


ELECTRONIC RESOURCE PRESERVATION AND ACCESS NETWORK 

Digital Curation—Visions from ERPANET and the UK Digital Curation Centre (DCC)

LUCAS, University of Liverpool
Liverpool, 28 February 2005

Dr Seamus Ross
Director HATII & ERPANET,
Associate Director Digital Curation Centre, &
Professor Humanities Informatics and Digital Curation, UofGlasgow

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool



Humanities Advanced Technology and Information Institute (HATII)



George Service House, HATII's home

- Undergraduate & Postgraduate Teaching
- Systems & Laboratory Development and Management
- Research (e.g. digital preservation studies, ICT in the heritage sector, virtual reality, evaluation studies)
- Consultancy
- Summer Schools, Conferences & Workshops

<http://www.hatii.arts.gla.ac.uk/>

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool

Humanities Advanced Technology and Information Institute (HATII)

- Example Research and Support Activities
 - ERPANET
 - <http://www.erpanet.org>
 - DigiCULT
 - <http://www.digicult.info>
 - Digital Curation Centre
 - <http://www.dcc.ac.uk>
 - DELOS—Digital Preservation Cluster
 - <http://www.dpc.delos.info>



George Service House, HATII's home

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool

Eight Questions

- Why should we preserve digital materials?
- Who should do it?
- What should they do?
- Where should they do it?
- When should it be done?
- How should it be done?
- For how long should we do it?
- How should we fund preservation?

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool

Objective of digital preservation

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



© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool

Making the Information & Knowledge Environment Work

If the national and global network infrastructures are to provide a suitable environment for education, e-business, e-government, and sustained public enjoyment then systems must be in place to guarantee that the requirements for: integrity, authenticity, reliability and the archiving of digital information can be carried out easily and effectively.

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool

Key Preservation Issues

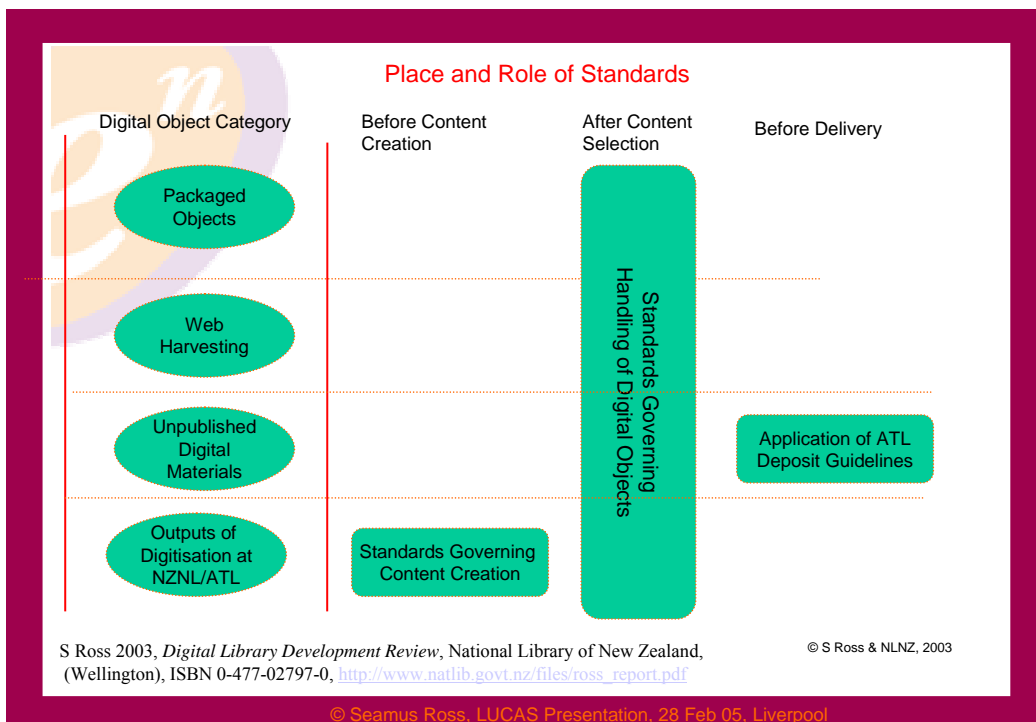
- **Medium**
 - storage media naturally decay or become obsolete
- **Technological (e.g. hardware/software)**
 - hardware and software obsolescence makes data/information inaccessible
- **Intellectual**
 - validation of integrity and authenticity
- **Contextual**
 - avoid loss of meaning with metadata
- **Legal Impediments**
- **The organisation and its staff**



© HATTI, UniGlasgow, 2005


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

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool



What's Happening?

- Collaborative Projects—
 - For example, Pittsburgh, InterPARES, NEDLIB, CEDARS/CAMiLEON, PrestoSpace,
- National Initiatives led by Libraries & Archives
 - Such as UK (NDAD), NARA, BL, Library of Congress, BNF, NLNZ, Danish State Library,
- Gap between commercial activity and the knowledge in the public sector about these activities
- Legal Challenges to digital preservation (IPR, privacy) but FOI & e-citizenship (or e-government) may provide counterbalances
- Increased awareness
 - Recognition that our cultural memory is at risk and that it is composed of many types of digital objects (e.g. audio, VR)
- Institutional Missions often without synergy of effort
- Emerging Concepts and Standards:
 - OAIS
 - Trusted Repositories



© HATTI LukGiesgow, 2002


Margaret Hedstrom (University of Michigan) & Birte Christensen-Dalsgaard, Staatsbibliothek Denmark at Delos Preservation Research Workgroup Meeting in Washington DC, Nov 2002.

