

The Challenges of Migration as a Long-term Preservation Strategy: The Findings of TEAM Norway and LongRec



Feng Luan
Norwegian University of Science and Technology

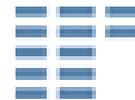
InterPARES 3 Project, 1st International Symposium
Seoul, South Korea
5 June 2009

Agenda

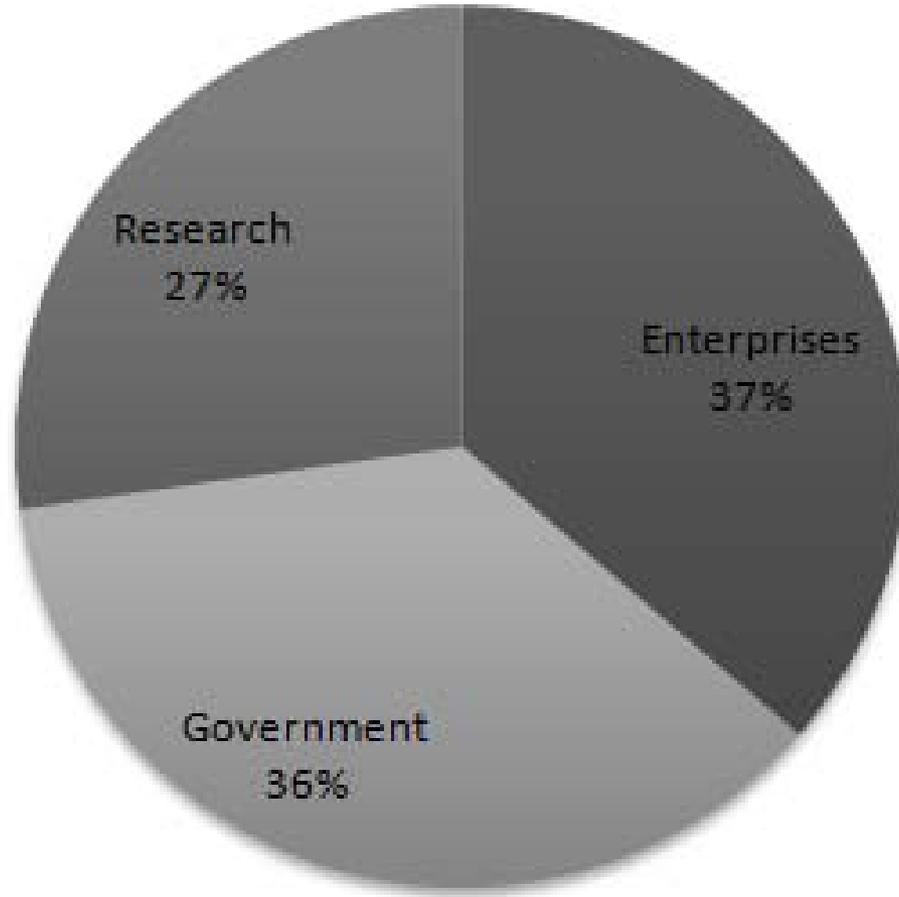
- An introduction of LongRec
- Case study: the National Library of Norway
- Corrent preservation strategies
- Migration challenges

LongRec project

- 3+ projects, research and case studies:
 - Start from 2006, end at 2010
 - Sponsored by the Norwegian Research Council
 - Hosted by DNV
- 12 Partners:



Brønnøysundregis
trene



Design a “persistent, reliable and trustworthy long-term archiv[es] of digital information records with emphasis on availability and use of the information”.

