Automatic retention of records

Norwegian Research Project - LongRec

17 September 2010

MANAGING RISK DNV

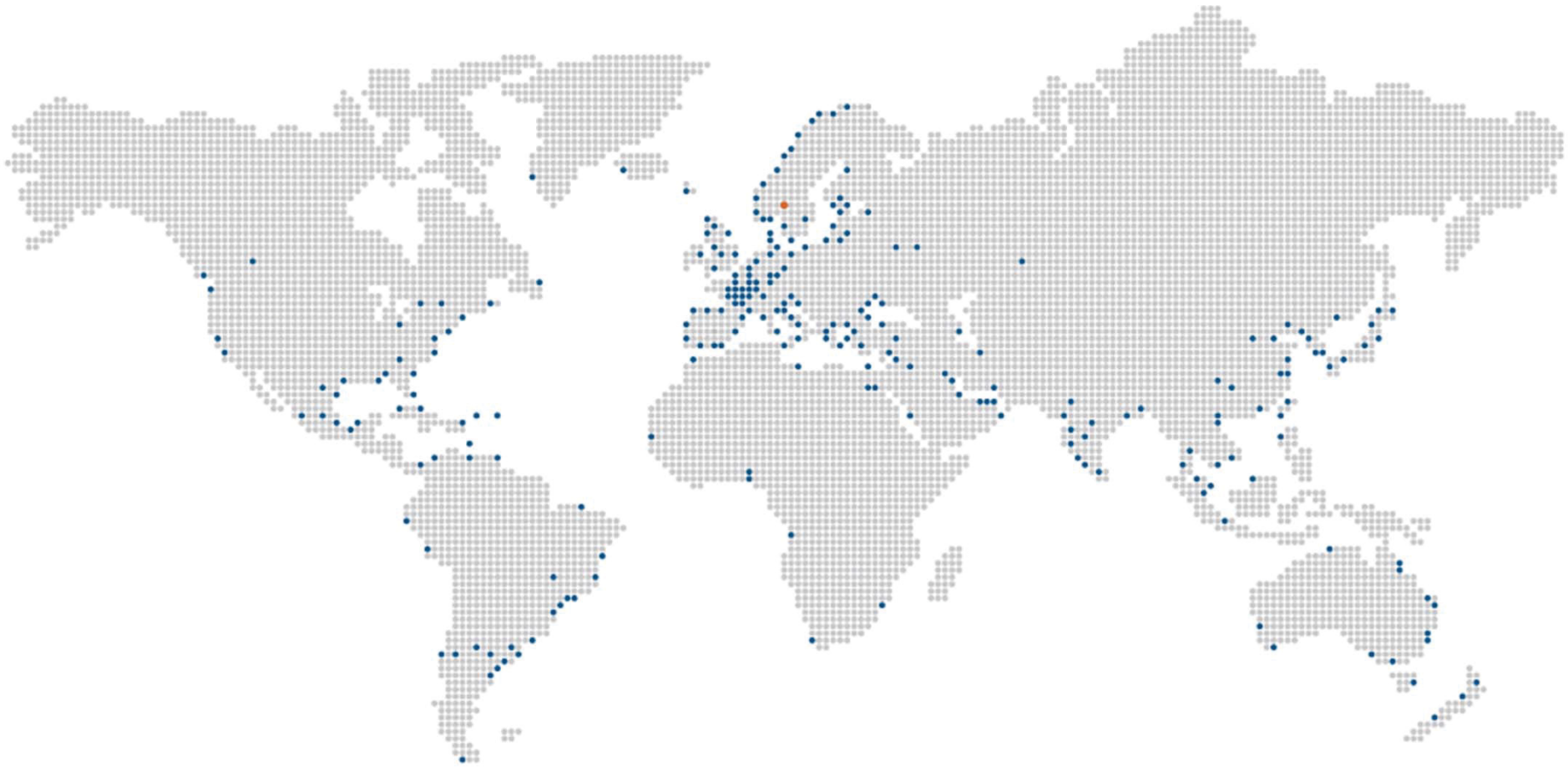
Agenda

- Who is DNV?
- Why LongRec?
- Digital Trends
- Why is semantics important for Preservation?
- How context is linked related to automatic retention?
- Challenges and recommendation
- Wrap up

DNV – an independent foundation

"Safeguarding life, property, and the environment"

DNVs locations

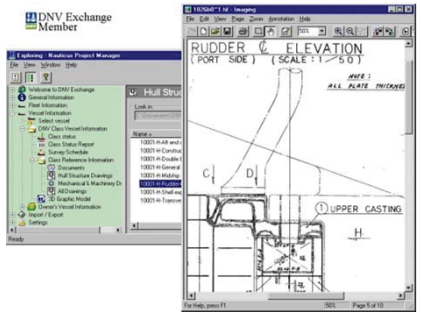


9000 employees

● Head office ● Local offices

DNVs motivation – LongRec

- Transition to digital documents and work processes
- DNV requirements
  - Documents to be stored and updated for at least 40 years
  - Textual documents, drawings, photos, multimedia information
  - High demands for availability, integrity, authenticity and confidentiality
- DNV interoperability requirements
  - Information from/to many actors (wharfs, ship owners, flag states, port states, insurance companies etc.)