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool

Recurring Value of Digital Objects

- Industry dependent
- Evidential value
- Product liability
- Competitive advantage
- Recurring value through reuse
- Costs of re-creation vs storage
- Accountability
- Corporate memory
- Foundation for scholarly endeavour

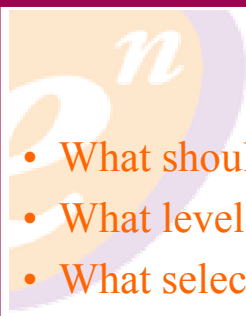
© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool



Difficulties Facing Creators and Users

- What information should be retained?
- Where should it be stored?
- What about the diversity of document types?
- How do I access it if I need it?
- How long should it be kept?
- What is its value?
- What are the costs and justifications?
- Does record creation equal retention?


© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool



Preservation questions

- What should be archived?
- What levels of documentation will be required?
- What selection criteria should be used?
- What standards should be used?
- Who pays? Who uses? Who appraises? When do they appraise?
- Medium, environment, context, integrity


© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool



Sample Organisational Obstacles to Preservation

- Tendency towards decentralisation & networked organisational structures
- Lack of collaboration between creators, business managers, records managers, and IT staff
- Need to link records management strategies with organisational objectives
- Lack of organizational commitment (social, economic, political)
- Failure to link Preservation to corporate objs.
- Failure to acknowledge the investment needed
- Failure to identify recognizable benefits

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool



Information Risks

- Uncontrolled growth in data, documents, records, & resources
- Possibility of accidental loss
- Security (e.g. information leakage)
- Record duplication and authentication
- Unauthorised modification of records
- Loss of integrity and authenticity of digital resources

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool

Appraisal and Selection

- Administrative value
- Evidential value
- Informational value
- Reusability & Integration
- Technical viability
- Anticipated costs of preservation
- Usage restrictions



Hans Hofman (Dutch National Archives) and Charles Dollar at ICA2004 Wien.

© ERPANET, 2005

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool

Preservation Options

- Hardware and software preservation
 - technically complex and expensive
- Conversion
- Media Refreshing
- Data/Document/Record Migration
 - can lead to data and information loss
 - can lead to loss of functionality
 - how can we define acceptable loss
- Software & Hardware emulation
 - Practical at what scale
- Virtual Machines
- Binary Retargetable Code
 - Transmogrifying Adaptable Preservation (TAP)
- Digital Archaeology



Hard disk undergoing one stage of forensic analysis as part of data recover planning at the Tunstall and Tunstall (Nepean, Ontario, Canada) in 2003.

© ERPANET, 2005

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool

Migration strategies

- Sequence of tasks undertaken periodically
- Change media
- Converting format or encapsulating
- Incorporation standards
- Time & labour dependencies
- Costs vs. value
- Influenced by processes, systems, & best practice

See for example, El Archivo General de Indias
<http://www.mcu.es/lab/archivos/index.html>



© ERPANET, 2005

Julian Bescos, technologist who led migration for Archivo General de Indias Project, speaking at the ERPANET Seminar in Budapest, October 2004

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool

Preparation strategies


- Design systems for long term accessibility (e.g work with developers)
- Control, monitor, document, and audit migration
- Avoid proprietary systems (e.g. hardware, software, applications, standards)
- Avoid emerging technologies



© ERPANET, 2005

Hard disk undergoing forensic analysis as part of data recover planning at the *Tunstall and Tunstall* (Nepean, Ontario, Canada) in 2003.

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool



ERPANET

- raise awareness of preservation needs and strategies;
- appraise and evaluate information sources and developments in digital preservation and increase the accessibility of this work;
- provide an EU-wide enquiry and advisory service on preservation issues, practice and technology;
- organise 20 developmental seminars and workshops across EU (so far 37 leading organisations and more than 175 professionals contributed material)
- stimulate research in digital preservation;
- promote awareness among software producers of the preservation needs of the user community; and
- research to improve method and practice.

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool




Partner Institutions:

- The Humanities Advanced Technology and Information Institute (HATII)
- Institute for Archival and Library Science (Università degli Studi di Urbino)
- Nationaal Archief van Nederland
- Schweizerisches Bundesarchiv


				
Niklaus Bütikofer	Maria Guercio	Hans Hofman	Seamus Ross	Peter McKinney

+ Content Editors based at these institutions

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool



General Findings



- Off-the-shelf policy statements
- Business cases & strategies
- Digestible guidance on technologies and their preservation implications
- Improved models (reference, costs, standards, functional requirements)
- Simple Guidelines on digital survival
- Guidance on creating data repositories
- IPR support and guidance

Maria Guercio and John McDonald at the *Preservation Policy and Procedures* co-sponsored with the Archives de France – Le Centre des Archives Contemporaines, Fontainebleau, January 2003.

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool



OBJECTIVES

Thematic Workshops & Seminars

- Exploring core topics with experts and specialised institutions
- Increasing the level of awareness within the communities of practice and need
- Developing recommendations
- Providing documentation of new research directions and results
- Provide skills development opportunities to the community
- Creation of authoritative sources of information



Speakers at the Lisbon Seminar on *The Selection, Appraisal and Retention of Digital Scientific Data*, Weber Amaral, William Anderson, Terry Eastwood, John Faundeen, Pedro Fernandes, Luigi Fusco, Francoise Genova, Myron Gutmann, Gail Hodge, Jürgen Knobloch, Meredith Lane, Seamus Ross, Kevin Schürer, Alex Szalay, Peter Weiss

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool



ERPANET Content

- Literature Evaluation Mechanisms
- Commentaries & Assessments
- Case Study Instruments
- Case Study Reports
- Evaluation of erpaEvents
- Workshop and Seminar Reports
- Materials from all Workshops & recordings of most




© ERPANET, 2005

Alex Szalay, The Johns Hopkins University (USA) speaking at the ERPANET Lisbon Seminar on *The Selection, Appraisal and Retention of Digital Scientific Data*, December 2004.


<http://www.erpanet.org>

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool



erpaStudies – Insights

- Build a picture of current methods matched against their context of use underpin the definition of best practices;
- Accumulate and make accessible information about how approaches to digital longevity;
- Produce materials for seminars and workshops;
- Enable comparisons of approaches across sectors;
- Enable the development of assessment tools; and,
- Identify issues for further research.