Agenda

- ✓ An introduction of LongRec
- Case study: the National Library of Norway
- Corrent preservation strategies
- Migration challenges

Preservation objects at the National Library of Norway



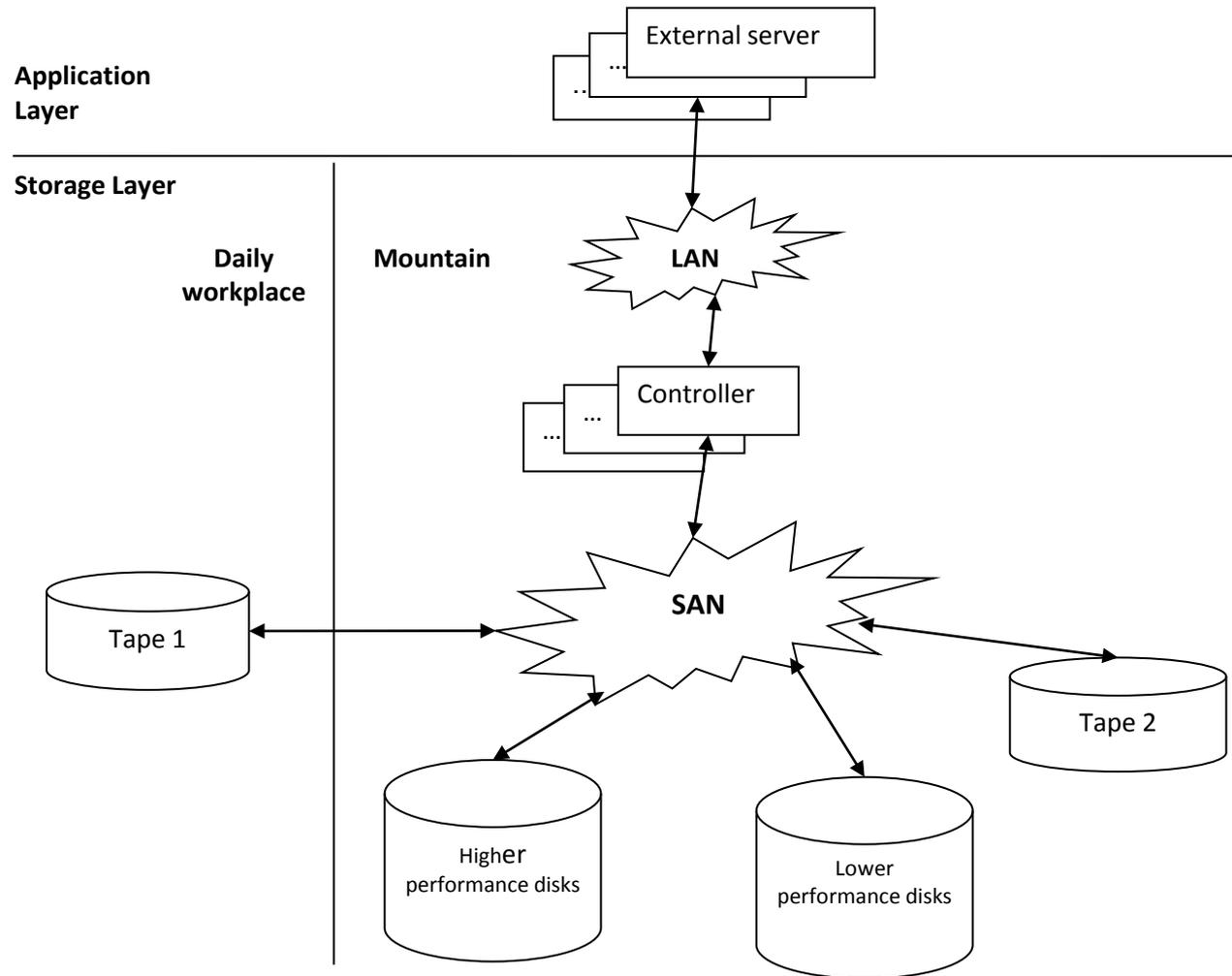
If it's not digital, it's not accessible



- Preserved digital objects:
 - 4 formats for one digital object
 - Three copies.

