out to assess requirements for digital curation
- **Advisory Service**: offers on-demand advice ranging from questions of a complex technical nature to the pursuit of relevant legal advice; helpdesk@dcc.ac.uk
- **International Journal of Digital Curation**: peer-reviewed; devoted to papers, articles and news items on the curation of digital objects and related issues: <http://www.ijdc.net/>

DCC: Tools and Standards

- Tools and Test-beds: A catalogue of digital curation tools and technical information is available
 - Data Description Tools
 - Data Storage Tools
 - Digital Repository and Library Models
 - Documentation and Standards
 - Management Software
 - Strategy Tools and Guidelines
 - Relevant Projects
 - Web Archiving Tools
- Standards Watch: to ensure that digital curation services are based on sound and tested practice

DCC: Associates Network

- **Goals:** Develop understanding, share best practice, advance research, promote recognition, develop consensus
- **Membership:** Free, International groups, national bodies, industry partners, funders, research groups, HEIs, FEIs, individuals.....
- **Benefits:** Early access to R&D outputs, advisory services, training, input to definition and design, community participation
- **Discussion Forum:** <http://www.dcc.ac.uk/>
- **Please join us!**

24

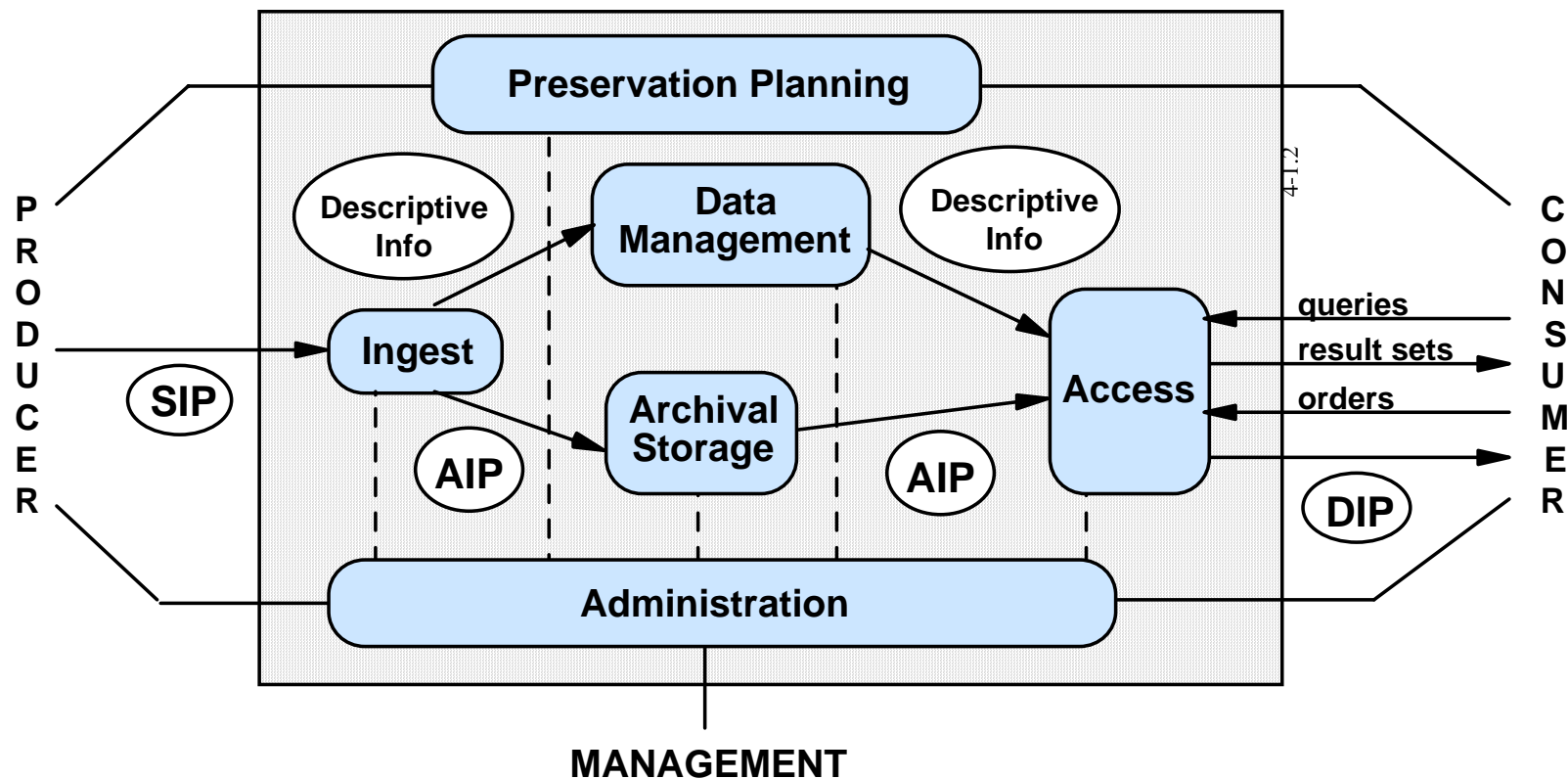
DCC: Research

- Led by Professor Peter Buneman
- Current research priorities, centred around science databases:
 - Annotation in databases
 - Data archiving
 - Socio-economic and legal issues
 - Metadata extraction and curation
 - Provenance and databases
 - Data transformation, integration and publishing
 - Security
 - Supporting technologies

