

Creating a pragmatic pan-European framework for permanent access to the records of science

Dr. Peter Tindemans
Chairman Task Force Permanent Access

Peter Tindemans, Nottingham 06-12-05

1

Summary

- 3 ICT Infrastructures: Networks, High Performance Computing, GRIDs
- 4th one will affect science as profoundly: an 'infrastructure' to provide long-term preservation of and access ("P&A") to Records of Science (and digital heritage in general)
- How does this look like?
- How should we build it?
- Major stakeholders from science, libraries, archives offer their strategic commitment to national governments and the EU to create in 3-5 years sufficient momentum

Peter Tindemans, Nottingham 06-12-05

2

Overview

1. Background: including Records of Science in Digital Heritage: Task Force Permanent Access.
2. The problem and range of technical solutions required
3. High-level, strategic, pragmatic approach: need and essence
4. How does 'European Digital Infrastructure for Preservation of and Access to Records of Science' look like?
5. Alliance in the making
6. Financial challenge

Peter Tindemans, Nottingham 06-12-05

3

1. Background

- Inclusion of records of science in digital cultural heritage evolved from 'records of history of science' to 'records of science in operation'; this concerns not just scope, but also nature of records: 'data' next to 'documents' (and other cultural physical artefacts)
- Particular culmination point EU Conference "Permanent Access to the Records of Science" (National Library of Netherlands KB, Netherlands EU Presidency), 1st November, 2004, The Hague.
- Participants agreed to need to create European infrastructure for long-term preservation to and permanent access to records of science. KB urged to create a Task Force.

Peter Tindemans, Nottingham 06-12-05

4

Composition Task Force

- Bertil Andersson, Chief Executive European Science Foundation;
- Lynne Brindley, Chief Executive The British Library;
- Wim van Drimmelen, Director General Koninklijke Bibliotheek;
- Norbert Kroo, Secretary-General Hungarian Academy of Sciences;
- Wolfried Stucky, professor Institute of Applied Informatics and Format Description Methods, Karlsruhe University, curator Max Planck Institute of Computer Science, Germany;
- Malcolm Read, Executive Secretary Joint Information Systems Committee, UK;
- Vincenzo Beruti, ESA/ESRIN;
- John Wood, Chief Executive Council for the Central Laboratory of the Research Councils, UK;
- Peter Hendriks, Board Springer Science and Business Media, Executive Board International Association of Scientific, Technical and Medical Publishers.
- Tomas Lidman, Director General The National Archives of Sweden;
- Peter Tindemans, chair, on behalf of the Koninklijke Bibliotheek. (reflects 'data' and 'documents')

Peter Tindemans, Nottingham 06-12-05

5

2. Problem

Science-angle

- Individual scientist
 - Maintaining and accessing databases built up by individuals: e.g. Madison database on GDP
 - Requirements of journals and funders with regard to supplementary or original data
 - New cultural paradigm challenges individuals, universities, funders, etc.
- Large research organisations and communities (CERN, ESA,...): volume of data
- European social sciences data archives

Libraries- and archives angle ("perennial storage")

- acidification threatened paper; obsolescence and volume explosion jeopardise digital heritage

Peter Tindemans, Nottingham 06-12-05

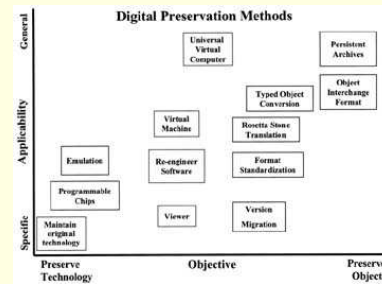
6

Range of solutions required: RDD programme

- Storing Petabytes to 100s of Petabytes to Exabytes, surviving changes in hardware and software technologies, retrieving information
- Standardised approach to describe information (metadata) and management of information as successive 'virtualisation' layers (hardware, data, knowledge, workflows, trust, management) to enable fully automated, distributed solutions
- Complex dynamic datasets and databases
- Legal solutions for digital access and rights management
- Economic business models based on value-chain analysis and public-good aspects
- Technical tools, e.g. to overcome 'museum of old ICT technologies', and cumbersome migration

7

Digital preservation methods suggested (Thibodeau, 2002)



8

3. High-level, strategic, pragmatic approach: need

- Increasing awareness among experts, sometimes institutions about size and complexity of problem of P&A.
- Many projects, standards, good practices, etc.

But:

- No recognition in Europe that preserving and making accessible the digital heritage on very long time scales is strategic issue for
 - Organisations (with few exceptions)
 - Governments,
 - as well as many private sector parties
 - ➔ no financing mechanism
- In USA since 2002 National Digital Information Infrastructure and Preservation Program: Library of Congress working together with NSF, research libraries, archives etc; 100 M\$ to start with

Peter Tindemans, Nottingham 06-12-05

9

High-level approach: essence

1. Make 'digital heritage' **stakeholders understand at 'board level'**
 - economic and cultural importance of P&A for their strategic development
2. Involve **public and private parties**: essential to find business model based on
 - private and user interests and cost allocations,
 - public infrastructure: important 'public good' aspect.
3. Adopt **non-technical 'model of world' as basis for the 'infrastructure'**
4. Adopt **practical way ahead**
 - a. Where is highest impact possible?
 - b. Involve initially not too few, not too many stakeholders
 - c. Connect to ongoing activities: don't replace, but integrate responsibilities

Peter Tindemans, Nottingham 06-12-05

10

a. Highest impact: focus on Records of Science, taken in broad sense

Two worlds

1. Cultural heritage
 - Begins to include digital heritage
 - UNESCO; 'memory institutions': archives, deposit libraries, museums.
 - Science = 'records of history of science'.
 - Politically increasingly visible: UNESCO, EU
2. Records of science.
 - S&T in digital age: 'data' next to 'documents'; small part to spill into traditional archives.
 - 'Science' = S + T; NSE+HMS+SSH; Large scale data collection for operational services and science (meteorology, GIS, census, ...); experiments + observations + simulations + surveys + census and poll + history records.