- Procedures
 - Scanning books
 - Extracting text from TIFF
 - Creating metadata
 - Preserving digital books

Long-term preservation system at the National Library of Norway



Agenda

- ✓ An introduction of LongRec
- ✓ Case study: the National Library of Norway
- Corrent preservation strategies
- Migration challenges

Preservation strategies

Strategy	Description
Computer museum	The entire old computer system that was used to create digital information is retained.
Emulation	Emulation contains two parts. The first part is to develop a software application to imitate all functionalities of computer hardware. The second part is to preserve all necessary application software to interpret digital objects.
Encapsulation	Encapsulation serves to preserve digital information in its original form with the specification of its format used. People must rewrite the source code of the interpreter so that the interpreters can run on new computers to access the old digital information.
UVC	UVC is similar to JAVA Virtual Machine. All digital information is saved with particular instructions of a UVC. The content can be rendered by a UVC interpreter.
Migration on request	For migration on request, people should preserve the original bit streams in the original format and do migration only when future end users want to access preserved digital information.
Batch migration	Batch migration transforms archived digital information to another format before the old format becomes obsolete.

Agenda

- ✓ An introduction of LongRec
- ✓ Case study: the National Library of Norway
- ✓ Corrent preservation strategies
- Migration challenges

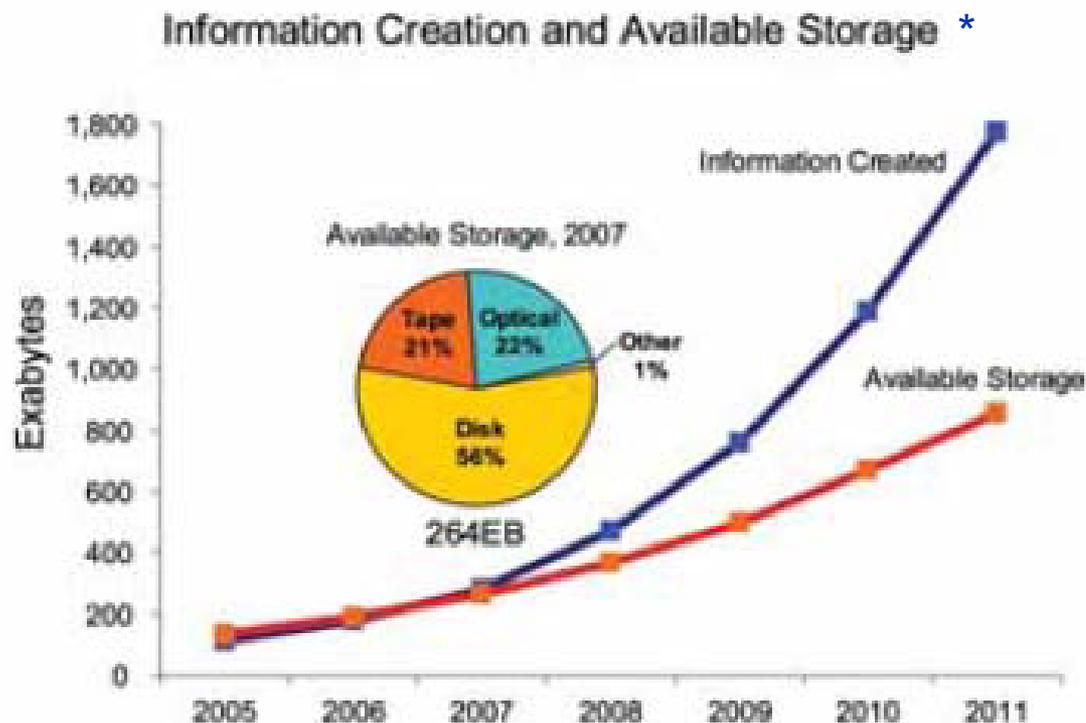
Challenges in Migration

- We have identified three challenges:
 1. Very large amounts of data in preservation systems
 2. Lack of a comprehensive migration framework
 3. What metadata should be preserved for migration?

C1: Very large amounts of data in preservation systems

■ IDC research:

- 281,000 petabytes data in 2007 and the annual growth rate is about 60%.
- In 2020, it will be more than 200 times of 2007.

