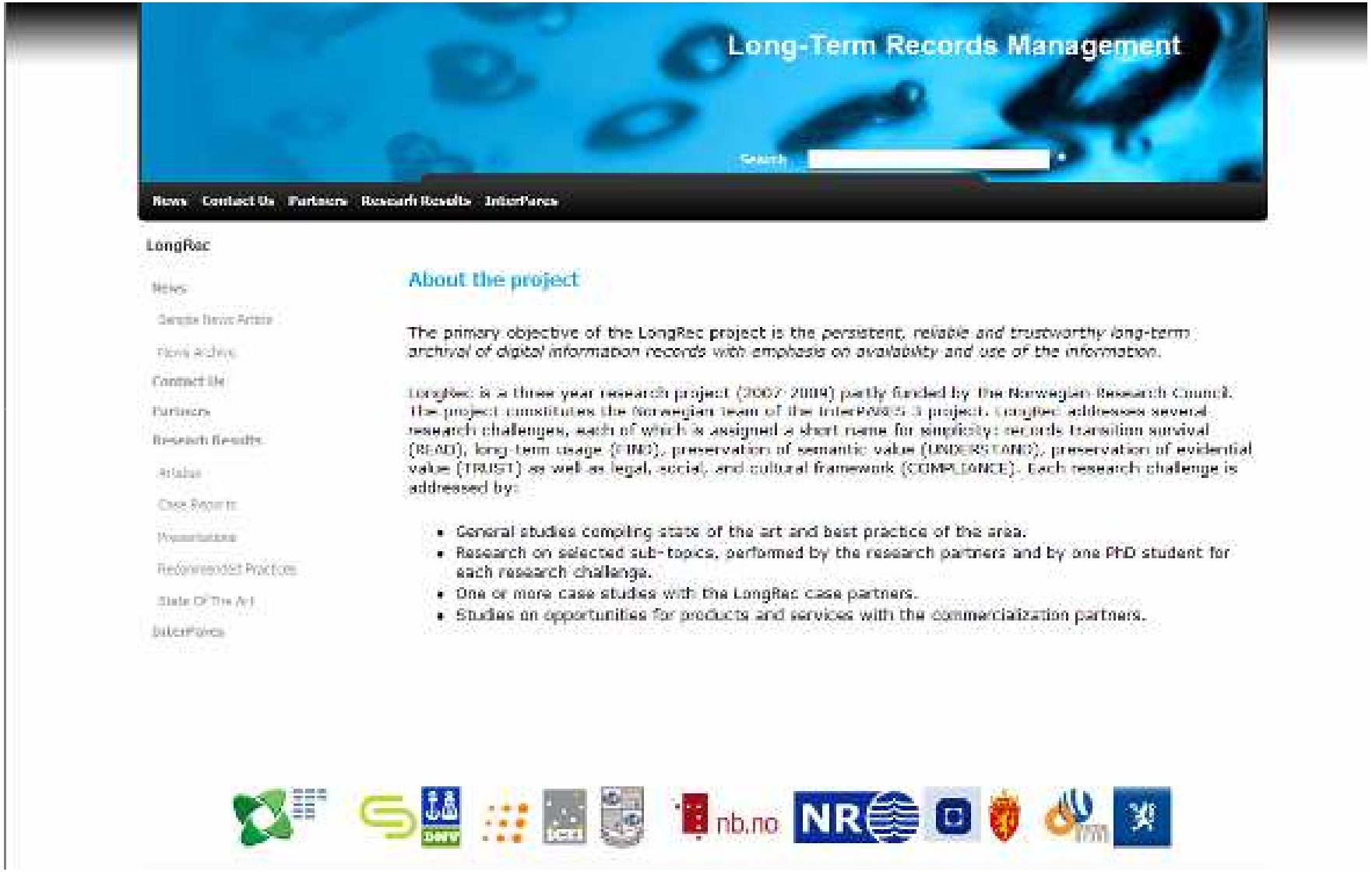
MANAGING RISK

Digital long-term challenges

Det Norske Veritas

Olga Cerrato
25 September 2008

Long-Term Records Management



The digital disease

- Symptoms
- Development of the disease
- The infectious agents
- The patients

- Wrap-up

DNV – an independent foundation

MANAGING RISK

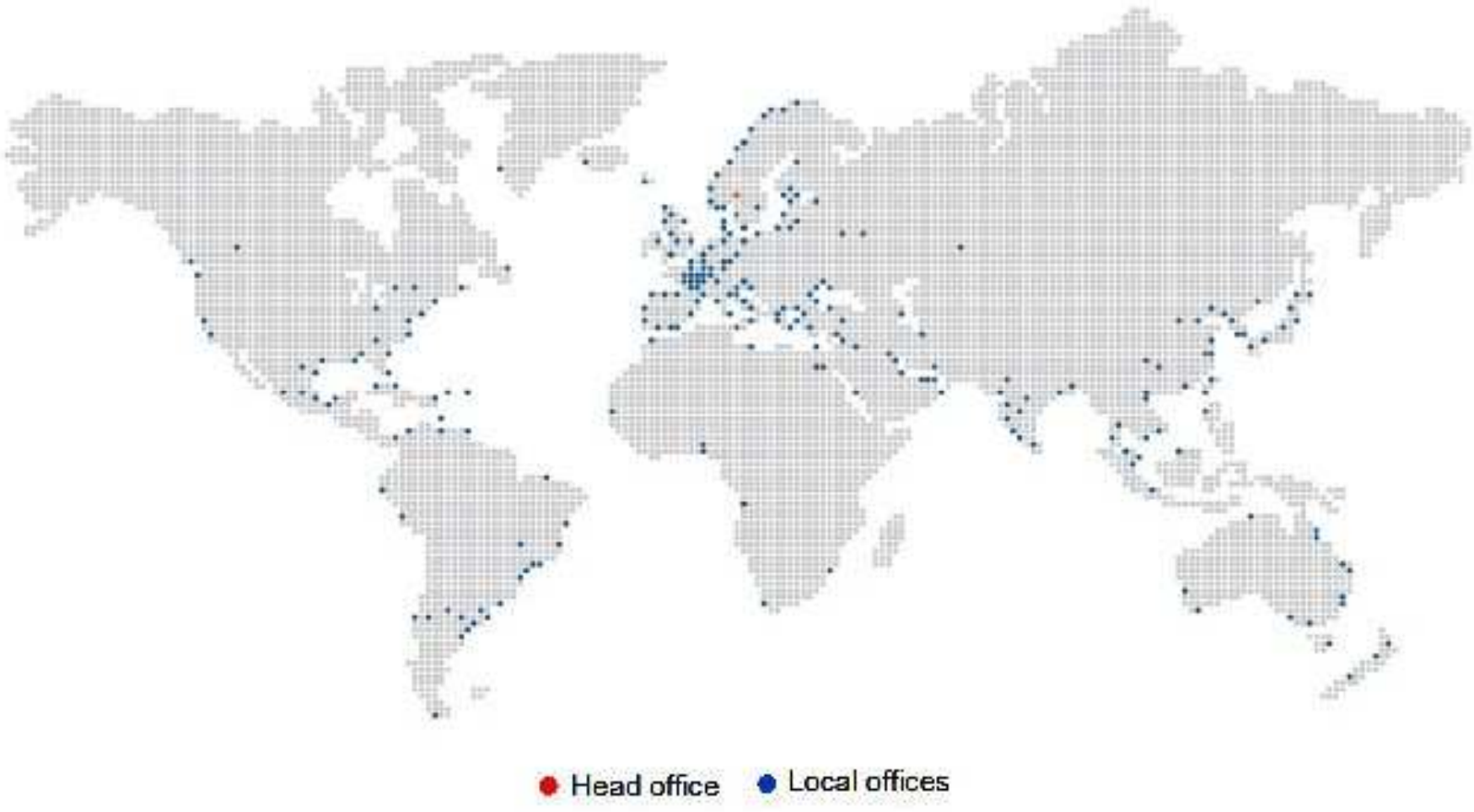
"Safeguarding life, property, and the environment"