LongRec

DATA =

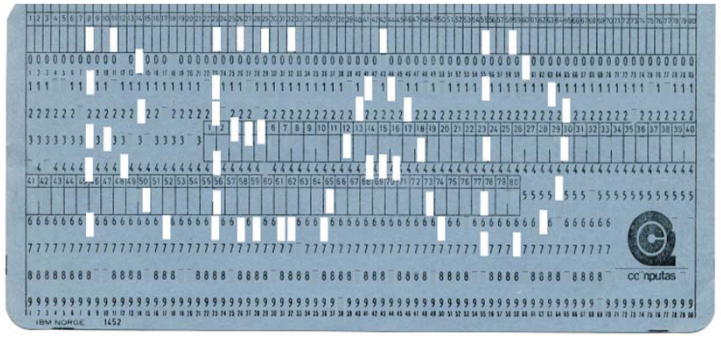
DIGITAL ACCESS THROUGH AEONS

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

www.longrec.com

LongRec

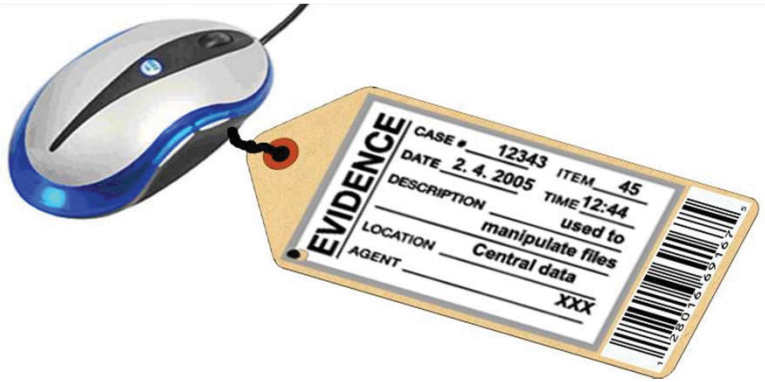
DATA = Digital Access Through Aeons



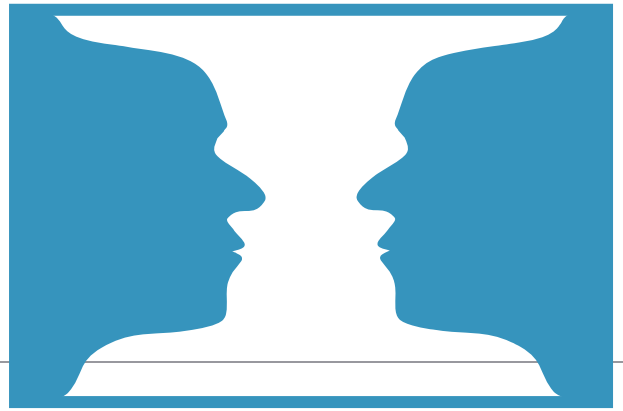
Read

Find

Trust



Understand



+ COMPLIANCE

Project partners

Nasjonalbiblioteket

Norsk Regnesentral

Utenriksdep.

Riksarkivet

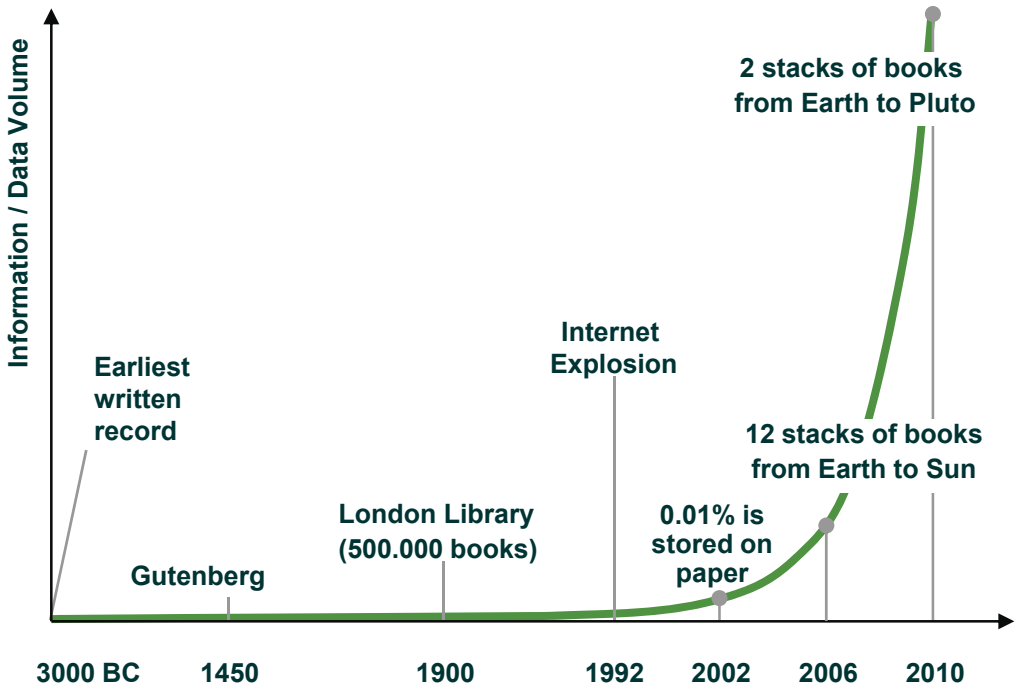
Brønnøysundregistrene

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

Digital Safe (Video)

http://research.dnv.com/longrec/digsafe.wmv

Information explosion



TREND

more and more is digital,
more and more is stored (temporarily),
more and more has to be documented,
more and more is non-textual

CHALLENGE

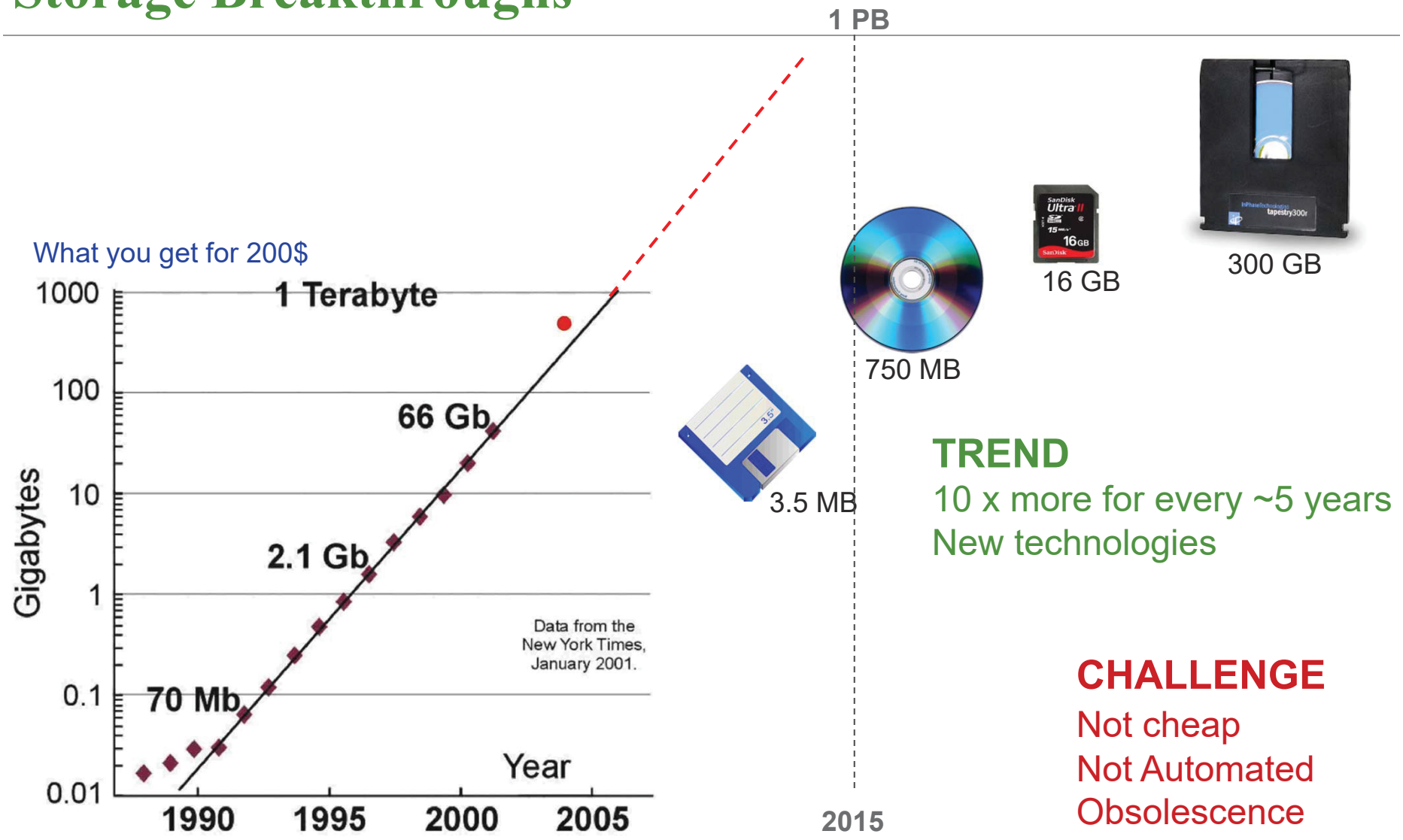
Too much to store
What to store
How to handle

