

**The Issue of Digital Objects:
Scale, Complexity and Opportunity**

Eileen Fenton
Executive Director, Portico

Curating Research: e-Merging New Roles
and Responsibilities in the European Landscape

April 17, 2009
The Hague, Koninklijke Bibliotheek

Today's Comments

1. Definition of digital preservation and its purpose
2. Characteristics of digital objects
3. Proposed guiding principles for digital preservation
4. Our opportunities

1. What is Digital Preservation?

- Digital preservation is a means to an end: long-term, assured access to important digital resources.
- Digital preservation is the series of management policies and activities necessary to ensure the enduring
 - usability
 - authenticity
 - discoverability and
 - accessibilityof content over the long term.
- Digital content does not remain viable by accident.

2. Characteristics of Digital Objects

- Digital objects are prone to loss,
 - 27 months after publication up to 13% of online cited sources are irretrievable.*
 - Tapes of U.S. census data from 1960's are now inaccessible.
 - The location of NASA's original moon landing recordings is (currently) unknown.

*Dellavalle, Robert P. et. al. "Information Science: Going, Going, Gone."
Science 302, no. 5646 (Oct. 31, 2003), 787-8.

2. Characteristics of Digital Objects

- and are growing exponentially in kind and number.
 - Formats now include: datasets, publications, blogs, micro-blogs, web sites, digitized materials of all types, maps, GIS data, video, audio, ...
 - In 2007, the “digital universe” totaled 281 billion gigabytes.*
 - Information created, captured, or replicated exceeded available storage for the first time in 2007... by 2011 almost half of the digital universe will not have a permanent home.*

*Gantz, John F. et. al. “The Diverse and Exploding Digital Universe.” *An IDC White Paper* sponsored by EMC. (March 2008), 2.

3. Guiding Principles for Digital Preservation

- Befriend selection.
 - Not all digital objects will be (or should be) saved. Curation requires selection.
 - Selection will necessarily be driven by organizational mission or business needs.
 - New selection tools will enable more highly automated selection processes.

3. Guiding Principles for Digital Preservation

- Recognize the demands of diversity and scale.
 - Solutions that work for video may not for GIS data or audio or datasets...
 - The infrastructure and tools appropriate for 1TB of data may not be a fit for 100 TB of data.
 - Scale impact moments or thresholds can be difficult to anticipate.

3. Guiding Principles for Digital Preservation

- One size does not fit all; multiple preservation methodologies and models are needed.
 - Formal publications: may be well suited to format migration.
 - Select dynamic web content: may require emulation tools.
 - Complex databases: may be initially limited to bit capture until satisfactory preservation methodologies are developed.

3. Guiding Principles for Digital Preservation

- Understand cost drivers and minimize, where possible.
 - Ingest costs may be greater than long term preservation costs (~42% vs 23% in one example).*
 - Early intervention with content creators may help to reduce costs.
 - Cost drivers will vary depending upon content type and preservation method.

* Beagrie, Neil, et al. *Keeping Research Data Safe: A Cost Model and Guidance for UK Universities*. JISC-funded study prepared by Charles Beagrie Limited, April 2008, 5.

4. Our Opportunities

- Experiment with new approaches.
 - Dioscuri project
- Share lessons with others.
 - planets project, Digital Preservation Coalition
- Right size the solution to the digital resource to be preserved.
 - Preservation of journals vs. newsletters
- Rely on partners rather than reinventing the wheel at home.
 - Third party archives vs. do-it-yourself

Eileen Fenton
eileen.fenton@portico.org
www.portico.org