More than 140 years of managing risk

MANAGING RISK

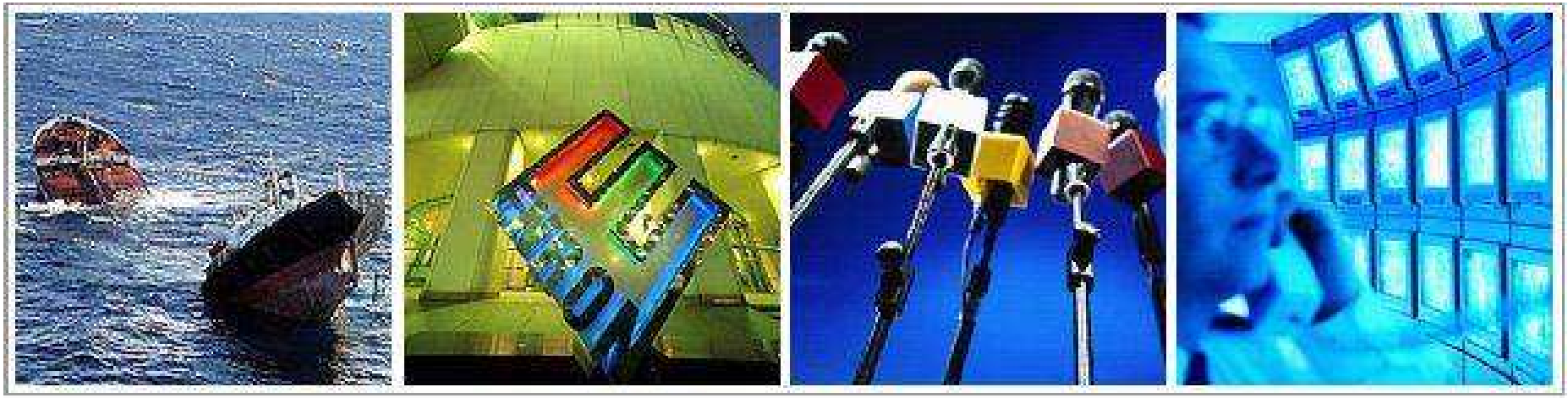
- Det Norske Veritas (DNV) was established in 1864 in Norway
- The main scope of work was to identify, assess and manage risk – initially for maritime insurance companies

300 offices in 100 countries



New risk reality

- Companies today are operating in an increasingly more global, complex and demanding risk environment



- Society at large is gradually adopting a “zero tolerance” for failure
- Increased demands for transparency and business sustainability
- Stricter regulatory requirements
- Increasing IT vulnerability

Core competence

Target industries

Managing risk

Maritime

- Ship classification
- Certification of materials and components
- Assessments and solutions
- Fuel testing
- Training

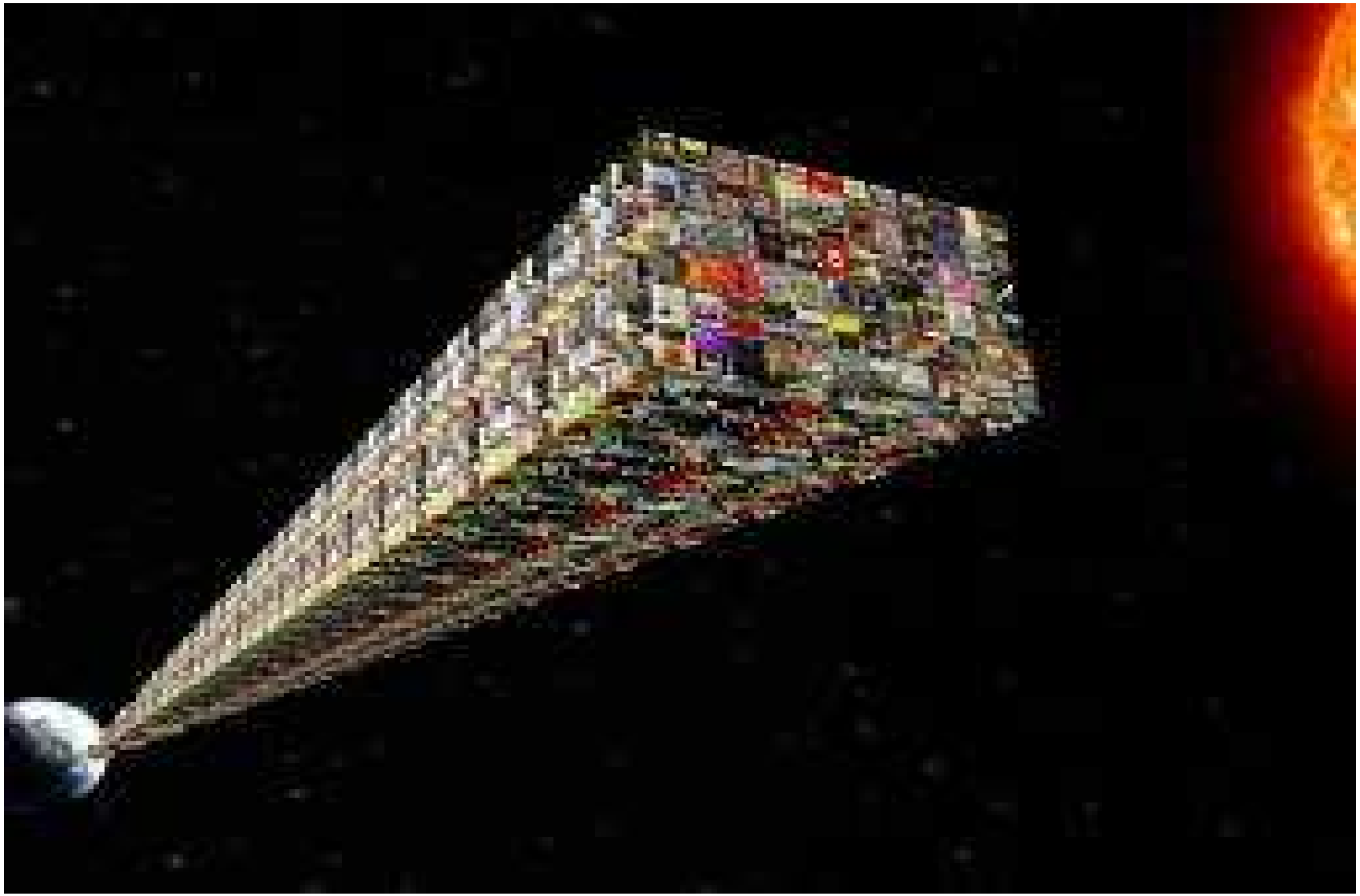
Energy

- Risk management consulting
- Qualification and verification
- Offshore classification
- Laboratory services
- Training

Other prioritised industries

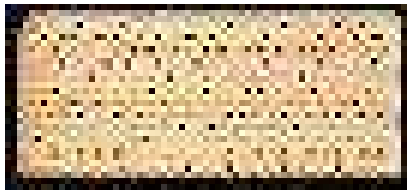
- Automotive 
- Defence 
- Finance
- Food and Beverage 
- Health care
- ICT and telecom 
- Public sector 
- Transportation 

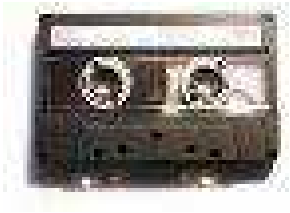
Digital Information Production in 2010 in the World



Does anyone remember how to use this?

Robotron 1370





Or these?

Symptoms?

Or…very soon these?