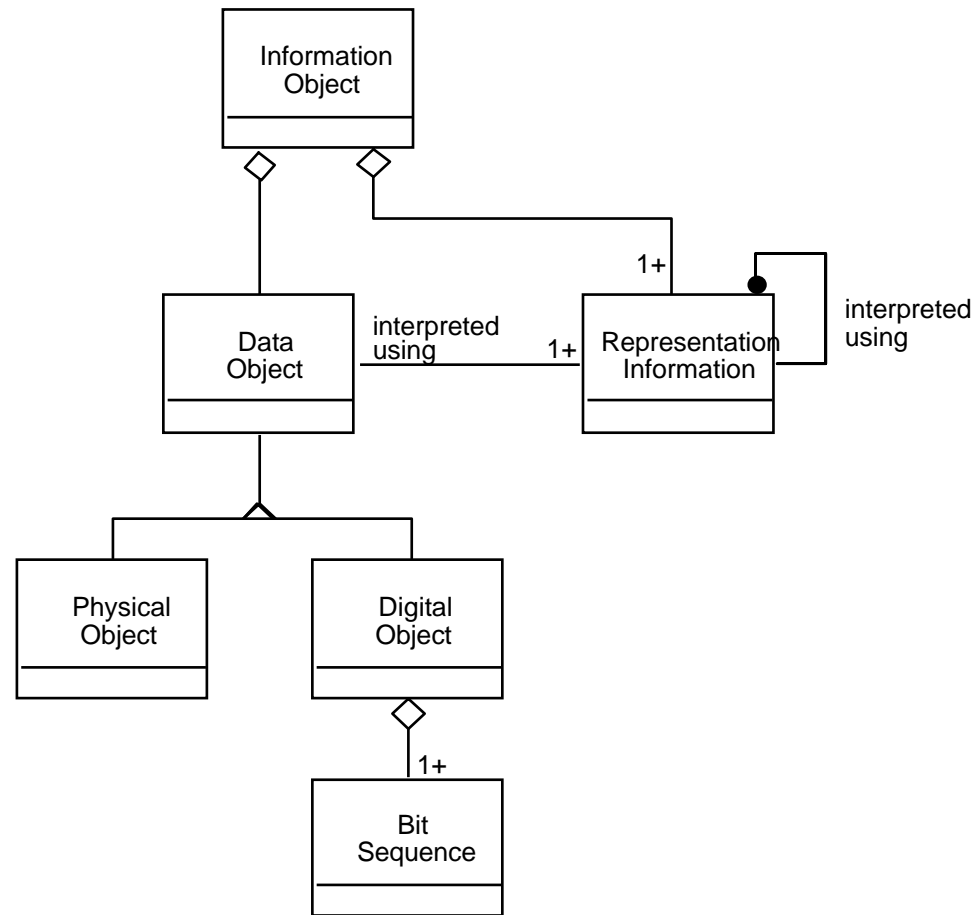
DCC: Development

- Led by Dr. David Giaretta, Head of CCLRC's Astronomical Systems and Services Group
- “DCC Approach to Digital Curation” white paper sets out the path for development activities based on the Reference Model for an Open Archival Information System (OAIS); ISO standard, 14721:
 - Monitoring international standards
 - Development of a Representation Information (RI) registry/repository (DCC-RR)
 - Recommendations for tools and methods for generating Representation Information
 - Creating test-beds for digital curation tools
 - Creating auditing and certification processes for trusted repositories