2.1 Mega Pixel

8 yrs

12 Mega Pixel

Storage Breakthroughs



TREND

10 x more for every ~5 years

New technologies

CHALLENGE

Not cheap

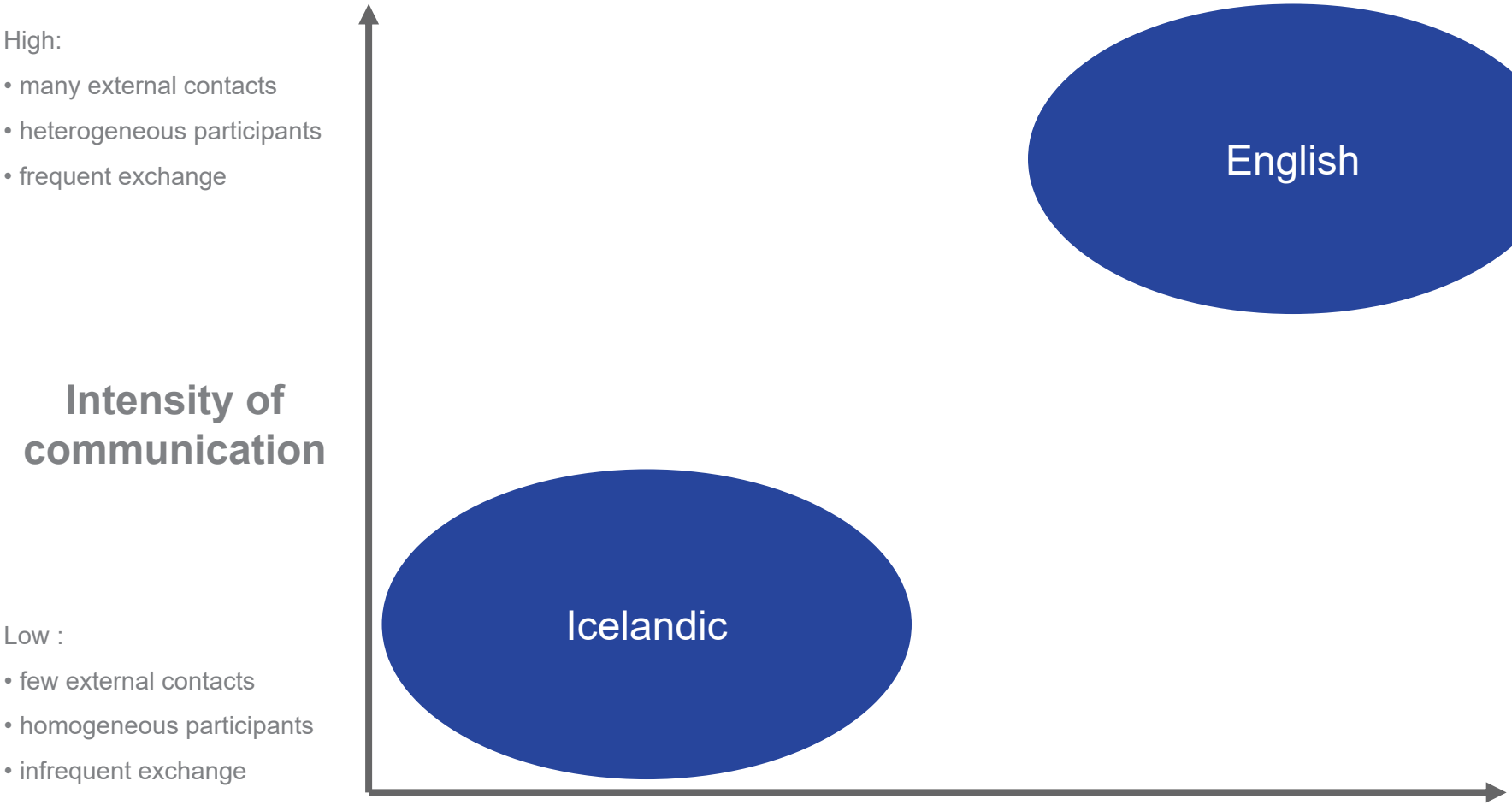
Not Automated

Obsolescence

Questions

- If the record is created digitally is that the original?
- Do we think that more and more originals will be in digital formats?
- Will there be more and more tools that creates digital formats?
- Is it important to capture preservation information when the digital record is created?
- Do we believe that we can preserve the museum solution HW/SW and information?
- Is it better to preserve the original in the archive than garbage?
- Etc.

The impact of communication on language change



The impact of globalization on changing information needs

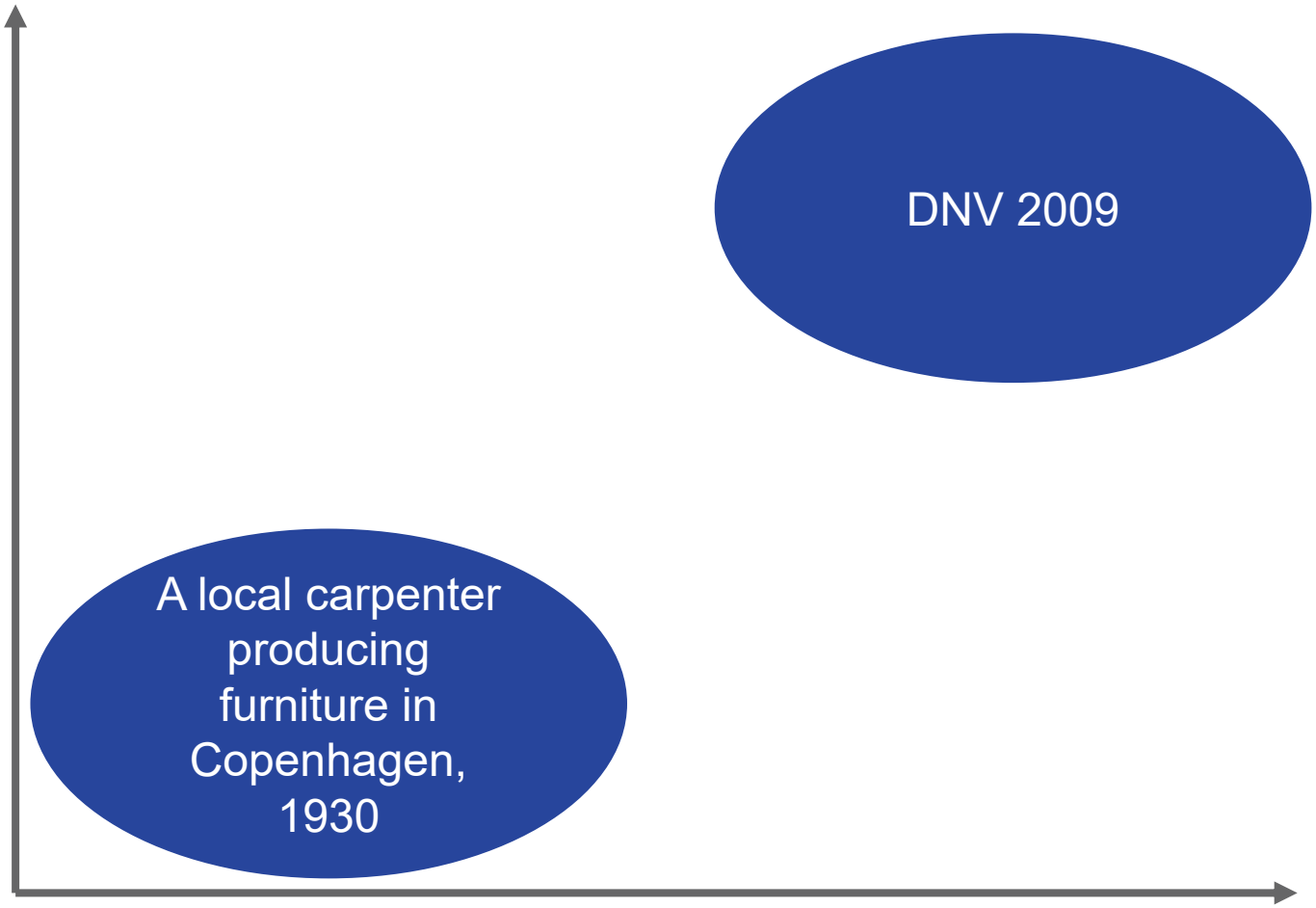
High:

- many international subcontractors
- complex, heterogenous customer base
- internationalized technological development

Degree of globalized business

Low :

- limited dependency on business partners
- no transnational, few local business partners
- homogenous and stable customer needs



Low:

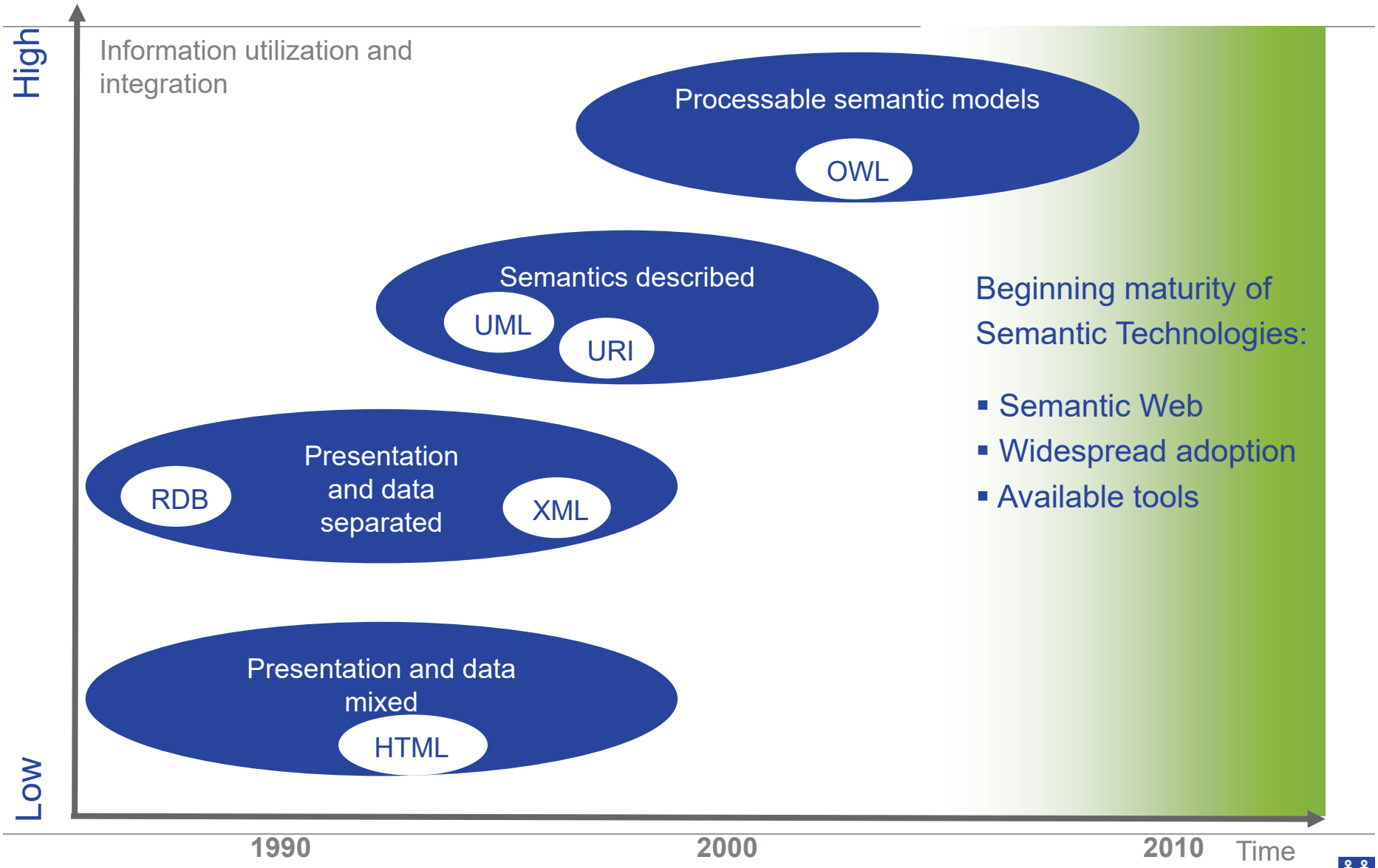
- slow development of new knowledge
- stable, predictable information needs