Accessibility of content

Rosetta stone

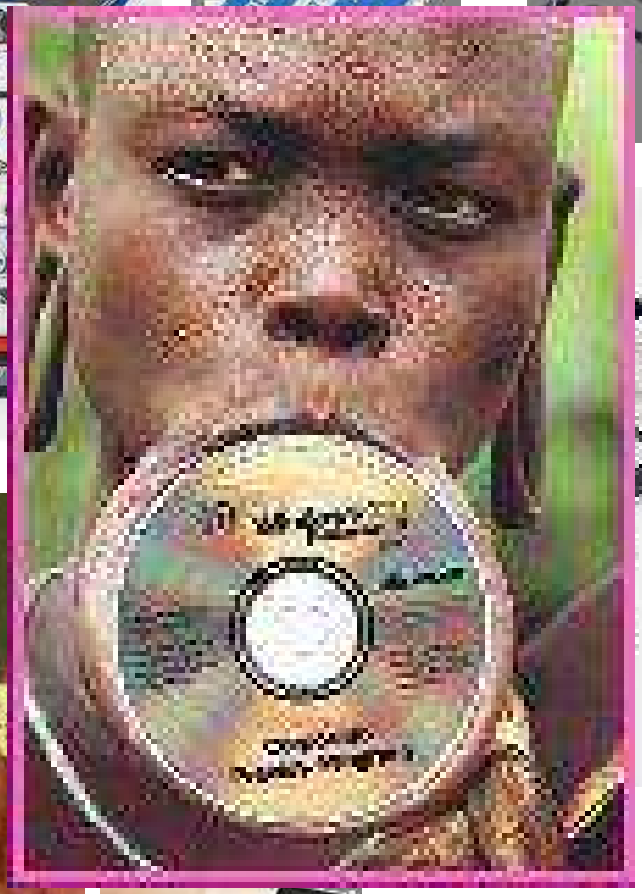
- Text understood today

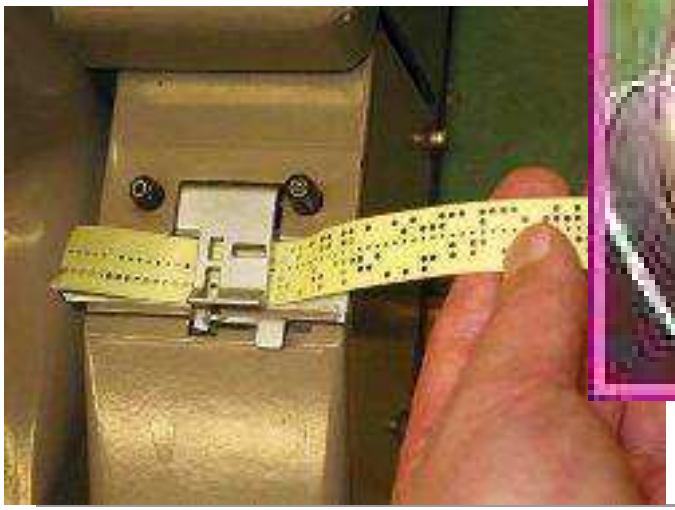
Robotron

- East German computer
- No-one knows how to use it

BUT INFORMATION PRODUCED AFTER 1990 will be lost if we don't do anything!

Symptom: Hardware Obsolescence





Symptom: File Format Obsolescence

- **Proprietary, closed specifications, e.g. Word.doc.** Evolve quickly, exist in many different versions for different platforms, with only limited backward compatibility

- **Proprietary, open specifications**, e.g. Adobe.pdf. Vulnerable to market forces as they can be abandoned for commercial reasons.

- **Non-proprietary, open specifications**, e.g. JPG. Guaranteed long-term availability, specifications published by international standards bodies. BUT these standards must be widely adopted by both user and developer.

Other Symptoms: Traceability over time

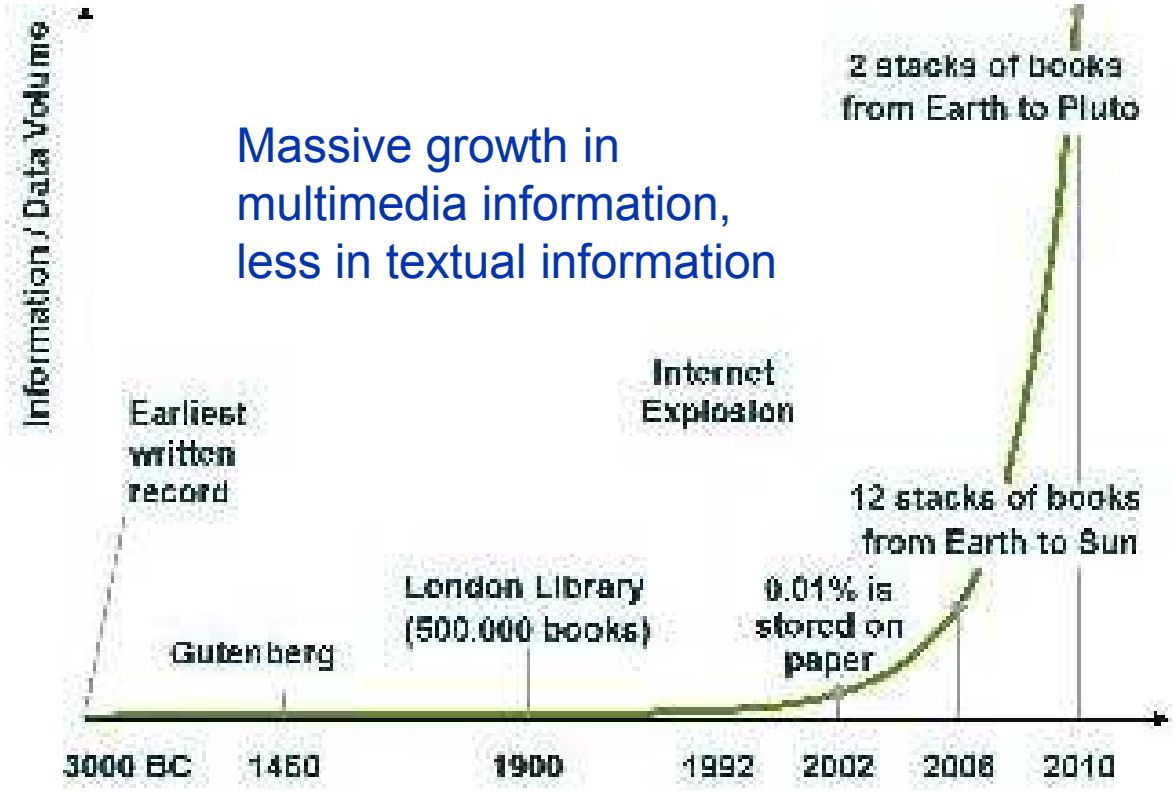
The Norwegian branch of Nordic bank Nordea vows a full investigation into how bank account statements for Princess Märtha Louise and other celebrities wound up in the hands of reporters at magazine *Se og Hør.* (Aftenposten, 12/2-2007)

The bank, regulators and other media are crying foul after newspaper *Dagens Næringsliv* reported over the weekend that the royal bank account statements were leaked to the magazine. It's the latest in a string of revelations about reporting techniques at *Se og Hør,* most of which have been revealed in a new book by a former staff writer at the magazine.

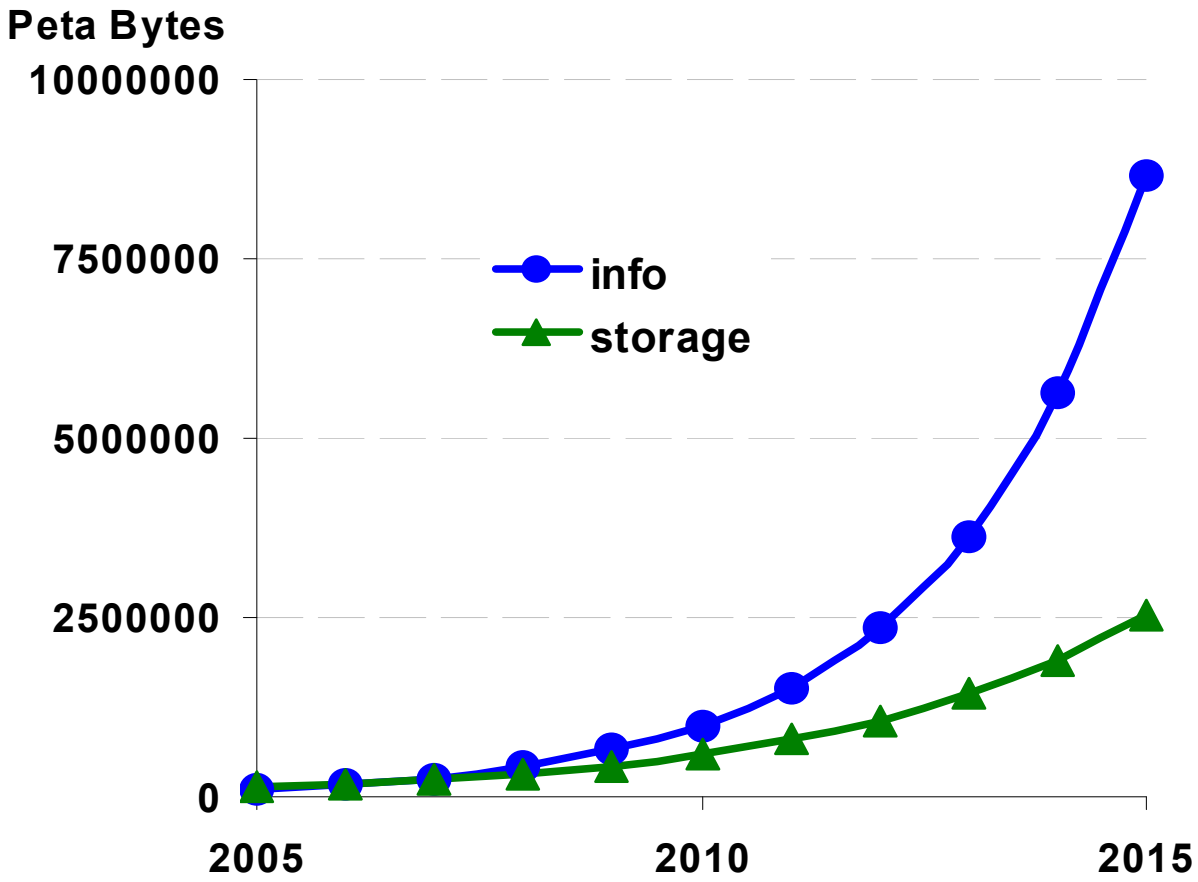
- **No special security**
  Account information for members of the royal family or other public officials or celebrities isn't subject to any stricter security controls, meaning that anyone dealing in customer service at the bank can have access to the accounts. Nordea has nearly 4,000 employees in Norway. …**it's possible to track who may have accessed the accounts, but that it may be difficult to track such information if the access occurred many years ago.** Other banks in Norway have much the same practice as Nordea, meanwhile, with all customer service employees able to access all accounts.