OAIS Functional Model



OAIS: Information Object

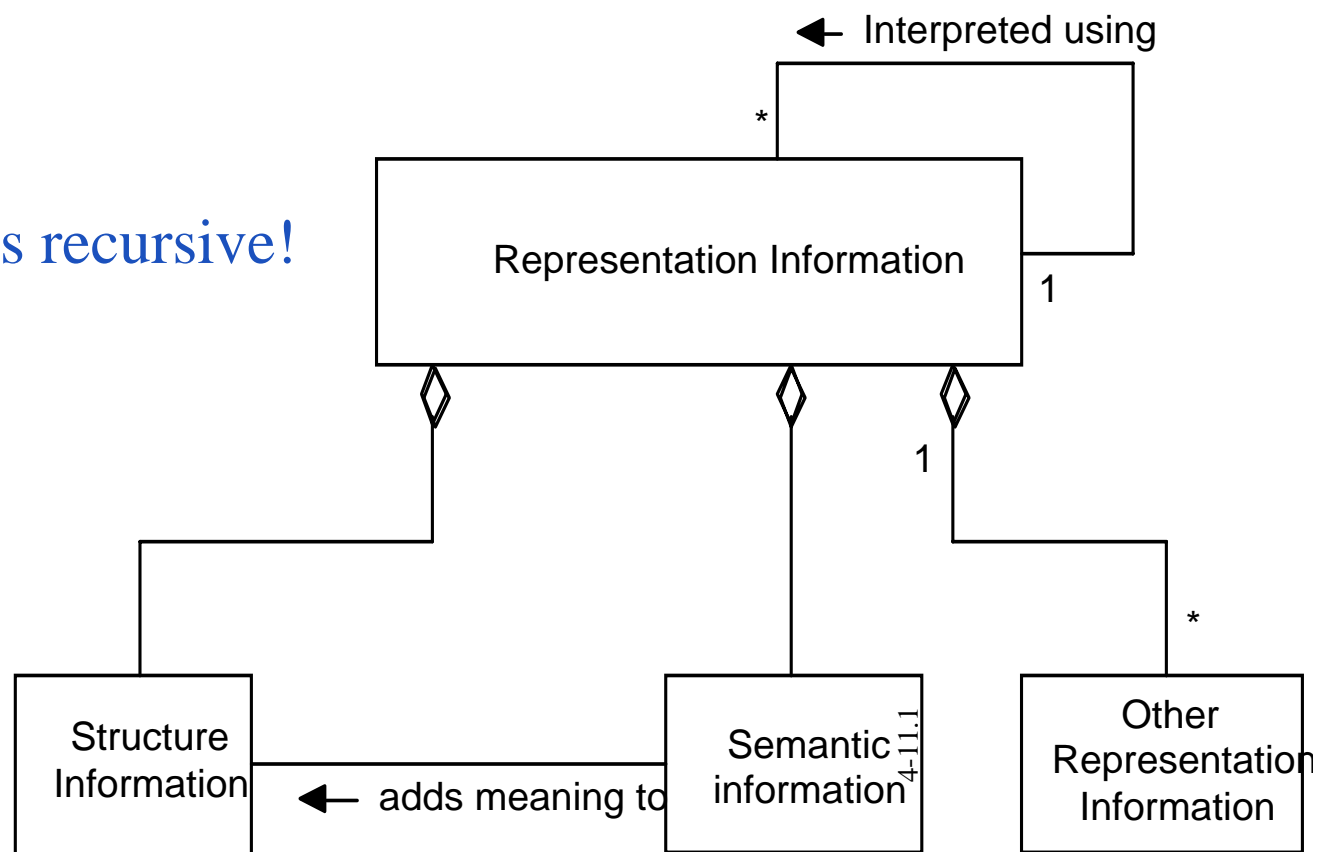


DCC-Dev: Agenda

- Representation Information (RI) is the key to long-term access
- RI is any information required to **interpret and understand digital data** (includes file formats, software, standards, semantic information etc.)
- DCC-RR should be OAIS compliant
- Emphasis on interoperability and automated use
- Vision is to have a global, distributed network of RI
- Investigate how RI fits into the work of other projects and initiatives

OAIS: Representation Network

RI is recursive!



30



DCC-Dev: DCC-RR Overview

- Use concepts and terminology from OAIS Model
- Prototype implemented using MySQL and PHP scripts
 - Register and view RI
 - Create RI Label
 - Add a classification entry
- More robust long-term version based on JAXR clients and ebXML server under development
- Classification scheme for RI under development
- RI label –connects RI to an information object

DCC-RR: A basis for collaboration

- Idea of RI is the key
 - Information Object: a specific object to be archived
 - RI: all information required to interpret and render the object
 - RI Label: used to connect RI to an Information Object
- Need to see how RI and RI label might fit into work of other projects and architectures
- RI label serves as a mechanism for accessing RI in the DCC-RR
 - A label attached to each digital object
 - Label should identify RI (persistent identifier)
 - Provides mechanism for combining individual RI components