© ERPANET, 2005


Stuart Weibel of OCLC at the grave of George Boole following the ERPANET seminar on *Persistent Identifiers*, Cork, June 2004

See for example: S Ross, M Greenan, and P McKinney, in press 2004, 'Digital Preservation Strategies: The Initial Outcomes of the ERPANET Case Studies' in the *Preservation of Electronic Records: New Knowledge and Decision-making*, (Ottawa: Canadian Conservation Institute). (a pre-print is available at the ERPANET Website).

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool

Case Studies Parameters

- Case studies in public and private sectors
- Aim to achieve spread across organisational size and location
- Diversity of organisational type, activity, regulatory framework, and culture
- Perception across records and resources
- Outcomes:
 - Vertical View: within sectors
 - Horizontal View: across sectors



© ERPANET 2005

Mr Allemann, Trivadis AG Switzerland co-sponsor of the ERPANET Bern Workshop on *Long term preservation of databases*, April 2003.

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool

Case Study Practices

- Research by questionnaire and interview (normally by telephone)
 - A team of six research assistants in four countries
- Examined how other projects had conducted surveys to inform our own design
 - The Pittsburgh Project and InterPARES
- Designed focused template which we hope other projects will adopt when conducting similar studies
- Interviews with three staff members sought: Business Manager, IT Manager, Archivist/Records Manager

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool



Our Sample

- Just over 500 organisations contacted
- Current participation rate: 15.6%
- Many initially said 'yes' or 'may be' and then 'no'
 - “currently developing a digital preservation programme and at an early stage”
 - “no time and capacity for participation”
 - “too early to participate”
 - Pulled out when seeing questionnaire: common excuse, no one able to complete form
 - Dropped out after reading report (e.g. BNF, International Court of Justice)—occasionally on legal advice

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool



Participating Organisations

- 1st Round Participants (20)
 - Astra Zeneca PLC, Aventis Pharma Germany GmbH, Bayer AG, Böhlinger Ingelheim, GlaxoSmithKline plc (GSK), Organon NV, Pfizer, RAI (Radiotelevisione Italiana), BBC (British Broadcasting Corporation), ERT (Elliniki Radiophonia Tileorassi SA), SF DRS (Schweizer Fernsehen der Deutschen und Rätoromanischen Schweiz), Gruppo Mondadori (), Hachette Filipacchi Médias, HMSO (Her Majesty's Stationary Office), Kluwer Academic Publishing, Office for Official Publications of the European Communities (Publications Office), Oxford University Press, Random House, Swisscom, Orange
- 2nd & 3rd Round Participants (58)
 - A European Banking House, Amnesty International, ARPAT - Agenzia Regionale per la Protezione Ambientale della Toscana, Banca d'Italia, Belastingdienst, Bio-Inspecta, Biblioteca Nazionale Centrale di Firenze (BNCF), BT Archives, Canadian Institute for Historical, Microreproductions - CIHM, Centraal Bureau voor de Statistiek, CINECA, Comune di Pesaro, Council of Europe, Deutsche Presse-Agentur (DPA), Dutch Kadaster (Land Registry), Edinburgh University Library, ENEL, Electronic Records Management Training at the European Investment Bank (EIB), EMI Music, Engineering Simulation, European Investment Bank (EIB), European Patent Office, Europol, Ferrovie Dello Stato, FIFA, Infocamere, International Labour Organisation (ILO), ISSN International Centre, Kernkraftwerk Leibstadt (KKL), Koninklijk Nederland Meteorologische Instituut (KNMI), Legacoop, Marine Accident Investigation Branch (MAIB), Meteorological Office UK, National Archives of Scotland, National Centre for Public Administration Informatics (CNPIA), National Library of Wales, Nederlandse Aardolie Maatschappij (NAM), Netherlands Historical Data Archive (NHDA), OECD, Office for Metropolitan Architecture, ORF Radio Archive, Portable Antiquities Scheme (PAS), Project Gutenberg, Provincia di Pesaro e Urbino, Regione Emilia Romagna, Schweizerische Depeschagentur (SDA), Swedish University of Agricultural Sciences, Tate, Tessella Support Services PLC, Theater Instituut Nederland (TIN), Universal Postal Union, World Intellectual Property Organisation (WIPO)

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool

Focus of Investigations

- Perception and awareness of risk posed by digital materials
- Understanding how digital preservation affects the organisation
- Identifying what actions have been taken to prevent loss of digital materials
- Develop an understanding of the way preservation activities are monitored (policies or programmes)
- Understand how organisations plan future preservation requirements (e.g. costs)
- What technologies are in place

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool

Perception and Awareness (1)

- Awareness varied across both public and private sector institutions, e.g.
 - Broadcasters, high, possibly reflecting reuse recognition
 - Drug companies, high, but for economic and compliance reasons
 - Public organisations, aware, but often still paper-based
- Some sectors did not see the benefits or were trying to identify where responsibility should lie (e.g. publishers)



© ERPANET, 2005

Peter Robinson and Manfred Thaller at the ERPANET and the Accademia dei Lincei workshop on *Trusted Repositories for Preserving Cultural Heritage* (Rome, November 2003)

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool

Perception and Awareness (2)

- Drivers to do something:
 - Primary business (National Library of Wales, National Archives of Scotland, oil companies),
 - Reuse (Swiss News Agency, EPO, meteorological institutes),
 - Legal compliance (pharmaceuticals, MAIB, banks, kadaster),
 - Experience with loss of information
- Cultural and historical value given lowest priority

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool

Policy, Strategy and Standards

- Few organisations have developed comprehensive policies yet
- Where there were policies interviewees acknowledged they were not always implemented across organisations
- Standards and guidelines developed by other communities often considered as not attractive
- Few organisations articulate preservation strategies as part of system and software specifications