Development of disease: Volume explosion

- 90% of all data is unstructured (pictures, video, e-mails, blogs, …)
  - no data model, no meta data
- 70% of all data belongs to individuals and are de-centralized stored
  - Video, Photos, web pages, ect



Development of disease: Storage Shortage



- 2015 annual growth:

  8 Zettabytes (= 1 million Petabytes),

  = ca 20 x higher than 2008.

- **2015: Data created will be three times amount of available storage.**
- **Lots of data will be for immediate consumption only**

The infectious agents

Disease also known as:

- technological advances
- new developments
- new products

Challenges:

- Technology/systems life-time
- Software lifetime
- Formats' lifetime
- Processes' lifetime
  - ➢ Conversion, migration
- Volume
  - ❖ Search and retrieval
- Trust
  - Compliance (laws and regulations)

In 2015 **80%** of today's employees will still be working but **80%** today's technology will be replaced by the new one

INFORMATION outlives most of us and most of it will live forever!

The ticking digital bomb…

- 2010 six times the data we produced in 2006
- In 3 years we will produce the same amount of information as we have produced so fare in life
- Hidden information cost
  - Massive volumes
  - Unstructured information
- More rules and regulations
- More integrated tools
- Increased organized internet crimes

The information outlives the information carrier !

Are we prepared for this?

We need to find routines, procedures on how digital information can be read and understood into eternity

LongRec – one step closer to the solution…

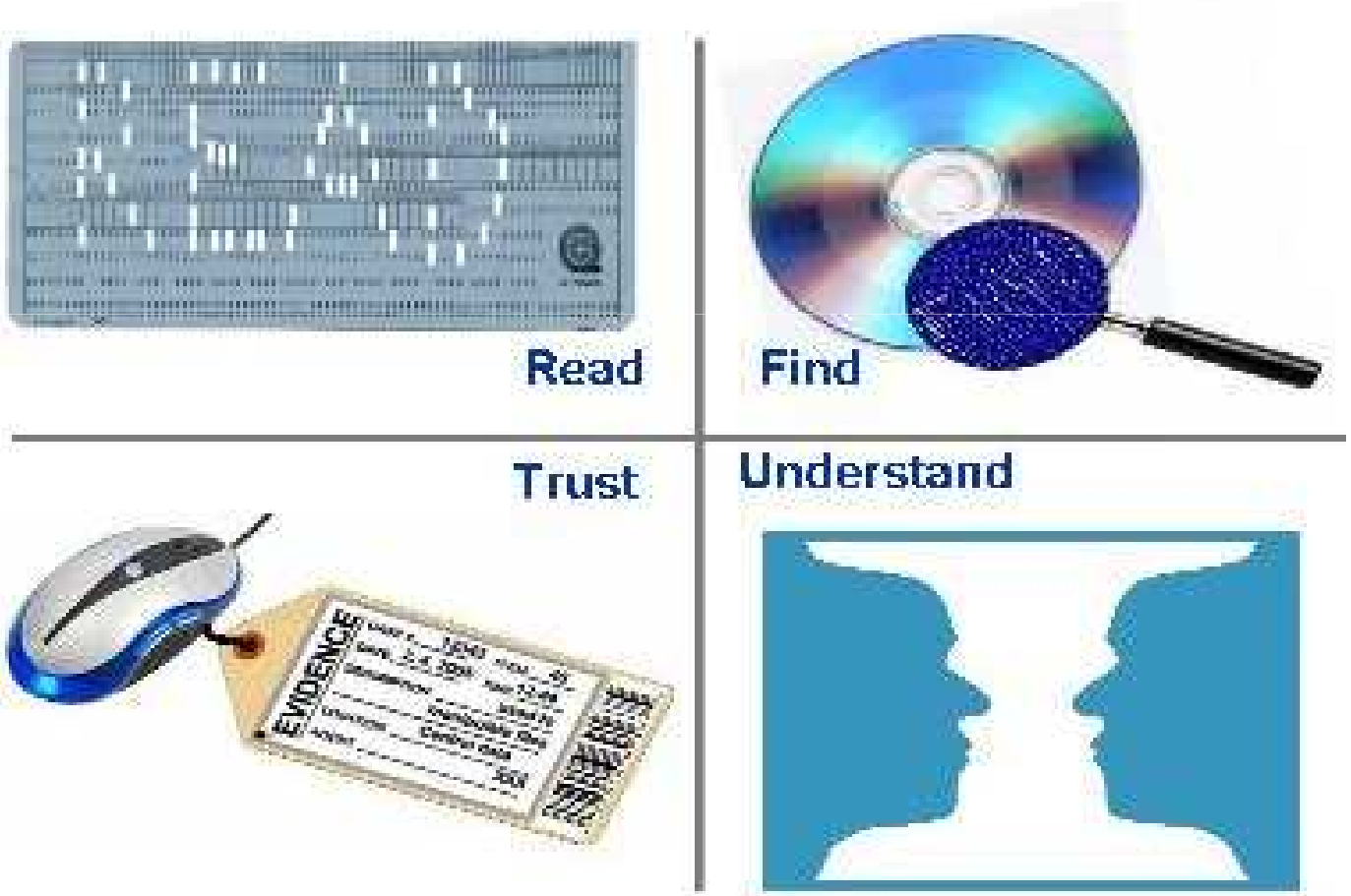
DATA =

DIGITAL ACCESS THROUGH AEONS

- 3+ year project, research and case studies
  - DNV R&I lead, 10 partners
  - Start October 2006, end November 2010
  - Overall budget 27,6 MNOK, Norwegian Research Council grant 9.2 MNOK
  - 3 PhD theses in work

http://www.longrec.com

DATA = Digital Access Through Aeons



+ COMPLIANCE

Project partners

- InterPARES 3: http://www.interpares.org
- ICRI (Interdisciplinary Centre for Law and ICT), Katholieke Universiteit Leuven

The primary objective of LongRec

- Persistent, reliable and trustworthy long-term archival of digital documents, with emphasis on availability and use of documents
  - Enable transition to digital original documents and digital work processes even for information that must be available and in use over decades
  - Explore the potential for commercial products/services in this area

Why

➔ **Digital storages are not yet trusted for long term storage (20+ years) of original documents**

- Blocks transition to digital work processes for organizations that require storage and use of documents over several decades
- Or organizations make the transition in the hope that “future development” will handle the problems => potentially high risk

- Some key requirements
  - Documents need to be available for their entire lifetime
    - Technology changes need to be transparent to the user
  - Security and access control need to be maintained for the entire lifetime
    - Document owners and responsible will most likely change
    - Digital signatures must be maintained and be verifiable
    - Digital original documents are becomming legally binding
  - Changes and amendments might be carried out at any time in the lifecycle
  - Regulatory compliance and management of operational risk
    - Demonstrate legal and regulatory compliance
    - Support operational risk management

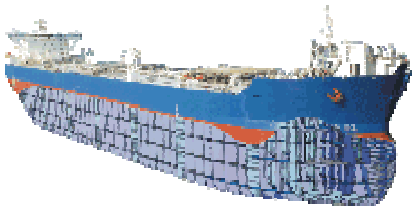
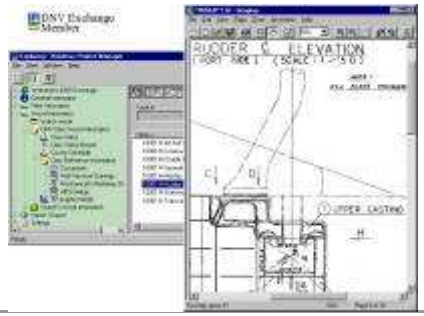
Is this not solved already?