Degree of changes in information

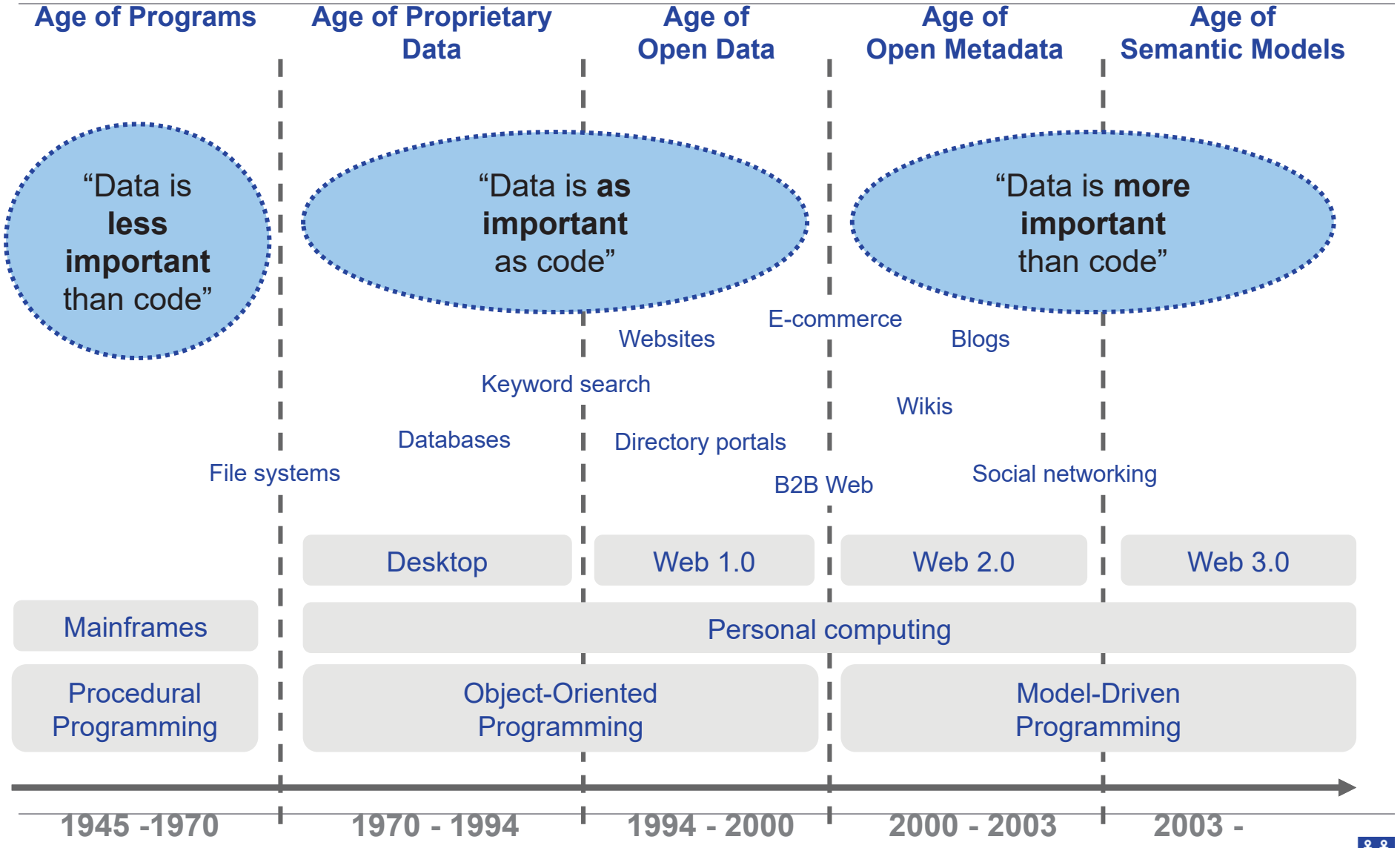
High:

- high change rate of knowledge
- volatile information needs

Maturity of the Semantic Technologies



Semantic Technologies: From code to meaning

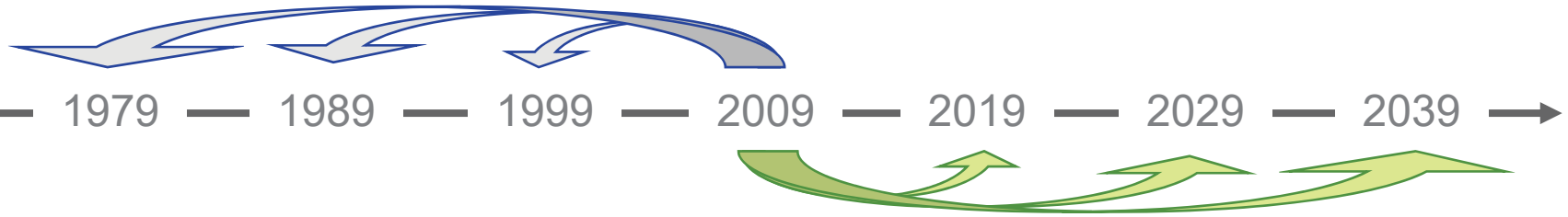


Inspired by: Nova Spivack (2007): How the WebOS evolves? See http://novaspivack.typepad.com/nova_spivacks_weblog/2007/02/steps_towards_a.html

Preserving semantic value: Two perspectives of time

1. Help information consumers understand historic data

information consumer



information provider

2. Help information providers make their information understandable over time

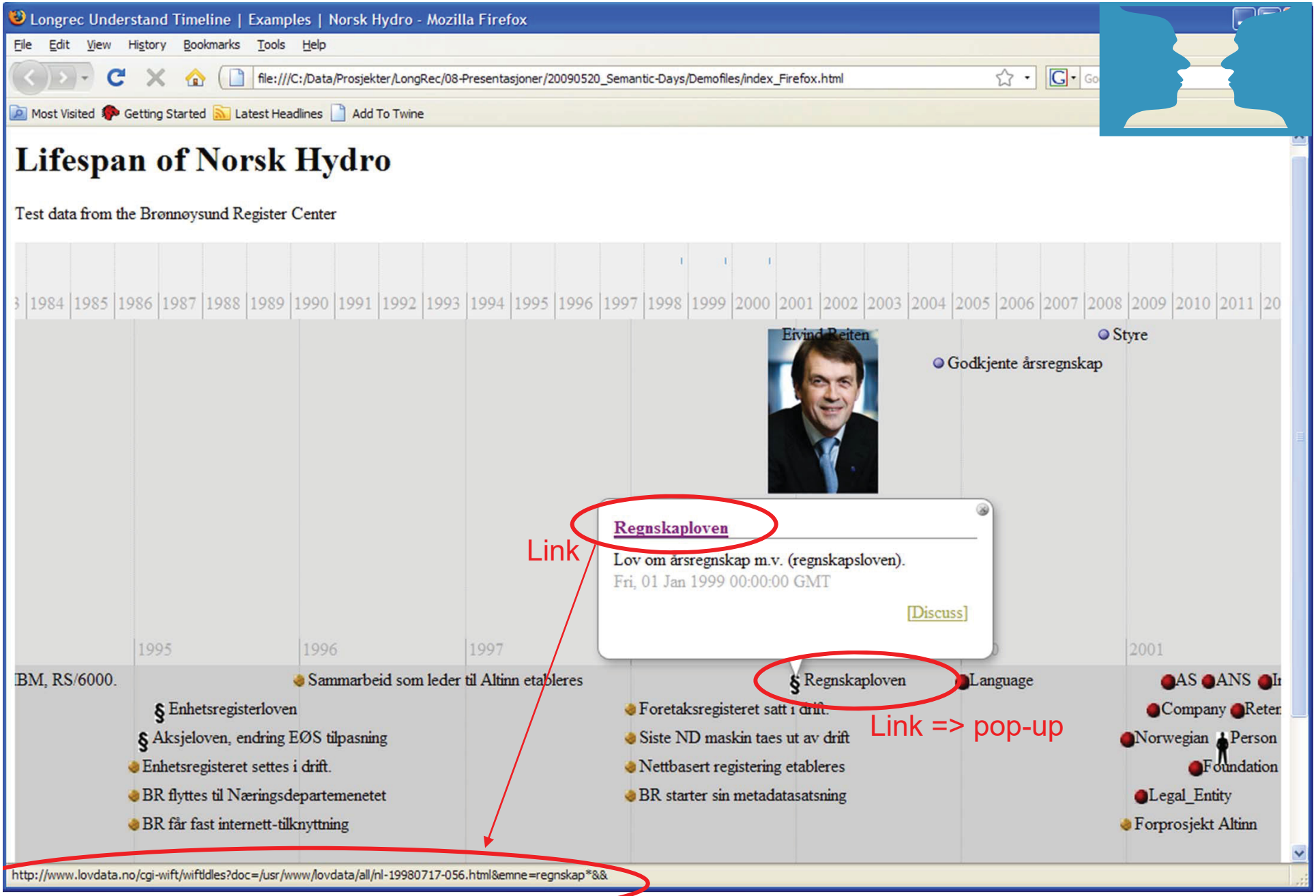
Empowering the knowledge worker (I)

