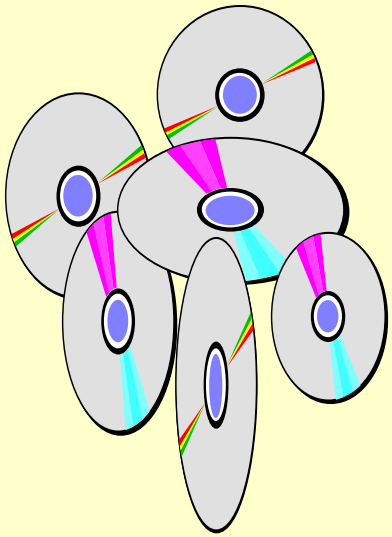
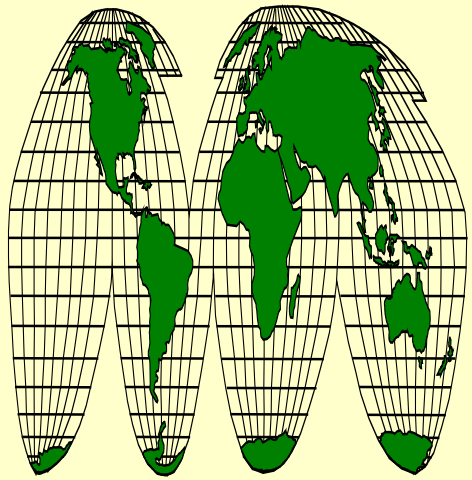
Future Trends and Opportunities



DPC Industry Forum June 2002



Neil Beagrie

Programme Director, JISC London Office

Secretary, Digital Preservation Coalition

Overview

Digital preservation --preservation of digital materials (born digital or surrogates)-not necessarily in perpetuity varying timescales and retention -but through change so more than “archiving”for disaster recovery/business continuity

- **How much information**
- **How much needs to be retained**
- **The US**
- **Europe and UK**
- **The Digital Preservation Coalition**
- **Conclusions**

How Much Information?

http://www.sims.berkeley.edu/research/projects/how-much-info/summary.html

The world produces between 1 and 2 exabytes of unique information per year, which is roughly 250 megabytes for every man, woman, and child on earth. An exabyte is a billion gigabytes, or 10^{18} bytes. Printed documents of all kinds comprise only .003% of the total. Magnetic storage is by far the largest medium for storing information and is the most rapidly growing, with shipped hard drive capacity doubling every year. Magnetic storage is rapidly becoming the universal medium for information storage.

- ``democratization of data." A vast amount of unique information is created and stored by individuals.
- "dominance of digital" content. Not only is digital information production the largest in total, it is also the most rapidly growing. Even today, most textual information is "born digital," and within a few years this will be true for images as well.

How much needs to be retained

- Focus has been on current access and opportunities
- over next 5-10 years current digital trend and growth will continue

- sea change in perception and recognition of challenges
- much of C20th information at risk
- changes in business model

- cultural memory institutions with long-term responsibilities first
- other sectors and private individuals will follow

US NDIIPP

US National Digital Information Infrastructure and Preservation Program
http://www.digitalpreservation.gov/ndiipp/

US NDIIPP

- $175 million program
- initial planning stage of $5 million
- environment scans for: Web sites, electronic journals, electronic books, digitally recorded sound, digital moving images, and digital television.
- Research program

http://www.si.umich.edu/digarch/

US Cyber-infrastructure

- **National Science Foundation Advisory Panel on Cyberinfrastructure Draft Report**

https://worktools.si.umich.edu/workspaces/datkins/001.nsf

> "The importance of data in science and engineering continues on a path of exponential growth; some even assert that the major science driver of high end computing will soon be data…Collecting, organizing, storing, and providing access to vast quantities of data and other information (such as scholarly publications) is becoming as important as simulation has been and will likely grow faster over the next decade."

Europe - EU 6th Framework

http://www.cordis.lu/fp6/eoi-instruments/home.html

http://www.cordis.lu/fp6/eoi-instruments/infosoc.htm

Information Society Technologies

For cultural heritage, the effort will focus on intelligent systems for dynamic access to and preservation of tangible and intangible cultural and scientific resources.

Europe - EU 6th Framework

- Knowledge technologies and digital content:
  The objective is to provide automated solutions for creating and organising virtual knowledge spaces (e.g. collective memories) so as to stimulate radically new content and media services and applications.
- Complex problem solving in science, engineering, businesses and for society: The objective is to develop technologies for harnessing computing and storage resources which are distributed in geographically dispersed locations, and for making them accessible, in a seamless way, for complex problem solving in science, industry, business and society.

e-Science and Research Grid in the UK

‘e-Science is about global collaboration in key areas of science, and the next generation of infrastructure that will enable it.’

£120 million budget for first phase

“all of our systems will be enabled to work with the Grid”

Irving Wladawsky-Berger, IBM

concerns over scientific data curation: continuing access needed medium to long-term

e-Science Examples

- Bioinformatics/Functional genomics
- Collaborative Engineering
- Medical/Healthcare informatics
- Earth Observation Systems
- TeleMicroscopy
- Virtual Observatories
- Robotic Telescopes
- Particle Physics at the LHC

Medicine and Biomedicine

- Key research challenge- ability to archive,integrate, analyse,interrogate, massive data from diverse experiments
- core data resources
  - archiving underlying primary data. Gene sequences 100Tb in 5-10yrs
- new forms of experimental data
  - microscopy image archives - ?Pbytes in 5-10years
- published literature
  - accessibility, connection of literature and data, e-theses, e-prints

Digital Preservation Coalition

- Aim of Coalition is to develop a UK digital preservation agenda within an international context.
- Launch July 2001- 20 full & associate members,+ allies

JISC

THE BRITISH LIBRARY

ULCC

CLRC

re:source The Council for Museums Archives and Libraries

The NATIONAL ARCHIVES of SCOTLAND

curl

The Wellcome Trust

RESEARCH LIBRARIES GROUP

OCLC

DPC and Industry

- Forging strategic alliances with relevant agencies nationally and internationally, and working collaboratively together and with industry and research organisations, to address shared challenges in digital preservation (DPC Goals).
- *Vendor neutrality*: The goals of the Coalition are generic and will be vendor neutral. It will support the development of standards and generic approaches to digital preservation, which can be implemented by a range of hardware, software, and service vendors (DPC Principles).
- *Openness*: The Coalition and its members commit to promoting and disseminating information and sharing outcomes so that we can all learn and benefit as quickly as possible from transferable lessons and experience (DPC Principles).
- Members not for profit associations, exploring how to engage with individual companies

Some conclusions

- preservation is a growing concern and opportunity
- many technological challenges to be addressed
- major gaps in current products and services
- major future role potentially for industry
- today opportunity to start a dialogue between industry and DPC

Further information

- **Digital Preservation Coalition**
  - **http://www.dpconline.org**
- **JISC Digital Preservation Focus**
  - **http://www.jisc.ac.uk/dner/preservation**
- **Digital preservation e-mail list**
  - **http://www.jiscmail.ac.uk/lists/digital-preservation**