 - May be a structured digital object itself (to cope with packaging of multiple objects)

32

DCC: Collaboration

- Digital Library Federation
- The National Archives
- Global Grid Forum
- NARA
- Library of Congress
- Research Libraries Group
- Digital Preservation Coalition
- JISC community
- E-Science Community
- Associates Network
- METS
- LOCKSS
- DSpace
- PANIC
- Global Digital Formats Registry
- Storage Resource Broker
- ...and many more

Concluding Comments

- Growing awareness at **governmental, national, regional, local and project levels** of the need for digital curation
- Realisation that the demands of digital curation can only be met through **collaboration and development of infrastructures**
- Acknowledgement that digital curation requires **long-term funding and commitment**
- Recognition that digital curation needs to be **built into projects from the outset**
- Developments in UK influenced by **international arena**
 - US: OCLC, RLG, DLF; GDFR, SRB, DAITSS, LOCKSS, PREMIS etc.
 - Australia: NLA; PANDORA, PADI, PANIC, APSR etc.
 - Europe: DELOS, ERPANET; DINI/NESTOR, IBM-DIAS (KB) etc.
 - China: Chinese National Library; China Digital Library etc.



First DCC Conference

- Location: Bath, UK
- 29-30th Sept. 2005
- Keynote Speakers
 - Cliff Lynch (CNI)
 - Graham Cameron (European Bio-informatics Institute)
- More information:
<http://www.dcc.ac.uk/>



Reception at the Roman Baths

Digital | Curation | Centre

Thank you for your attention!

Manjula Patel

m.patel@ukoln.ac.uk

<http://www.ukoln.ac.uk/>

<http://www.dcc.ac.uk/>