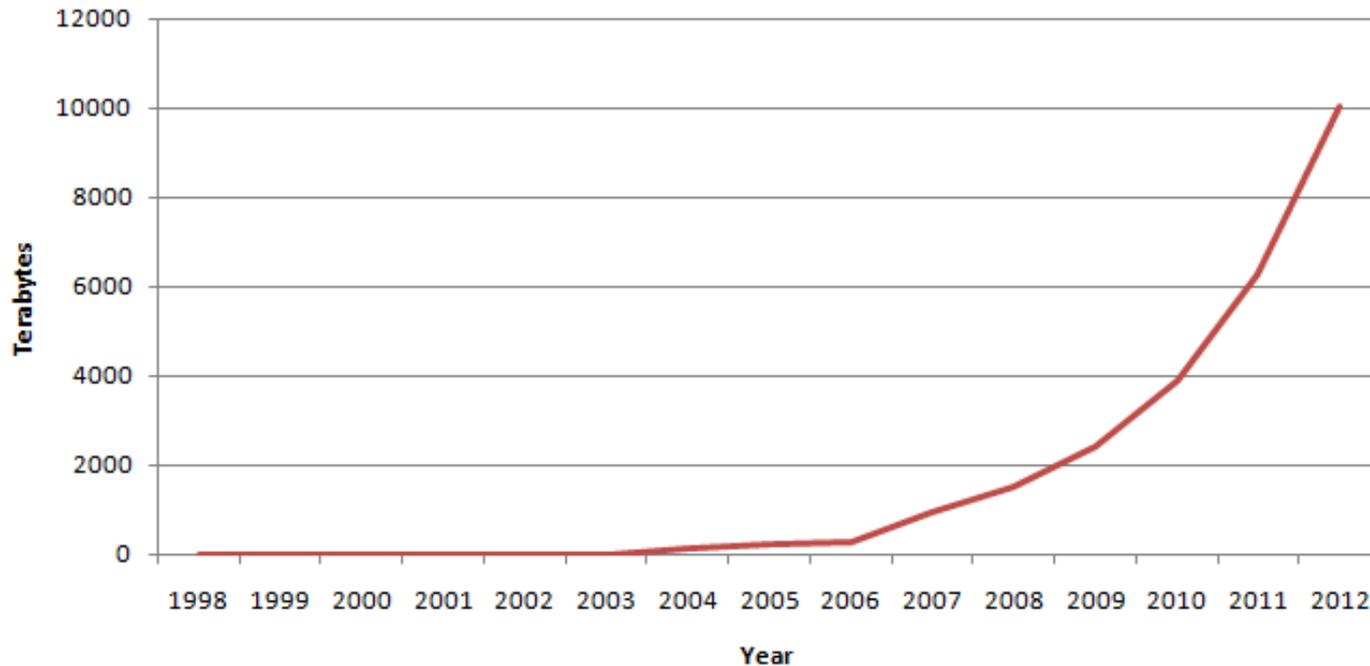


* J. F. Gantz, C. Chute, A. Manfrediz, S. Minton, D. Reinsel, W. Schlichting, and A. Toncheva, "The Diverse and Exploding Digital Universe: An updated forecast of worldwide information growth through 2011," 2008.

C1: Very large amounts of data in preservation systems

■ The National Library of Norway:

- In 2007, the library preserved **957 terabytes** data
- Annue growth is also circa 60%.