Enable the interpretation of primary data over time by relating it to and presenting it with relevant secondary data

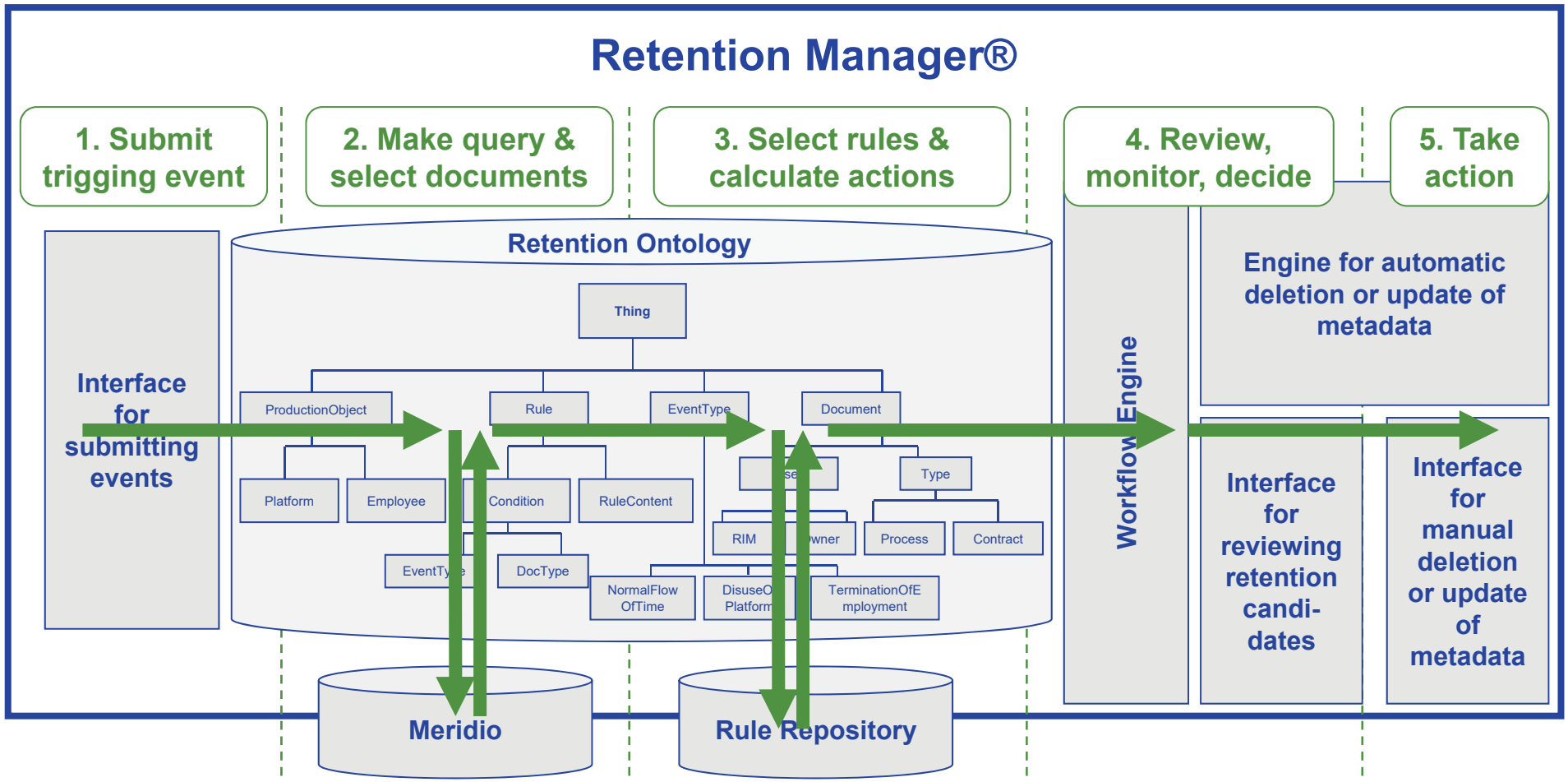
- Primary data = data in focus, e.g.
  - Norwegian business enterprise
  - Person holding a specific role in a Norwegian enterprise
- Secondary data = relevant context data, e.g.:
  - changes in law and jurisdiction
  - changes in BR's case practice
  - changes in language use (*styreforman* -> *styreleder*)

  (can be from both internal and external sources)

  - changes in ID used on referents, concepts, terms, records



Basic Scenario 1: Managing retention for a given event



”Disuse platform Oseberg”

”Employee Smith leaves the company

”Identify all documents related to Oseberg”

”Identify all documents related to Smith”

”Calculate all rules related to disuse of a platform AND the types of documents related to Oseberg”

”Calculate all rules related to termination of employment and types of documents related to Smith”

”Find RIM for Oseberg”

”Find RIM HR”

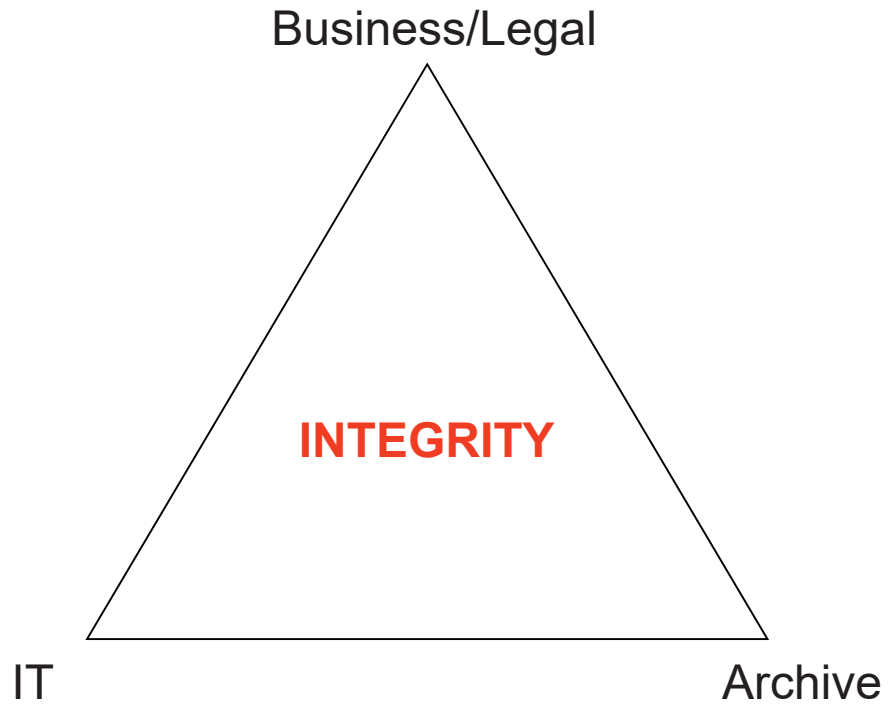
”Review Oseberg documents”