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool

Selection, Preservation & Storage

- **Selection & Disposal**
 - Awareness of the need to determine retention, but few retention policies; digital records not always included
 - Disposal of digital material often still to be implemented (too young)
 - Small number of organisations decided 'to keep everything' (e.g. to let the technology solve the problem, for revenue)
- **Preservation & Storage**
 - Respondents noted the complexity, diversity and size of objects
 - Proliferation of formats seen as a problem: standardisation
 - Long term perspective often not included
 - Pragmatic approach most popular (e.g. migration on demand, XML)
 - Still reliance on paper, certainly for document-like object types

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool

Costs and Access

- **Digital Preservation Costs**
 - Very limited information on costs, if any
 - Poorly predicted and not part of revenue expenditure
- **Access**
 - Mostly internal use; public access, only if (legally) required (e.g. EPO)
 - Different approaches: intermediaries, inter/-ranet portals, CD's
 - Technical issues (e.g. no central entry point, standard file formats), or procedural issues (e.g. authorisation required)
 - Security and privacy issues

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool



General Findings (1)

- Variation in awareness of risk
- Value of information depends upon recognition of business/organisational dependency, potential of re-use, or risk associated with information
- Responsibility rarely taken at corporate level
- Diversity of strategies, but few have them
- If not explicitly defined in policy, it will be unlikely digital preservation is really achieved
- Activity is fragmentary: practices tend to be *incomplete, ad hoc, and individual*

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool



General Findings (2)

- Recognised need for better sharing of knowledge
- Organisations are often waiting for external solutions
- Collaborative sectoral action rare (broadcasting is an exception), but if, then increased confidence and action
- Selection criteria from the paper world have been transferred to the digital environment
- Preservation and storage poorly understood
- Costs difficult to justify in the corporate environment

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool



General Findings (3)

- Lack of a single point of contact—no one knows who knows or is responsible
- The retention period does not necessarily influence preservation strategy: technically similar
- Many approaches are IT-driven, because of lack of ‘archiving’/preservation advice,
 - Due to lack of influence or knowledge of records manager?
- Profile of records or archives department needs to be raised (professionalism):
 - Level self-esteem and confidence
 - Level of knowledge and expertise

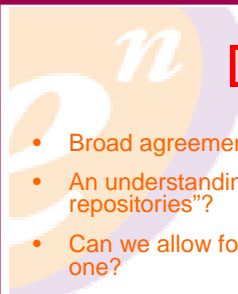
© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool



Objectives Achieved

- Accumulated materials to illustrate methodologies and practices
- Created a basis for establishing practical guidelines
- Raised awareness among more almost 500 institutions and individuals
- Stimulated sharing of information and experiences
- Added objective understanding to a profile which has often been based on single analyses and anecdotal


© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool



Digital Curation Requires

- Broad agreement on the features of a 'trusted digital repository'?
- An understanding of which current implementations can be called "trusted digital repositories"?
- Can we allow for multiple definitions trusted digital repository or can we only have one?
- How are the concepts of reliability, authenticity and trustworthiness interpreted in different contexts and why?
- Mechanisms to represent behaviour and functionality of digital objects in ways that we can test for loss or retention of those characteristics at different stages?
- Guidelines to benchmark acceptable loss?
- Strategies for defining and addressing the roles and responsibilities of different stakeholders?
- An understanding as to how do different communities see trusted digital repositories?
- The establishing and maintaining of trust in user communities of repositories to secure overtime and in the face of changing technologies digital assets.

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool



Preservation Models: OAIS

- OAIS = Open Archival Information Systems
- Key players in development
 - National Space Science Data Centre
 - Consultative Committee for Space Data Systems
- Now an ISO Standard
- Premises Underlying OAIS
 - Irreplaceable nature of much data (e.g. observation)
 - Need to move data and associated metadata across technologies
 - Recognition that representations and formats will change
 - Lack of consensus on adequate metadata standards

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool



Key OAIS Objectives

- Concern about lack of recognised framework for developing digital archive standards
- Address the need for a reference model
- Acknowledges the the hybrid nature of archives
- Do work in collaboration with archival community
- Conducted with a focus on data resulting from space missions
- Model must support near-term and indefinite storage of digital data
- Ensure that the framework is independent of the implementation model
- Model must be archive aware in that it should address the full range of archival processes


© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool



To Build the OAIS Model

- Examined other models
- Defined Data Archiving
- Established key functional areas (FAs) including ingest, storage, access, and preservation
- Set out the interfaces between the Functional areas
- Laid out a set of data classes
- Defined formal representation methods


© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool



OAIS Overview

- Defines how ingest of Information Packages provided by creators should be handled
- Establishes the concept of designated communities
- Reflects needs of identified user community
- Enables preservation in an understandable way
- Uses documented policies and procedures