- Technology lifetime is shorter than document lifetime
  - Up to 15 years is realistic lifetime for a Document Management System
- Preservation systems do not allow for changes and amendments
  - Do not maintain ownership and access control
  - Preserve content “forever”, not all documents suited for this
    - Proprietary content
    - Outdated and not correct any more
    - Sensitive information
- Organizational lifetime is shorter than document lifetime
  - Documents can pass from one owner to a new one
    - Ex: Oilrigs, ships and airplanes are all bought and sold during their lifetime
    - Mergers and acquisitions happen during document lifetime
  - Organizations change
    - Reorganizations
    - People quit, retire, and new people start
- Documents migrated without history and context
  - Mostly snapshots migrated today, history left in old systems

➔ Only partial solutions today!

The Patient: DNV (1)

- Transition to digital documents and work processes
  - Not just digital representation of paper originals
  - To gain full benefit from the technology, processes must change
- DNV requirements
  - Documents to be stored for at least 40 years
  - Textual documents, drawings, perhaps photos and multimedia information
  - High demands for availability, integrity, authenticity and confidentiality
  - Digital signatures needed for some documents (DNV certificates)
- DNV interoperability requirements
  - Offices in more than 100 countries
  - Information from/to many actors (wharfs, ship owners, flag states, port states, insurance companies etc.)


The Patient: DNV (2)

- In 40 years, everything will have changed
  - Software, computers, formats, organization, personnel, roles
  - Records management must handle this
- Service development (external services from DNV)
  - Validation and notary services (trusted third-party roles taken by DNV)
  - Information Quality Management
  - Risk management in an information or document life cycle perspective

The National Library of Norway

or

How to store the memory of the nation?

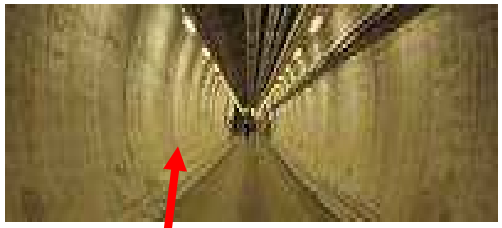
The Legal Deposit Act

The memory…

90 m long

4 floors

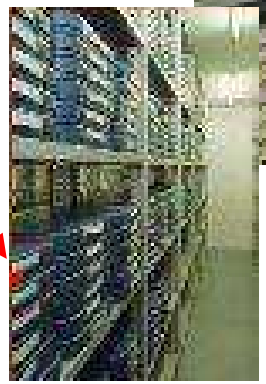
42 km shelves



100 m inside the mountain

Cold storage

Many tons of film



Automatic storage

Place to 1.500.000

documents

The patient: multitude of record types

73 000 music printings,

75 000 hrs/160 000 records

410.000 Norwegian

The national memory and multimedia knowledge centre…

Systematic digitalization of EVERYTHING

Digitalization started a while ago…

Status:

200 000 of 4 700 000 newspapers

365 000 of 1 800 000 pictures

47 000 of 410 000 books

500 of 400 000 hrs film/video/TV

1000 of 75 000 hrs music

5000 of 40 000 posters

300 000 of 1 200 000 hrs radio

0 of 4 000 000 manuscripts

0 of 55 000 maps

0 of 2 500 audio books

….

Digitalization

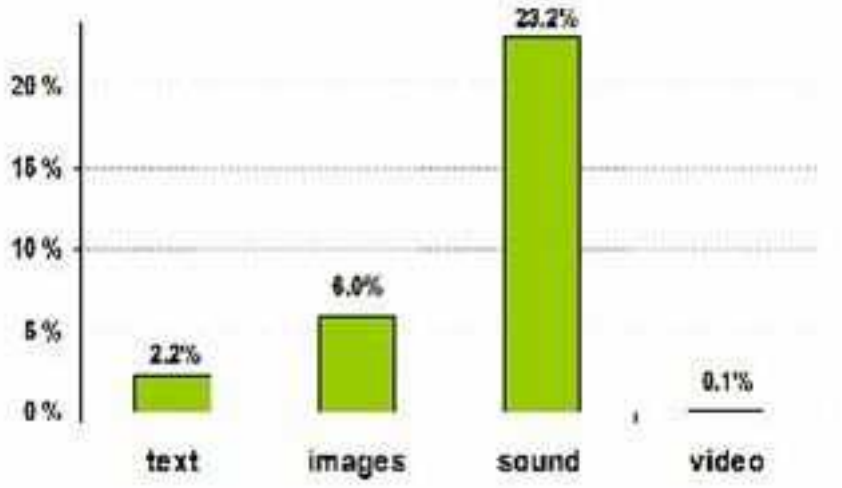
Current state of digitalization: ≈5%

Total volume when today's collections are digitalized (≈ 2018)

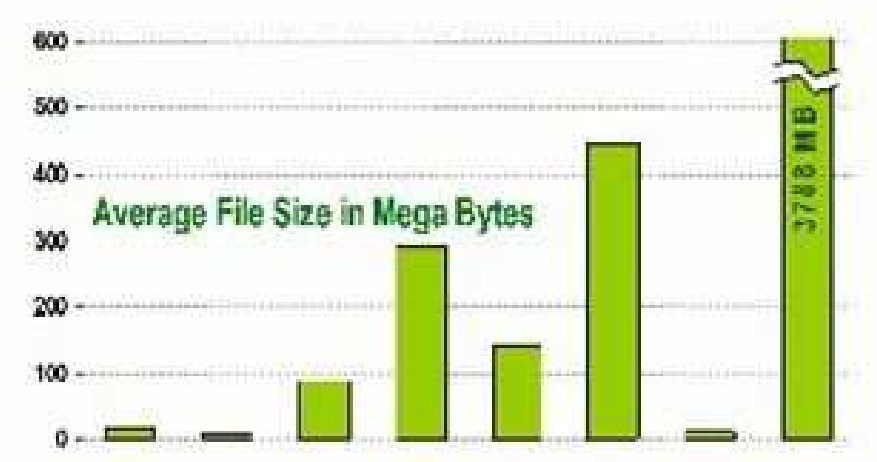
- Estimated total volume: 37 Petabyte
- Estimated number of files: 564.000.000

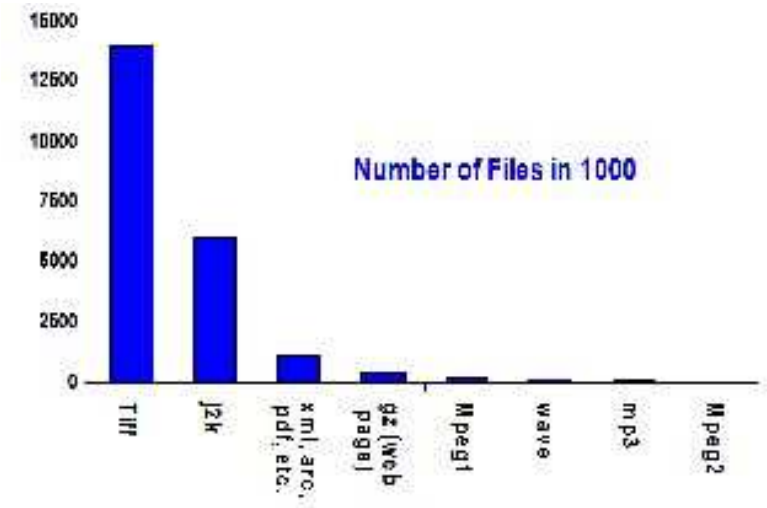
- **In addition:**
  - newly submitted materials
  - TV broadcasts, e.g. digital TV
  - web harvesting (.no domain)

Percentage of completed digitalization



File formats and volumes





File format obsolescence:

not yet an issue

Hardware Support:

3 (4) years only!!

=> copying of ALL files to new storage

(server and 2 tape)

Migration: Moving all the files to a new storage

Estimated:

- 40 Petabytes (≈ 1000 TerraB ≈ 1000*1000 GigaB)
- 560 million files

Assume:

- 1 sec per file transfer
- => 17.7 years !!