Gearing up best done by focusing on 'Records of Science'

- Greatest momentum:
 - Inherent needs of scientific community and organisations
 - High 'specific mass' (including financial mass)
- Covers broad field
 - Academic and deposit libraries, scientific publishers straddle two worlds.
 - Archives linked to e.g. historical, social and economic sciences.

Peter Tindemans, Nottingham 06-12-05

11

b. Not too few, not too many

- Common European approaches:
 - 'call for tender' for projects,
 - All-inclusive approach: all stakeholders from 25 member states plus Commission, resolutions, communications, agency,
- Instead focus on critical mass of stakeholders and focused action, i.e.
 - Emphasis on preservation (though preservation cannot be separated from building digital collections)
 - Aim to create 'infrastructure'
 - Aim to create growing consensus among and conditions for 'communities' and organisations and their particular preservation projects.

Peter Tindemans, Nottingham 06-12-05

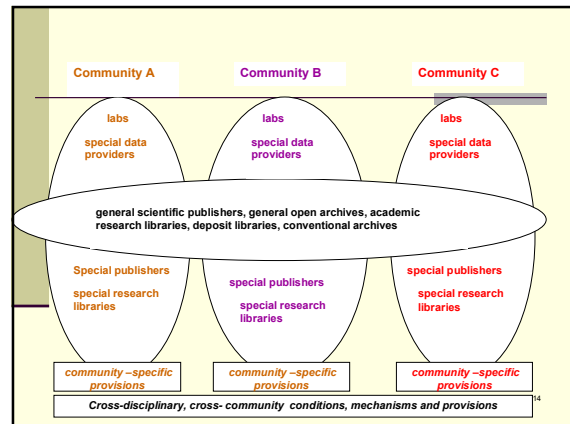
12

4. Model of the world → Framework of conditions and rules of conduct

- 'Communities' produce science (particle physics, social sciences, ...) are different, but have similar structural elements to house "Record of Science"
 - In some disciplines short-term role individual researchers
 - 'Laboratories'
 - Specialised data providers
 - Specialised publishers or web-based archives
 - Specialised research libraries
- Cross-cutting horizontal structure too exists:
 - Scientific publishers, multidisciplinary open archives
 - Academic research libraries
 - Deposit libraries
 - Conventional archives
- All are **digital archives or repositories** in digital world

Peter Tindemans, Nottingham 06-12-05

13



Transform into framework ('infrastructure') of real life organisations and operating conditions (for interoperability and collaboration)

1. Identify set of **core physical digital archives** in limited number of initial communities, and in horizontal layer ("critical mass" and 'high specific mass' are essential criteria)
2. These must be **OAIS-compliant** to ensure proper archiving, interoperability and long-term preservation
3. Framework for **metadata**, Framework for **persistent identifiers**, and number of **registries**
4. Cost-effective **preservation methods and services** must be available
5. Common framework of **principles and guidelines for management of access and rights** (underlying the technical tools to implement this framework)
6. **Financial mechanism** for developing and testing implementation tools, techniques and services
7. a. **Certification service providers**, accredited according to b. Common **European accreditation mechanism**.

Peter Tindemans, Nottingham 06-12-05

15

5. An Alliance in the making

Aims

- Establish wide consensus on framework ('infrastructure') for LTPA; initial focus on science
- Accelerate significantly creation of its main building blocks
- Work with national governments and EU to strengthen European strategies, policies and their implementation
- Strengthen role European parties world-wide
- Articulate and maintain ongoing R&D&D programme

3-5 years

A "Rolling Stone"

Peter Tindemans, Nottingham 06-12-05

16

Tasks

- Assisting communities: initial core set and others
- Enhancing and consolidating consensus on the building blocks of the 'infrastructure'
- Helping establish European funding mechanism
- Helping establish European accreditation mechanism
- Liaising with national governments and EU
- Promoting sustainable business models
- Raising awareness: funding bodies, professional societies, universities,

Peter Tindemans, Nottingham 06-12-05

17

Core Alliance Partners

- European Science Foundation
- Some of most active libraries: British Library, KB
- Some major scientific organisations: ESA, CCLRC, Max Planck Gesellschaft; CERN, EMBL, CNRS, CEESDA are among those approached
- Association of Scientific, Technical and Medical Publishers
- Some major national archives
- JISC,
- '**National coalitions**' for P&A, where they exist: UK, Germany,
- Corporate associate members (e.g. ICT industry): 'Customer-contractor' principle

Peter Tindemans, Nottingham 06-12-05

18

Strengthening emerging consensus; Building on what is being done

- **Conceptualisation and standardisation**, e.g.
 - OAIS
 - Dublin Metadata Core Initiative
 - Draft Audit Checklist for Certification of Trusted Digital Repositories (RLG, NARA plus European experts)
- **Practical development and implementation**, e.g.
 - Several EU-funded projects (but too much focus on co-ordination)
 - Strong national projects (but in few countries only)
 - Public-private agreements (e.g. libraries and publishers)
 - Audit and Certification of Digital Archives Project (CRL) to test audit 3 archives

Peter Tindemans, Nottingham 06-12-05

19

6. Financial model

1. Need to create European, but strongly distributed infrastructure;
2. Need to make Europe visible, strong partner in global efforts

Therefore:

- Partners continue current efforts and investments
- Partners contribute to establish small European organisation to co-ordinate Alliance efforts
- '100 M€' for the real action
- Leveraging national funding

Peter Tindemans, Nottingham 06-12-05

20