MIXED

MIGRATION TO INTERMEDIATE XML FOR ELECTRONIC DATA

Preservation by Migration to XML

iPres - Beijing - 20071011

Dirk Roorda

Data Archiving and Networked Services

DANS

Data Archiving and Networked Services
DANS

MIXED
MIGRATION TO INTERMEDIATE XML FOR ELECTRONIC DATA

work on a preservation strategy

- positioning of the XML preservation strategy
- implementing the strategy in software
- pursuing a standard
- with international partners
  - welcome to MIXED

Data Archiving and Networked Services
DANS

MIXED
MIGRATION TO INTERMEDIATE XML FOR ELECTRONIC DATA

Dynamics of preservation

- when the digital context changes
  - emulate: re-implement data and tools
  - migrate: re-represent data, use new tools
- whatever you do, do it smart

Data Archiving and Networked Services
DANS

MIXED
MIGRATION TO INTERMEDIATE XML FOR ELECTRONIC DATA

To emulate or to migrate?

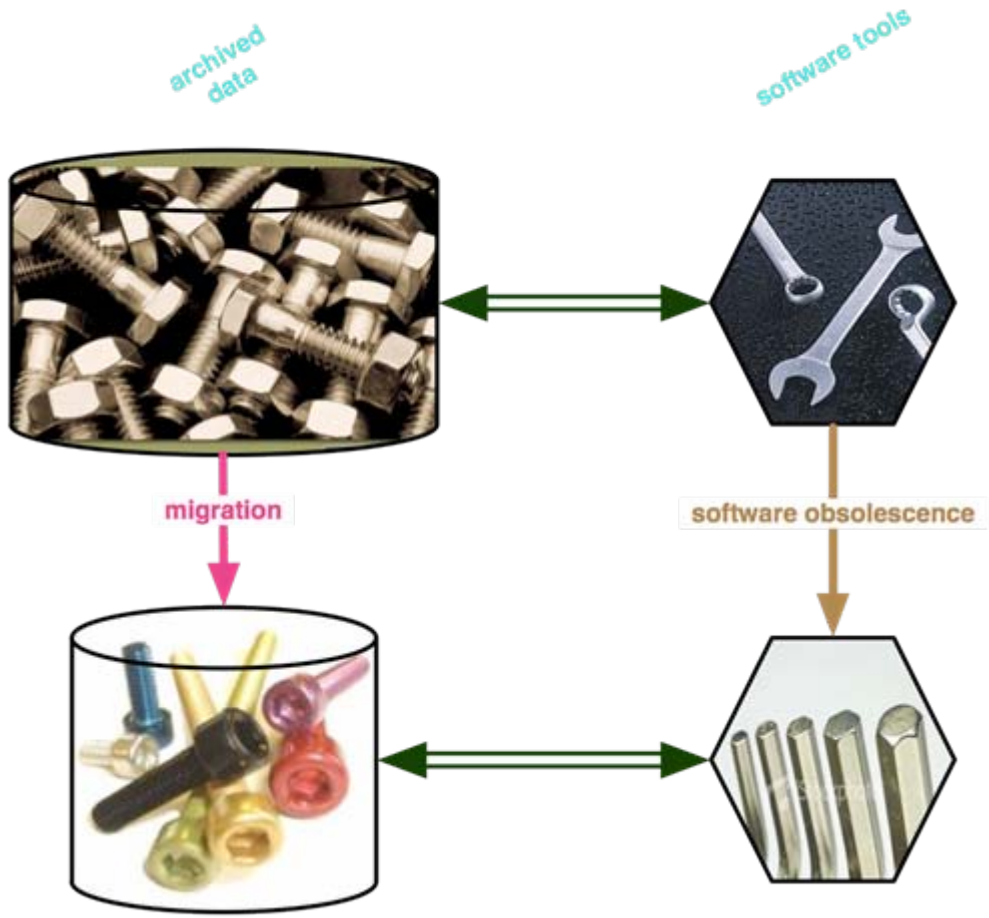
- archival material = data plus tools
- total recall ⇒ emulation
- new use of old data ⇒ migration

doing nothing is also an option

Data Archiving and Networked Services
DANS

MIXED
MIGRATION TO INTERMEDIATE XML FOR ELECTRONIC DATA

Data and tools



Data Archiving and Networked Services
DANS

MIXED
MIGRATION TO INTERMEDIATE XML FOR ELECTRONIC DATA

Smart strategies

- multiple related tasks?
- seek *normalization*
  - data representation: NF1-6
  - emulation: universal virtual computer
  - migration?