More than 4 times the hardware support period

Main challenges

Data volume

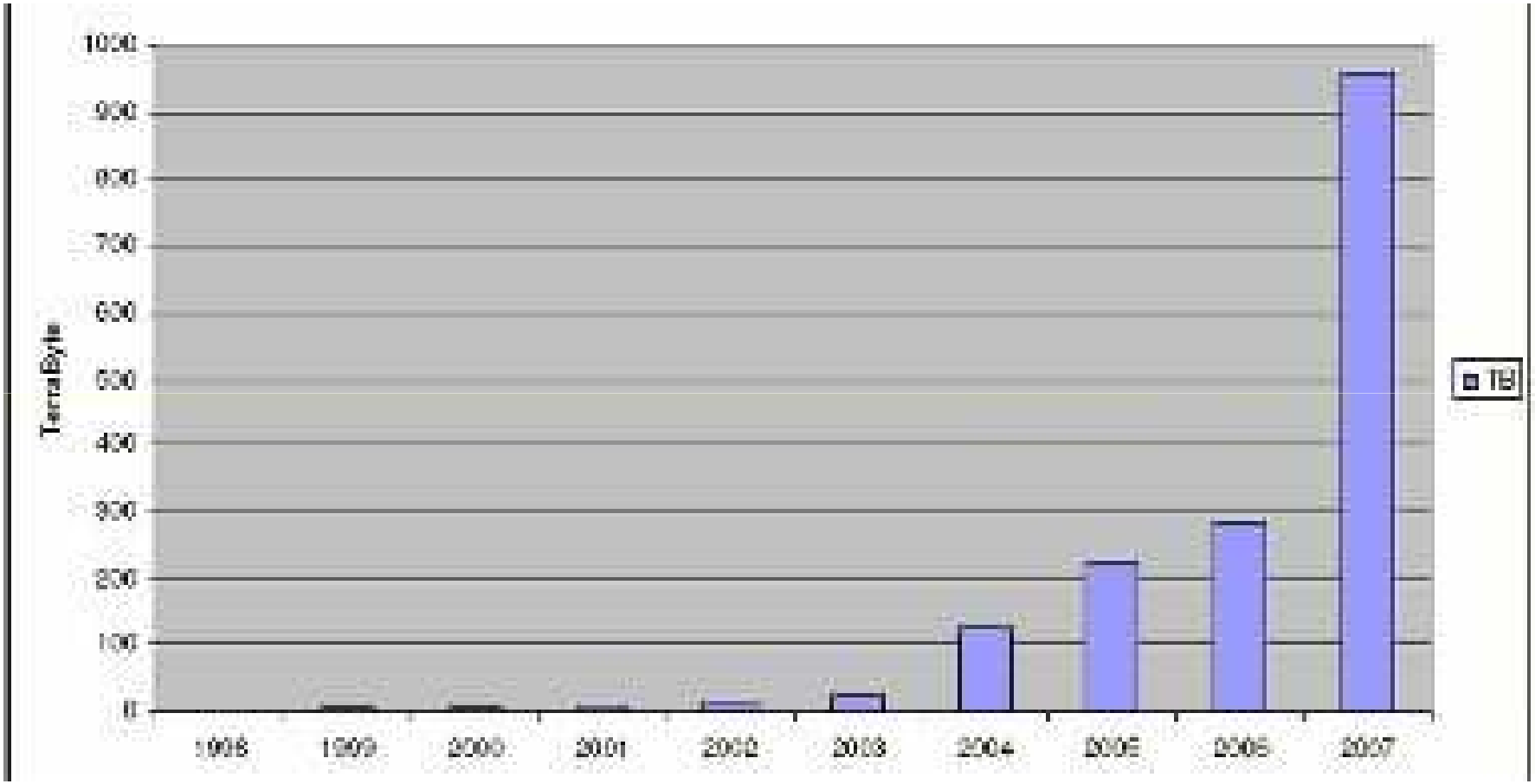
"This year there are TB, the next PB"

- The main principles: 3 copies, to different technologies, 3 places
- 1000 TB (x3) today
- + 750 TB growth annually
- Nothing can be deleted (incl. web-harvesting)

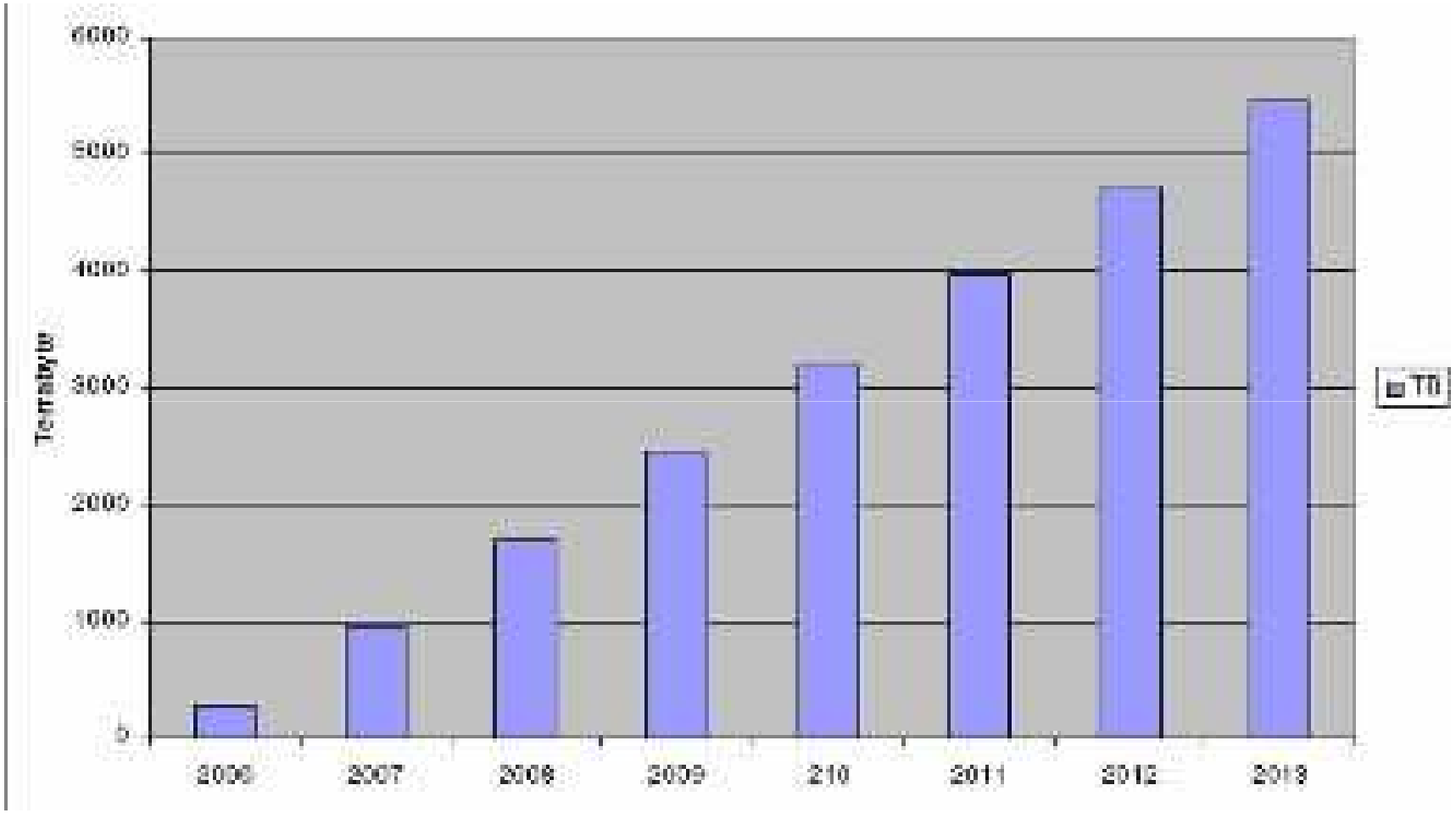
Long-term storage

- All digital content shall be preserved for at least 1000 years:
  - Searched
  - Retrieved
  - Shown
- The item displayed shall be as close to the original as possible
- Data integrity shall be secured