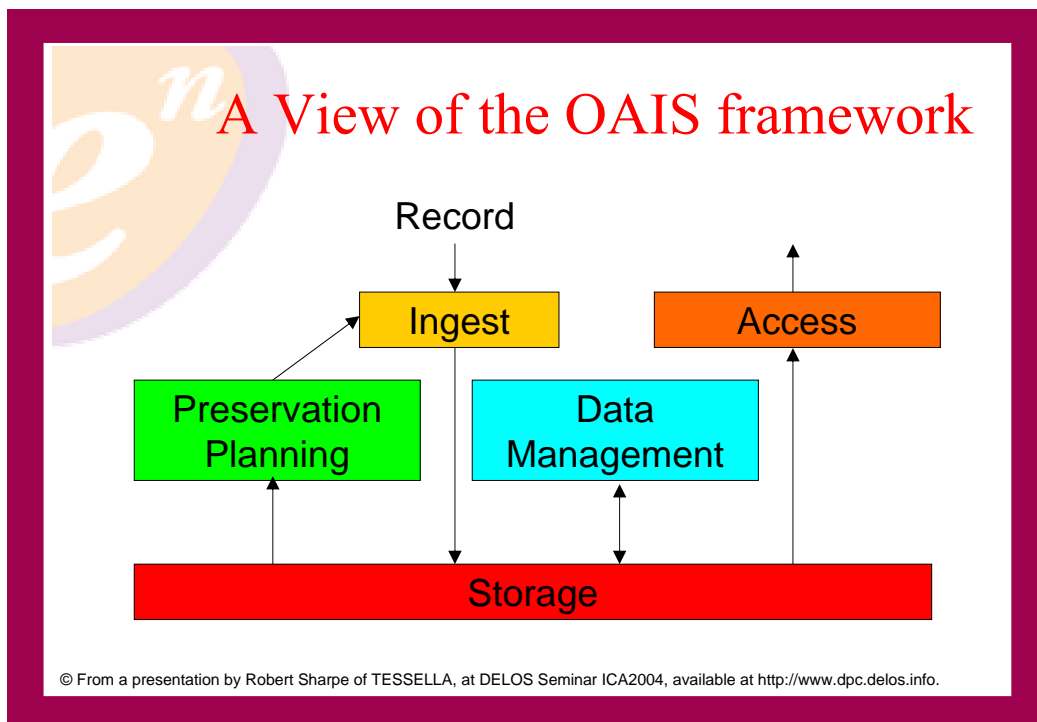
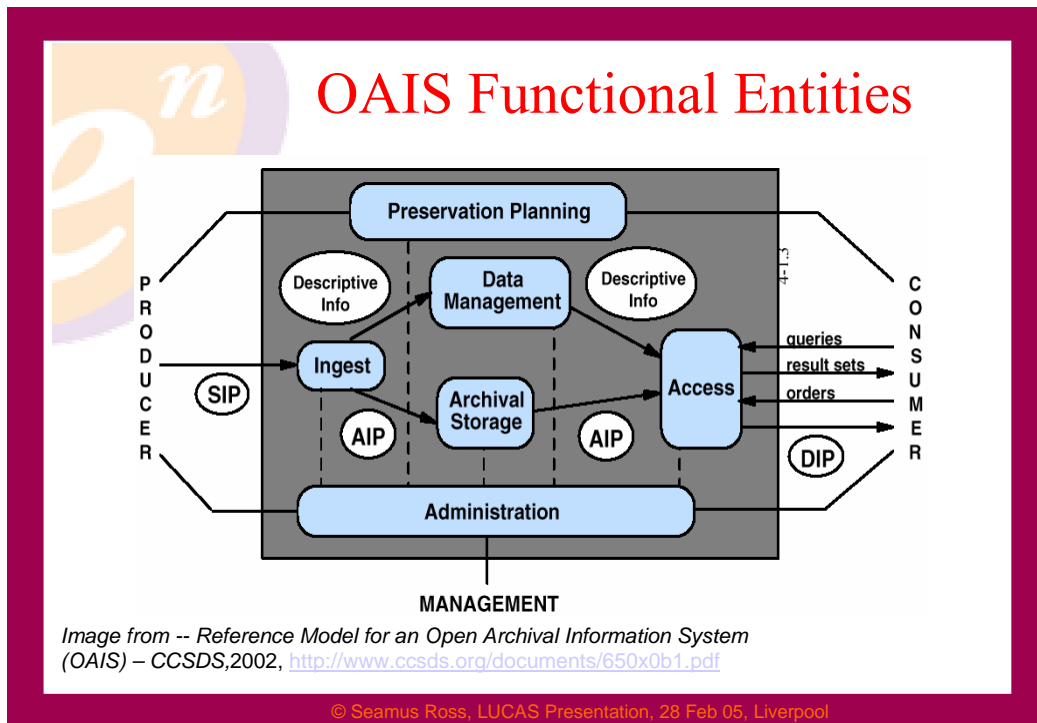
© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool



Advantages of OAIS

- Provides a model where one was lacking
- Facilitates procurement of systems
- Enables interoperability between OAIS compliant systems
- Supports the migration task
- Lays out a minimum set of responsibilities

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool



Who is working with OAIS

- Many communities: Libraries, Archives, science, business, engineering
 - Koninklijke Bibliotheek (KB) through NEDLIB—design and architecture of Deposit System for Electronic publications
 - CEDARS
 - NARA and the San Diego SuperComputer Center
 - National Space Science Data Center
 - Pharmaceutical & Aerospace Industries
 - French Space Agency for its plasma physics archive



© ERPANET, 2005
Participants at the ERPANET Seminar on Metadata for Digital Preservation, Marburg (September 2003)

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool

From Frameworks to Repositories



© ERPANET, 2004
Museums (Vienna)



© ERPANET, 2002
Archives (Wellington NZ)


© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool



Who will make use of Repositories

- Potential depositors
- Potential User communities
 - People
 - Machines
- Managing organisations
- Public authorities

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool




What must a Repository Do

- Handle a wide array of digital media types
- Be Secure
- Guarantee authenticity of the objects it holds
- Protect Integrity (from intended and unintended harm)
- Enable verification
- Ensure stuff ingested into the repository can be output (e.g. be accessible)
- Self-contained (in operation)
 - Must not rely on external infrastructure or services
 - Maintain all documentation in-house
 - Have disaster recovery functionality built-in

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool

Flexible Technical Infrastructure

- Hardware and software elements must be
 - Sustainable
 - Financially
 - Reasonable levels of complexity
 - Management
 - Open
 - Modular
 - Migratable
 - Secure (no unauthorized access)
 - Reliable (no data loss)
 - Available (the data is at hand when you need it)



© ERPANET, 2005

ERPANET Bern Workshop on *Long term preservation of databases* at the Schweizerisches Bundesarchiv, April 2003.

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool

Repositories must be trusted

- Processes:
 - Workflows
 - Operation (management of integrity, authenticity, intelligibility, and accessibility)
 - Automation (e.g. ingest, management, publication)
 - Documentation of procedures
 - Auditability
- Architecture and Implementation



© ERPANET, 2005


Robert Kahn (Corporation for National Research Initiatives) and Tito Orlandi (Centro Linceo Interdisciplinare) at the ERPANET and the Accademia dei Lincei workshop on *Trusted Repositories for Preserving Cultural Heritage* (Rome, November 2003)

RLG/OCLC *Attributes of a Trusted Digital Repository*, 2002, <http://www.rlg.org/longterm/repositories.pdf>

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool

The Trust Problem

- How do we know that our data are secure
- Questions of Trust
 - How is it established?
 - How is it maintained?
 - How is it secured?
 - What happens when it is lost?
 - How can it be verified?