”Set deletion date for Smith-documents to Jan 1, 2019”

”Delete Oseberg-reports, but keep contract”

Competence needs automatic retention

- Archive
- Business
- Development
- Operation
- Semantics
- Legal
- Management risk evaluation
- Management risk control
- Security/access control
- Cryptation
- Migration
- Etc.

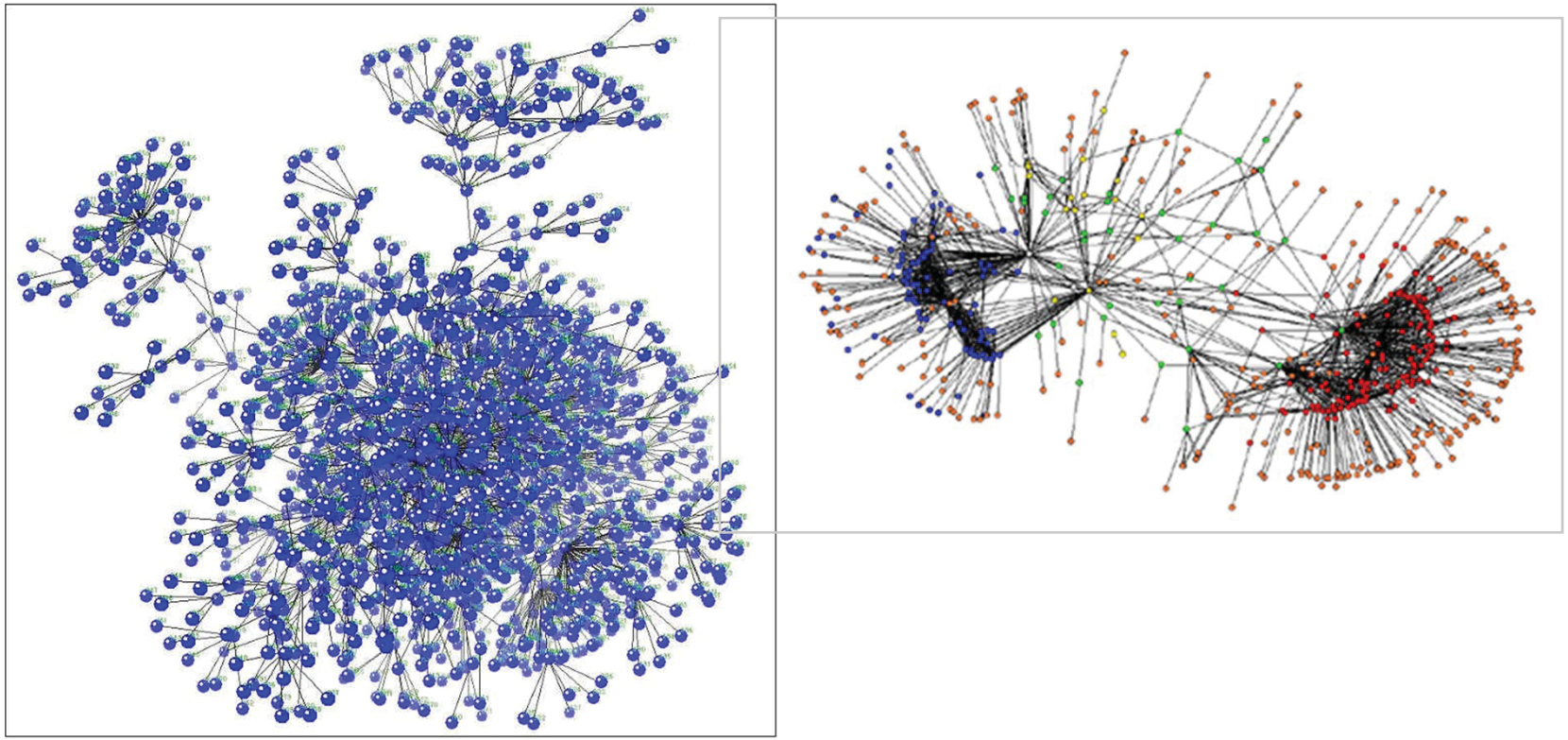


New Challenges

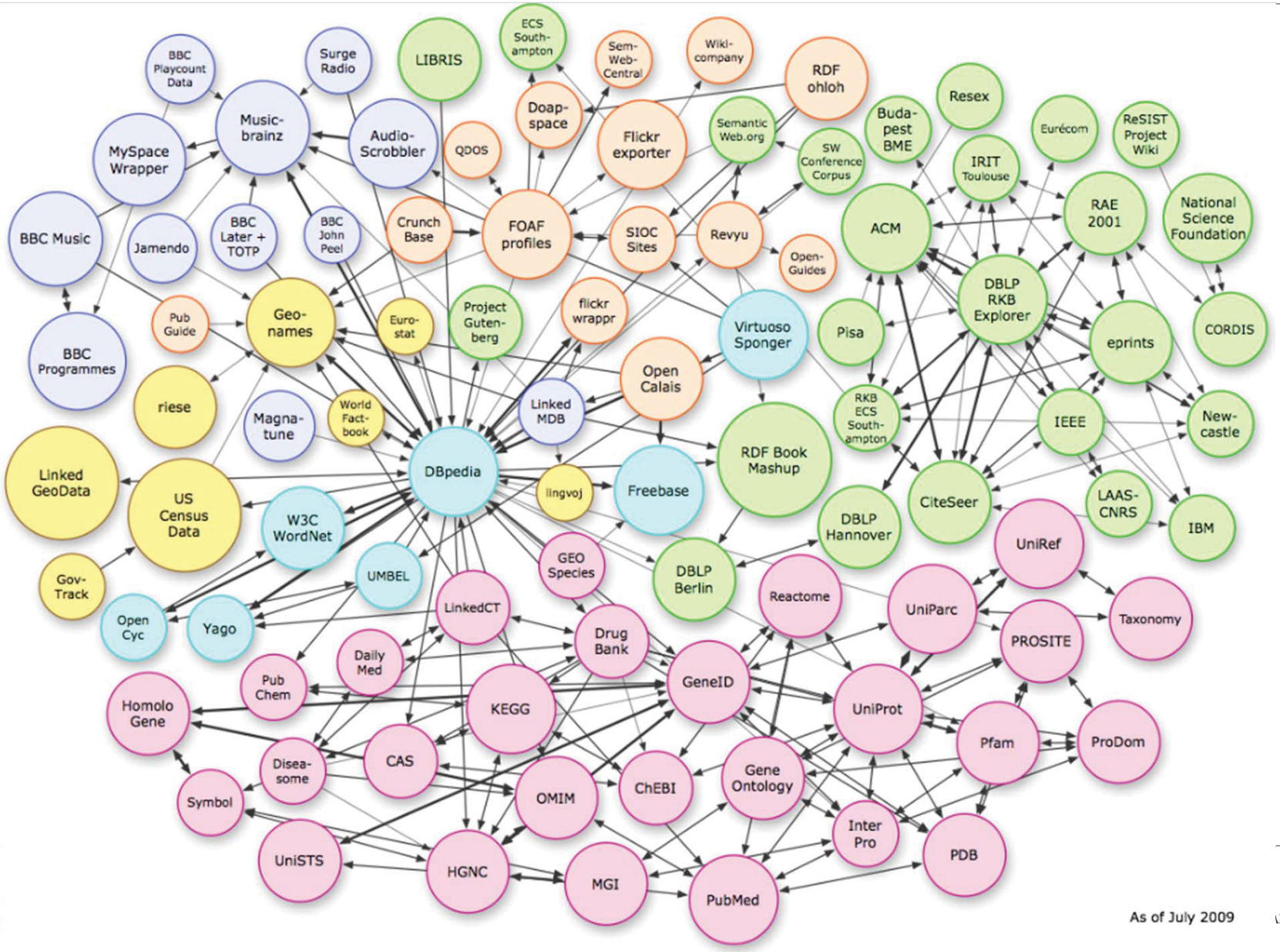
- Increased Volume
- We might not have enough disk space
- More and more compliance demands, even where deleting information is just as important as keeping the information
- More focus on information costs and benefits
- Critical users (response time, information about the information) – Interface
- Unawareness about companies needs related to keep original information
- Information that should be archived could be created at Twitter
- Etc