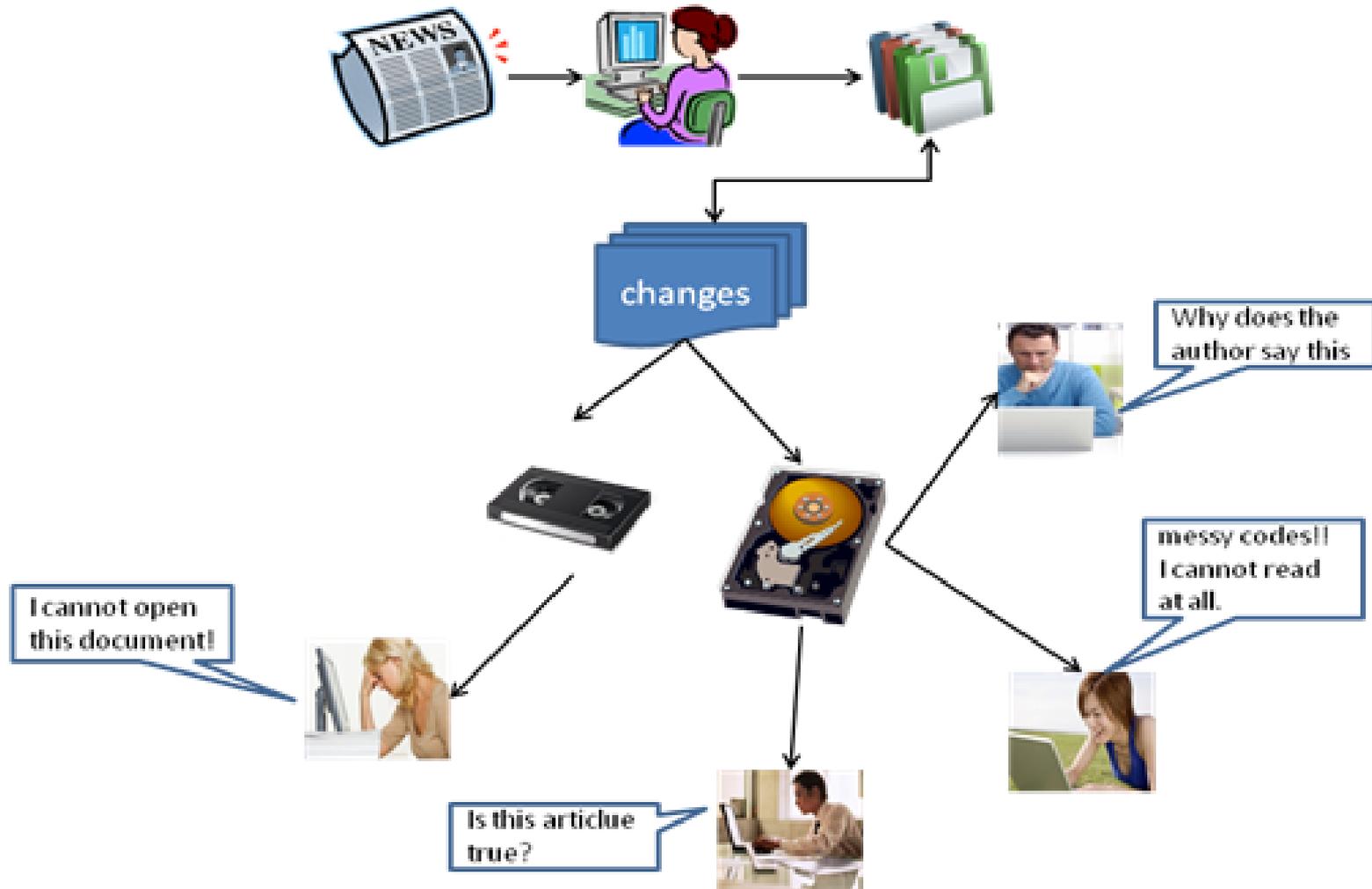
- Research Problem:

- *How much data can we afford to migrate in terms of the limited migration time and cost?*

- Methodology:

- *Mathematical equations to predict migration time.*

C2: Lack of a comprehensive migration framework



C2: Lack of a comprehensive migration framework

- Research Question:
 - What is an comprehensive migration framework?

- Methodology:
 - Survey
 - Questionnaire

C3: What metadata should be preserved for migration?

■ Research Problem:

- Many metadata for preserved digital objects, like metadata designed by CEDARS, NLA, NEDLIB and OCLC, respectively.
- No one investigate metadata related to migration.

■ Research Question:

- What metadata are affected migration?

■ Methodology:

- Survey
- Questionnaire

The end.

Thank you!

Feng Luan

Email: luan@idi.ntnu.no

Office Tel: +47 7359 1457