© ERPANET, 2005

Digital Repository Infrastructure, Swiss Federal Archives, Berne, October 2004

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool

Aspects need certification

- People
 - through developing competencies
- data
 - Quality management, policy, validation
- processes
 - OAIS model, IPR, FOI, organisational practices
- managing organisations
 - audit of approaches organisations take to data management

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool



Certification

- Statement of attributes to be measured
- Policies and Assumptions (e.g. practices, environment and security)
- Procedures against standards
- Relationship with depositors
- What processes are in place to manage fidelity checks for ingest
- What metadata processes are in place
- What user needs evaluation work is carried out

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool



Authenticity

- Requires control of ingest and its verification
- Depends on immutability of the data store
- Migration may destroy original byte stream
 - archives and stakeholders must identify significant properties and validate their migration
- Support Audit of the chain of custody, process history, and the descriptions of the migration processes
- Provide mechanisms to enable use



© ERPANET, 2004

Participants at the *File Formats* Seminar in Vienna, May 2004

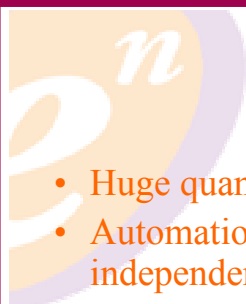
© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool



Repository Operation

- Change will be a feature of repositories
 - Storage technologies
 - Services, close down of some and initiation of others
 - Workflows
 - Verification mechanisms
 - Migration, refreshing, emulation

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool




Automation

- Huge quantities of materials to ingest and manage
- Automation of workflows allow integration of independent services
- Standardized logging/record creation
- Reduce human intervention
 - Cheaper
 - Less error prone
 - Enables higher level of security and reliability
- Intensive test and verification needed
 - Mistakes are very costly (financially but more importantly in terms of trust)

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool

National Archives of Australia

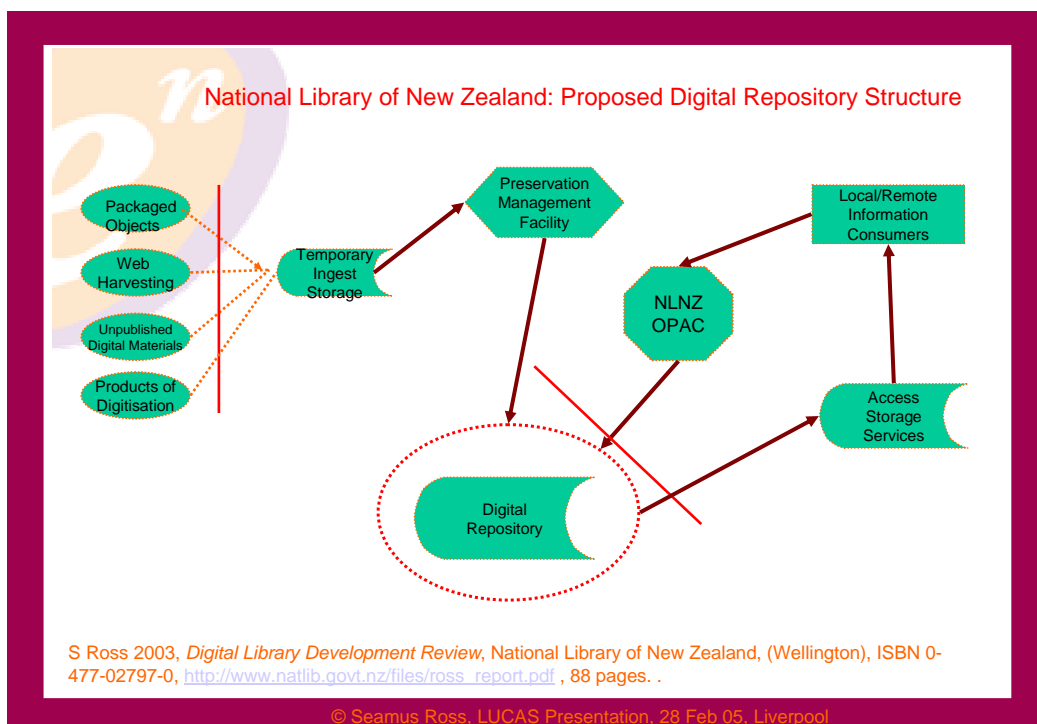
- NAA follows OAIS Framework
- Uses an open source-based solution
- Focuses on use of xml
- Uses open and well-documented data formats
- Not creating a digital archive, but a preservation approach
- Need to address: appraisal/selection, transfer, description, and retrieval and access

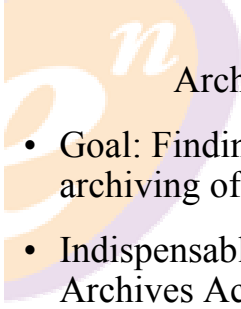


© ERPANET, 2004

Andrew Wilson of the National Archives of Australia speaking at the ERPANET Seminar on *Metadata for Digital Preservation*, Marburg (September 2003).