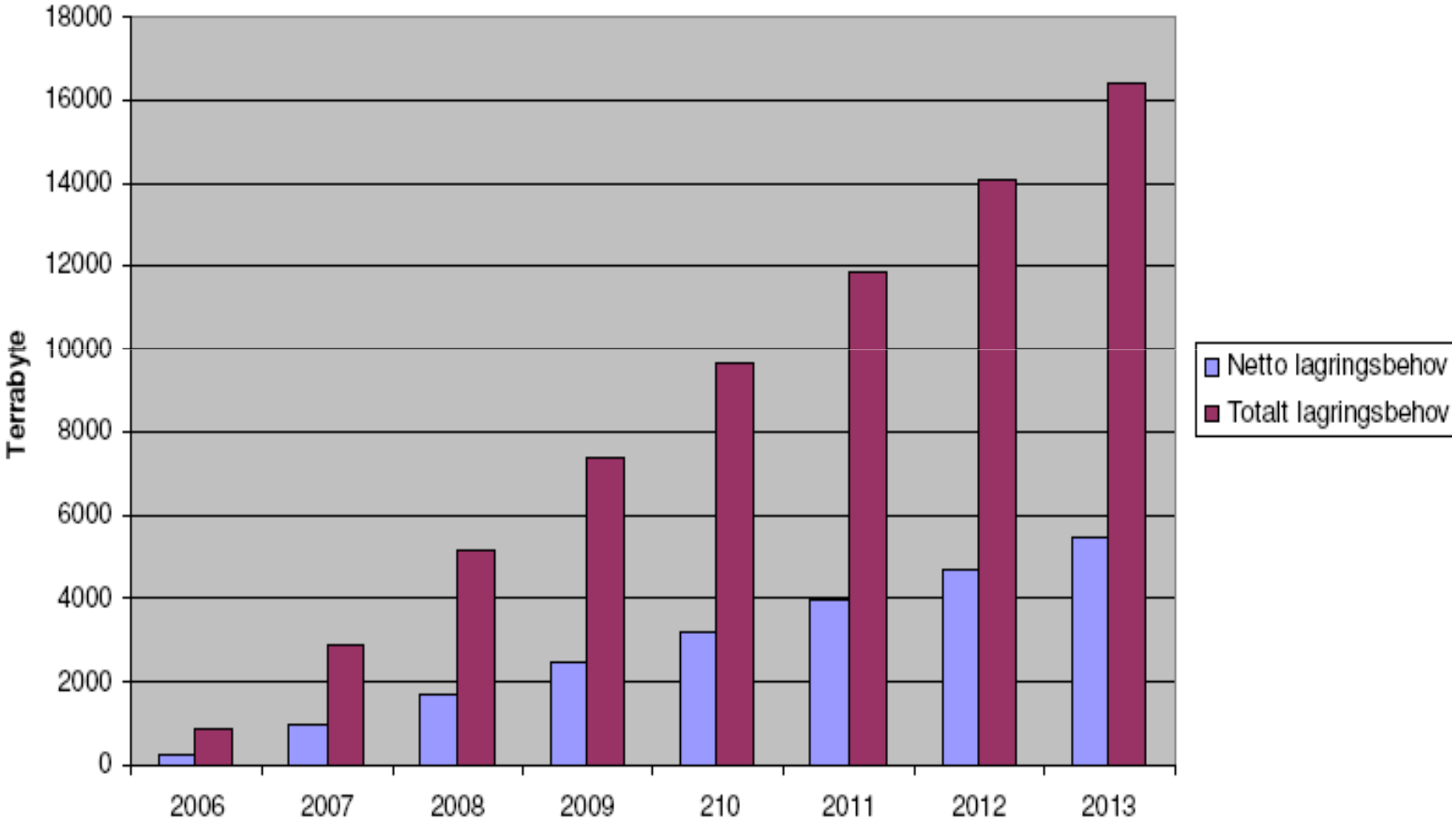
Data volume 1998 -2007



Data volume – prognosis, net

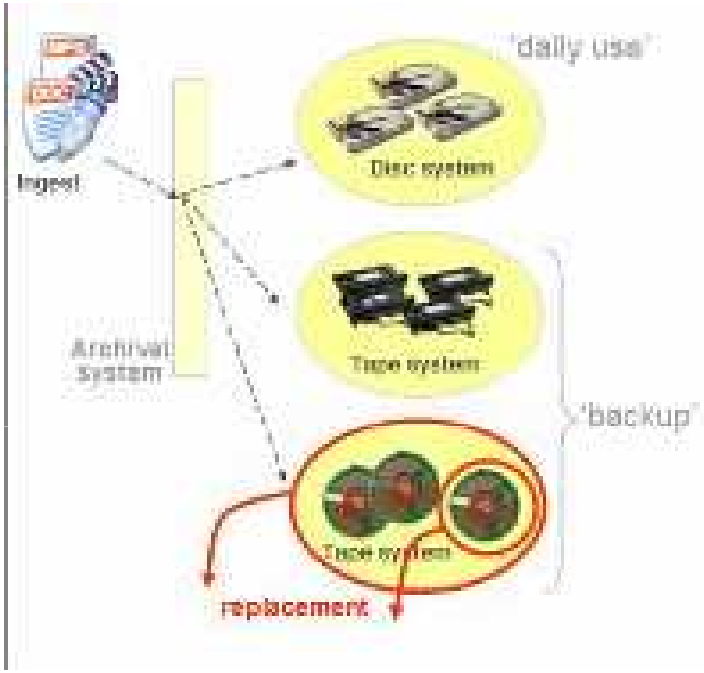


Data volume, prognosis - gross



What is being done today?

- The highest quality possible for the storage of digital objects
- Unique ID
- Metadata
- Minimum 3 exemplars, 2 technologies, 2 localities
- Data integrity check
- DSM (Trusted Digital Repository) application (developed in-house): handles preservation MD and physical placement of the objects



URN

- URN (Uniform Resource Name), IETF; refers to the permanent address of the net document independently of the physical location
- Assigned to all digital objects produced by the NB
- Use as a part of the file name
- Registered in meta DBs
- Internal resolution service
- Is a service to external partners

LongRec and the National Library: trustworthy migration

- **LongRec:** Migration is the process of moving digital objects from one storage media to another to ensure their continued accessibility as the medium becomes obsolete or degrades over time
- **Need:** calculate the migration time and strategy for the given record volume, type, desired quality level etc.

Why migrate?

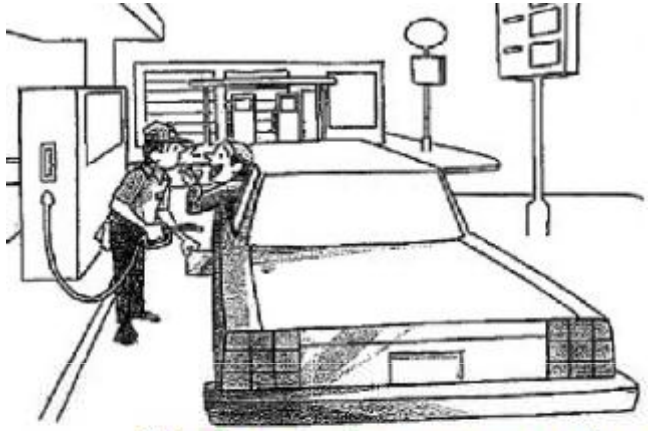
- Reduce risks of media breakdown (increasing failure rate into an unacceptable range)
- More reliable technology (damage or lower failure rate)
- Higher storage density (= less space)
- Cheaper technology (operational cost, energy cost etc)
- Faster technology (access speed)
- Consolidation of media (reduced variation in equipment, need less expertise, less ‘unknowns’)
- New vendors, relations, politics …

Cost are or will be too high in maintaining desired quality level

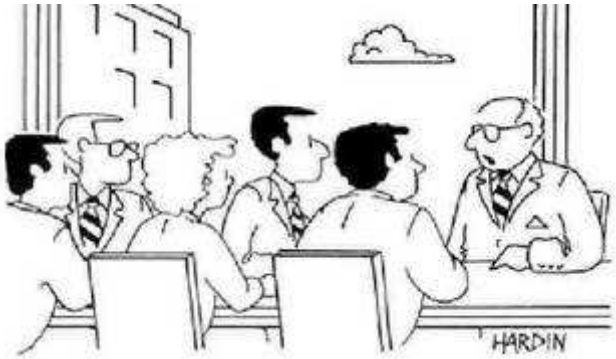
What migration strategy shall be chosen?

It depends on the desired quality level and cost frame. In principle 2 extremes:

- **Minimum Cost:**
  - Use old technology (cheap, known faults, …)
  - Mainly personnel cost
  - Higher risk, slower



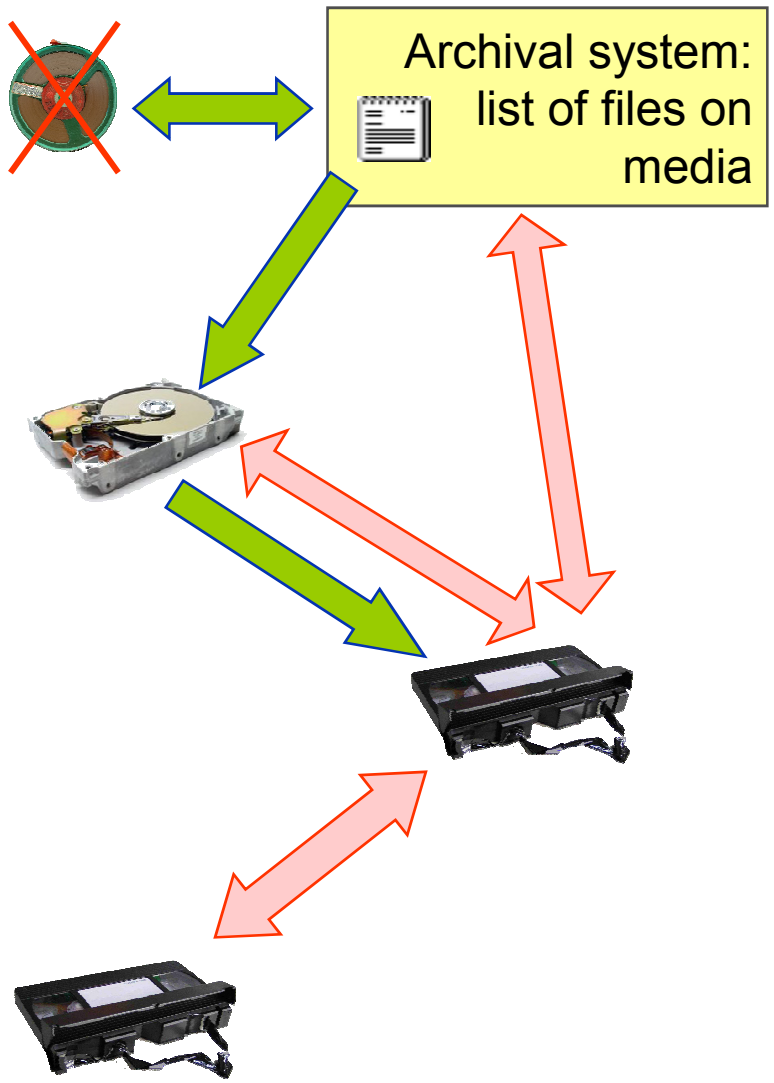
"Just enough to get me to the cheaper station"