Ontology Evolution

- How to maintain/assess complex ontologies?



Linked Open Data Cloud



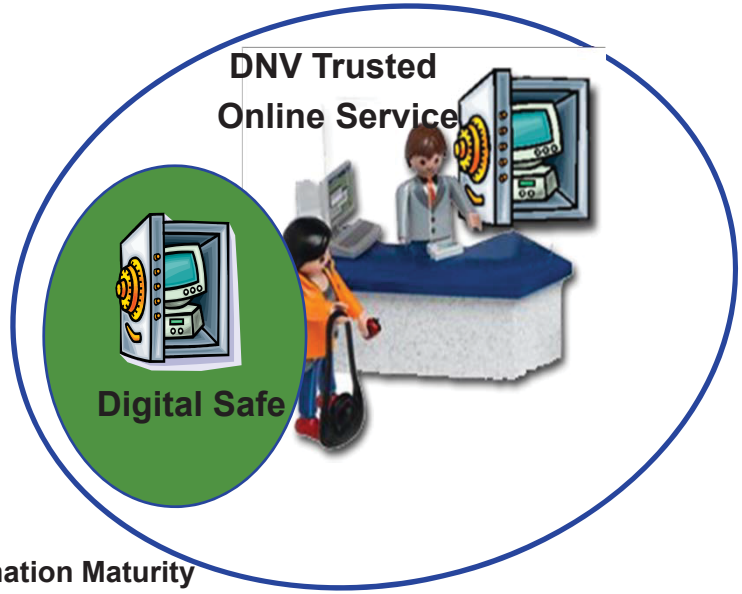
As of July 2009

Recommendation

- Identify the information your company needs to archive and what kind of security level your company needs
- Collect your demands related to preserve an original in eternity
- Talk with the business and the IT people and find out how you can deal with these challenges together
- Be positive to Open Source, everything that can make information handling more standardized
- Increase your companies risk awareness related to information handling
- **AND OF-COURSE USE SEMANTICS BOTH TO CAPTURE INFORMATION AND TO DO AS MUCH AS POSSIBLE AUTOMATICALLY!**

DNV are at the moment considering these services

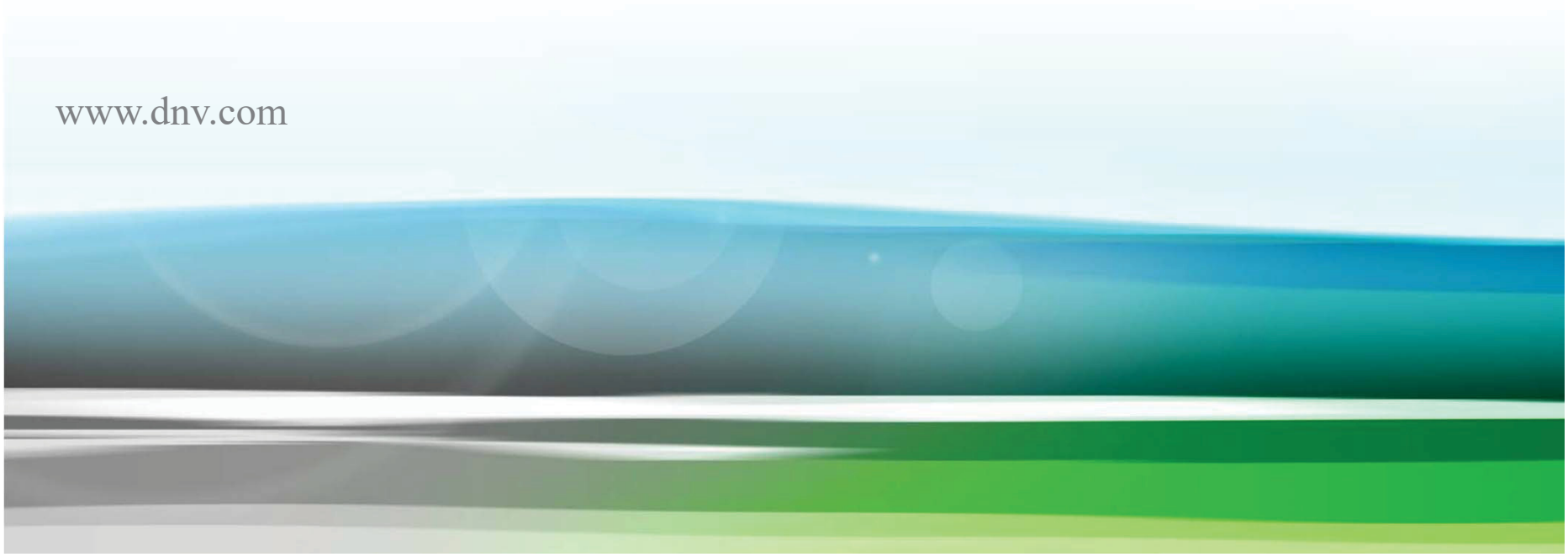
Information risk/analyse Management



Information Maturity Assessment Service

Safeguarding life, property and the environment

www.dnv.com



MANAGING RISK