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool





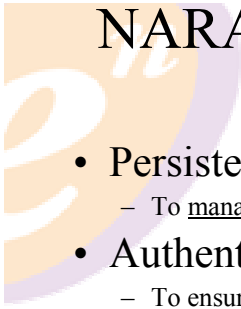
ARELDA

Archiving of Electronic Digital Data and Records

- Goal: Finding long-term solutions for the permanent archiving of digital records in the Swiss Federal Archives
- Indispensable for the long-term execution of the Federal Archives Act
- Development costs 2001 – 2008: ~ 11 Mio €
- Operational costs from 2005: 2.5 – 3.3 Mio € per year (expected growth: 20 TB/yr net)
- Today's project team: 7 people (4 CS engineers)

Thanks to Stefan Heuscher of Swiss Federal Archives © for this slide

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool



NARA's Strategic Response – ERA Requirements

- **Persistent**
 - To manage and access the records over time.
- **Authentic**
 - To ensure that these are the original records
 - Records that are created with attached documentary information
- **Scalable**
 - To grow and adapt to increasing volumes and evolving types of electronic records
 - To serve a variety of user groups (e.g. rich service layer)

Thanks to Fynnette Eaton of NARA © for this slide, which was used at the ERPANET Repositories workshop in Rome 2003. see <http://www.erpanet.org>

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool

Technical Challenges of ERA

- Receives, stores, preserves, and provides access to electronic records, regardless of type, format or media.
- Receives, preserves, and store electronic records in a manner and environment appropriate to their sensitivity level.
- Stores electronic records in a manner that allows for maximum possible independence from specific hardware and software infrastructures.
- Supports high availability.
- Provides viable long-term storage for electronic records.

Thanks to Fynnette Eaton of NARA © for this slide, which was used at the ERPANET Repositories workshop in Rome 2003. see <http://www.erpanet.org>

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool

NARA argues: repositories must

- Find records based on searches of descriptions of records
- Search the electronic records themselves.
- Accurately reproduce and output electronic records.
- Provide certified copies of electronic records.
- Manage requests for review of restricted materials.
- Implement the results of electronic records reviews.
- Enable users to request and receive assistance from archivists or librarians.

Thanks to Fynnette Eaton of NARA © for this slide, which was used at the ERPANET Repositories workshop in Rome 2003. see <http://www.erpanet.org>

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool

Planning for the Unknown

- Define objectives and aims of your repository and those of the services it will provide
- Develop and monitor the application of policies and procedures
- Define senior management steering roles and responsibilities in relationship to repositories
- Ensure that all services, technologies (hardware and software), meet exceptions
- Develop and maintain risk registers
- Status reports and minutes of meetings
- Define, implement, and monitor disaster recovery services



© ERPANET, 2004

Ken Thibodeau of NARA speaking at the ERPANET Berne Seminar, October 2004

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool


Are there tools

- LOCKSS (Lots of Copies Keep Stuff Safe)
<http://lockss.stanford.edu>
- Fedora --**FEDORA (Flexible and Extensible Digital Object and Repository Architecture)**
<http://www.fedora.info/>
- DSPACE--<http://dspace.org/index.html>
- Digital Asset Management Systems ???

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool

Thinking Toward the Future

- Design information for long term accessibility (e.g work with creators)
- Migration must be controlled, monitored, documented, and audited
- Avoid proprietary systems and emerging technologies
- better software
- intelligent record selection & appraisal tools
- mechanisms for maintaining links between business process and records created/used by them
- case studies of the cost-benefits analysis for data loss vs preservation



© ERPANET, 2004

Prof Luciana Duranti of University of British Columbia and Christine Pétillat of Le Centre des Archives Contemporaines at the launch of ERPANET at the DLM Conference in Barcelona (May 2002)

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool

Attentiveness to materials

- Access to digital materials depends on continued attention and maintenance
- Process requires close monitoring and must be controlled
- Interchangability of media, file formats,
- Peripheral devices
- Avoid propriety standards for encoding, software, & hardware

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool



Acquisition of Experience

- Develop test experimental frameworks
- Experiment with ingesting, managing, and providing access digital assets
 - Netherlands: Digital Repository Project
 - US: NARA
- Do something concrete -- gain experience
- Ensure parameters of the research are well-documented so that they can be duplicated
- Aim for 'recipe-like' descriptions of processes
- Digital Curation Centre as a mechanism to support developments in this area.

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool

The Digital Curation Centre

- *The quintessential purpose of the DCC is to support and promote continuing improvement in the quality of data curation and digital preservation activity.*
- *Specifically it..*
 - *Supports preservation of data as an aid to scholarship – reproducibility and reuse*
 - *Enables scholarly communication and eLearning*
- JISC and the e-Science Core Programme funding
 - for development, services, and outreach in digital curation
 - for a research programme
- Impetus to action
 - Growth in e-Science activity and data creation
 - Recognition that continuing access to digital materials crucial



Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool -- © This slide DCC Partners

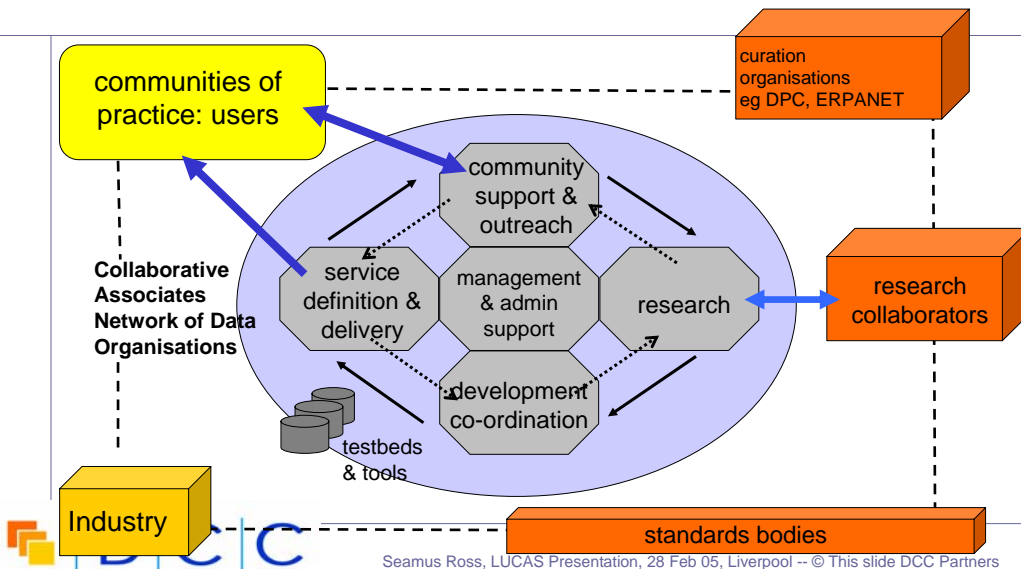
DCC Objectives

- Lead a vibrant research programme
 - addressing the wider issues of digital curation
- Create an active, innovative, and collaborative Associates Network
 - strong links across existing community of practice
 - engagement with curators (individuals & organisations)
- Deliver effective, efficient, and high demand services
 - to be constructed on cutting edge research, solid developments
 - to include the evaluation of tools, methods, standards and policies
 - to include repositories of tools and technical information
- Constructing a 'virtuous circle'
 - built around expertise, experience & requirement feed into the DCC research programme



Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool -- © This slide DCC Partners

functional management & collaboration



Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool -- © This slide DCC Partners

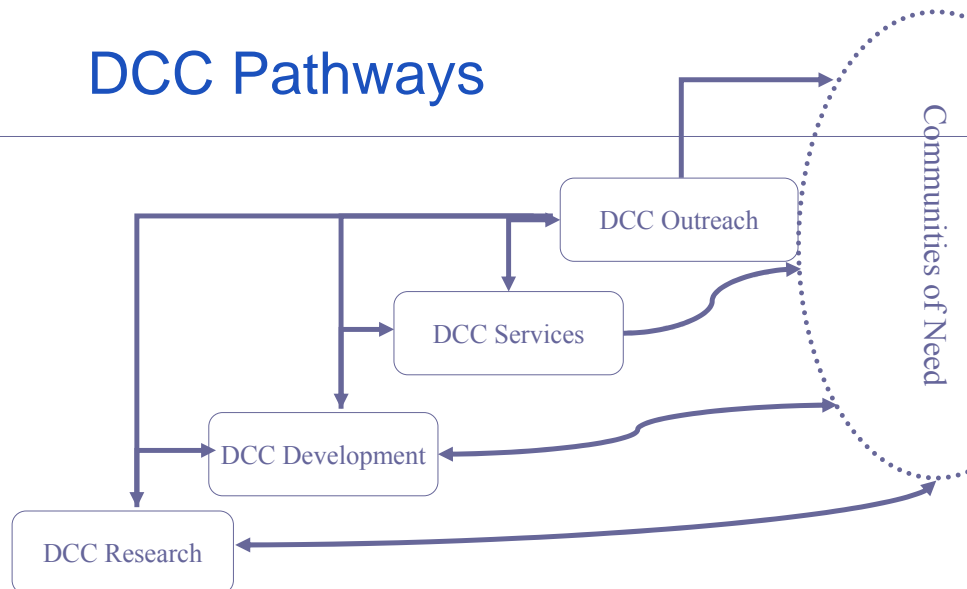
Partners

- University of Edinburgh (lead site):
 - Peter Burnhill, Prof Peter Buneman
- University of Glasgow through HATII:
 - Prof Seamus Ross, Director of HATII & ERPANET
- University of Bath through UKOLN:
 - Dr Liz Lyon, Director of UKOLN
- Council for the Central Laboratory of the Research Councils (CCLRC):
 - Dr David Giarretta
- Other key figures include --
 - Prof Malcolm Atkinson, Director, NeSC
 - Chris Rusbridge, Director, Information Services, U of Glasgow



Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool -- © This slide DCC Partners

DCC Pathways



Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool -- © This slide DCC Partners

Research Goals

- Bringing strands of curation, including
 - traditional archival functions
 - maintenance and publication of evolving knowledge as seen in scientific databases
- to identify key challenges and to deliver research results that help address them
- to conduct research in areas crucial to digital curation
- to interlock research activity and service provision
 - Drawn practical challenges to attention of researchers
 - Test research outputs in real world situations



Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool -- © This slide DCC Partners

Research

- Research
 - Annotation
 - Data integration and publication
 - Appraisal and long-term preservation
 - Socio-economic & legal context
 - rights, responsibilities and viability
 - Performance and Optimisation
 - Semantic metadata extraction
- Problems that:
 - involve fundamental problems
 - will generate highly visible research
 - have components that can be almost immediately carried into development and testing
 - interest to partners and to associates and clients.



Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool -- © This slide DCC Partners

Development Objectives

- Development planning & monitoring
 - Plan and organise Development thread of the DCC
- Registry/Repository development
 - Define and create registry and repository for metadata
- Representation Information development
 - Development of recommendations for tools and methods for generating Representation Information
- Testing and Certification processes
 - Create a foundation for DCC testing and certification work, including running testbeds



Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool -- © This slide DCC Partners

Services Objectives

- Service Planning and Monitoring
 - Plan and organise Services stream of the DCC
- Standards, tools, testbeds, and Repositories
 - Ensure access to standards, tools and repositories of Representation Information
- Advisory Services
 - Provide guidance on curation and preservation
- Audit and Certification
 - Deliver Audit and Certification Services to the UK HE, FE and research Community



Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool -- © This slide DCC Partners

Outreach Objectives

- Outreach planning and monitoring
 - Plan, manage and monitor the Outreach task area of the DCC
- 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.



Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool -- © This slide DCC Partners

Planning & Progress

- Long-term planning essential
- Substantial territory of action
 - multi-disciplinary, multi data type, multi tradition/profession
 - national and international, but also local and hidden from view
 - Increasing amounts of activity
- Focus on community needs
 - Ensure that we maximise outcome for investment
 - who/what should we plan to affect/effect?
 - policy-makers; 'responsible curators'; (researchers?)
- Collaboration is the key—creating a community of actors
- Officially Launched 5 November 2004



Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool -- © This slide DCC Partners



For Resources & Guidance Visit

The ERPANET Website:
<http://www.erpanet.org>

&

The DCC Website:
<http://www.dcc.ac.uk>

© Seamus Ross, LUCAS Presentation, 28 Feb 05, Liverpool