"We have considered every potential risk, except the risk from not taking risks."

- **Minimum Risk:**
  - Use tested (by others) technology
  - Through verification and QA
  - Use experts

Migration and Quality Assurance



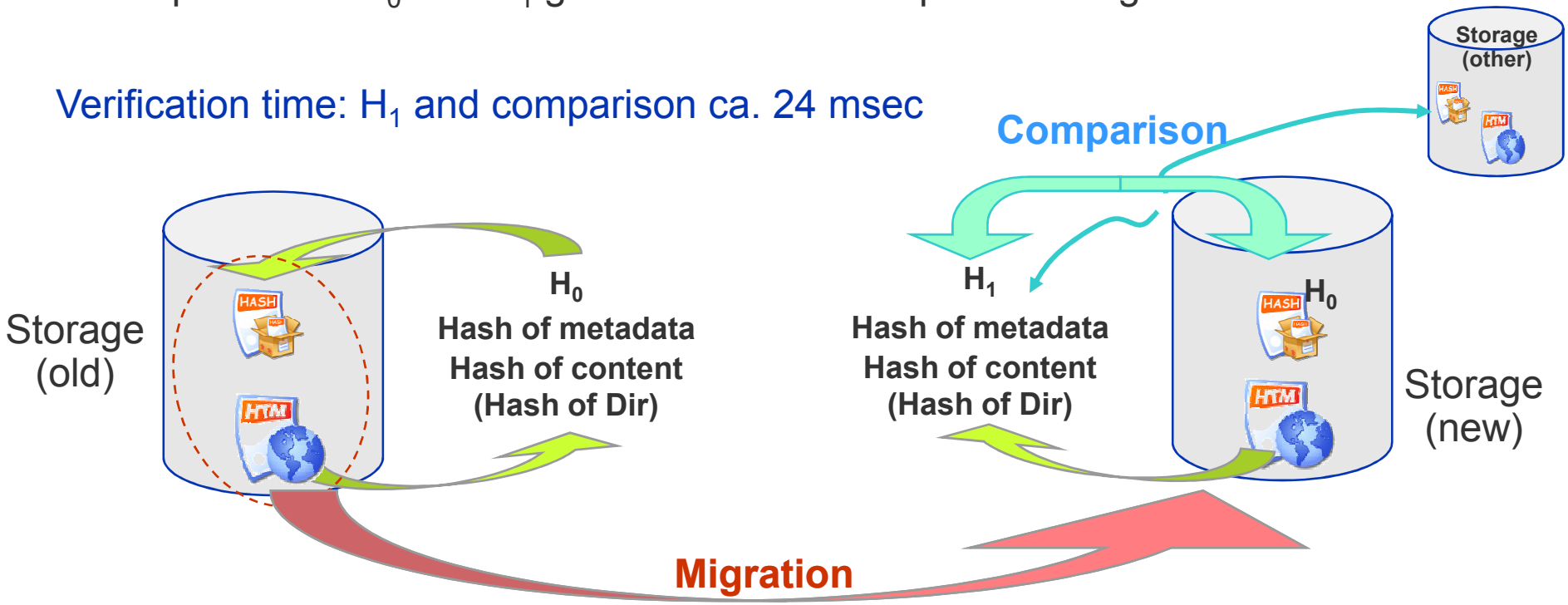
- File system catalogue gives info about files on the old media
- Finding corresponding files (preferably on disc)
- Copying file to new media

- Verification of successful migration
  - Access to new media (catalogue and media check)
  - Comparison to source (individual file check)
  - Comparison to other media (global file check)

Trustworthy Migration

- A HASH value H_0 was created for each file at ingest
- A new Hash value H_1 maybe computed after migration
- Comparison of H_0 and H_1 give indication about possible migration errors

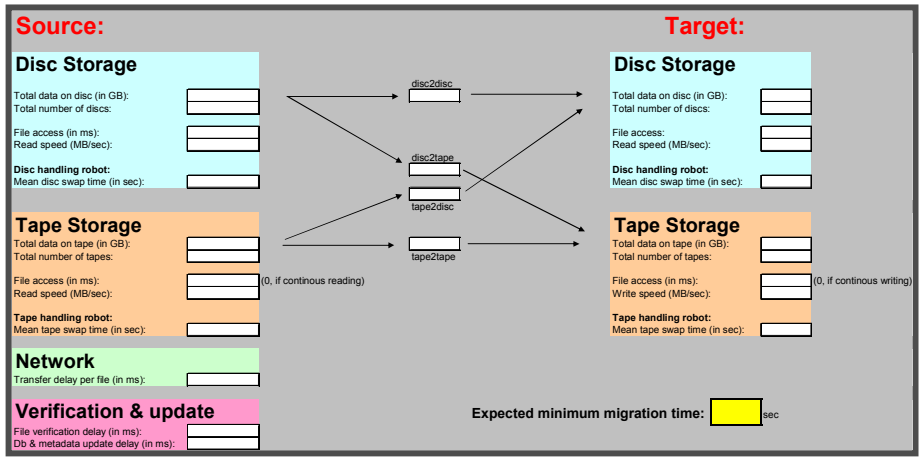
Verification time: H_1 and comparison ca. 24 msec



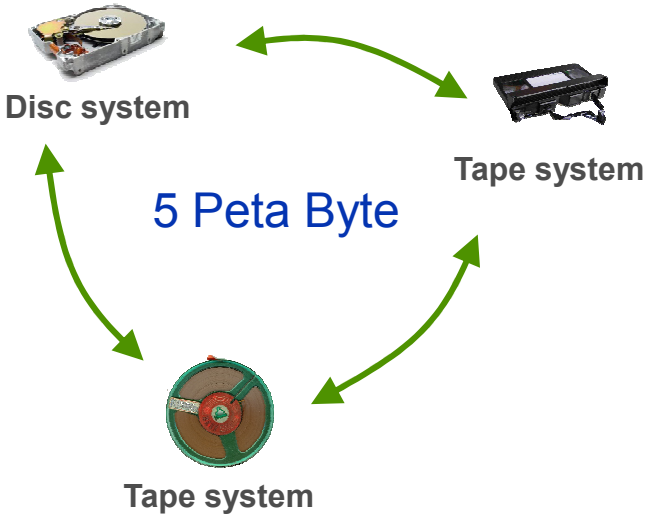
Mitra et al. Trustworthy Migration and Retrieval of Regulatory Compliant Records, 24th IEEE Conference on Mass Storage Systems and Technologies (MSST 2007)

The way ahead…

- ‘Calculator’ - estimates migration time based on a series of parameters



- ‘The Migrator’ – investigate various strategies wrt.
  - reliability
  - modularization
  - automation and practicability



Wrap-up

- Challenges: HW, SW, format, processes obsolescence, organizational changes
- Volume explosion and storage shortage
- The Digital Bomb Metaphor and the LongRec project
- Patient 1: DNV (drawings, at least 40 yrs)
- Patient 2: the National Library of Norway (migration, calculator)

Contact

- Inger-Mette Gustavsen, DNV Research & Innovation
  inger.mette.gustavsen@dnv.com
  +47 6757 7049 / +47 917 08 230

- Jon Ølnes, DNV Research & Innovation
  jon.olnes@dnv.com
  +47 478 46 094

- Olga Cerrato, DNV Research & Innovation
  olga.cerrato@dnv.com
  +47 957 35 880