Data Archiving and Networked Services
DANS

MIXED
MIGRATION TO INTERMEDIATE XML FOR ELECTRONIC DATA

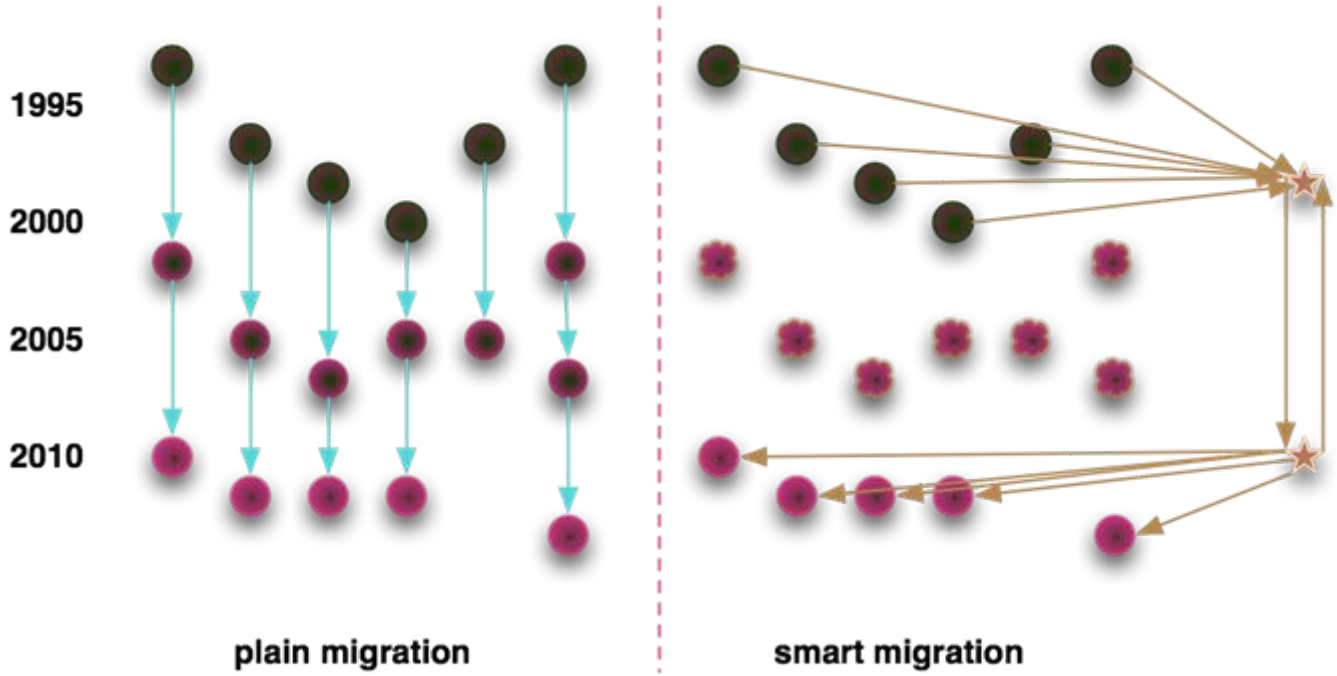
Diachronic and synchronic

- migration across time (diachronic)
  - original is nearly obsolete
  - original intention might be unclear
- migration within time (synchronic)
  - many different formats
  - vendor specific peculiarities

Data Archiving and Networked Services
DANS

MIXED
MIGRATION TO INTERMEDIATE XML FOR ELECTRONIC DATA

Smart migration



Data Archiving and Networked Services

DANS

MIXED

MIGRATION TO INTERMEDIATE XML FOR ELECTRONIC DATA

MIXED explained

Migration to
Intermediate
XML for
Electronic
Data

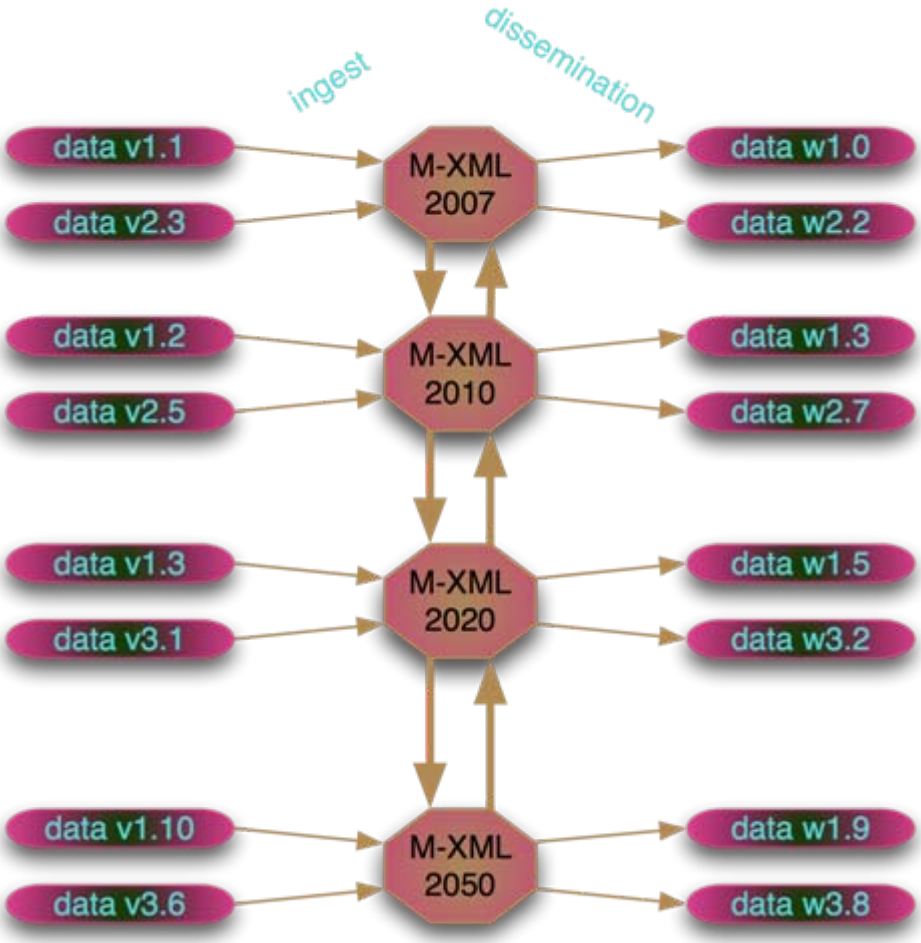
Data Archiving and Networked Services

DANS

MIXED

MIGRATION TO INTERMEDIATE XML FOR ELECTRONIC DATA

MIXED scenario



Data Archiving and Networked Services
DANS

MIXED
MIGRATION TO INTERMEDIATE XML FOR ELECTRONIC DATA

Kinds of data

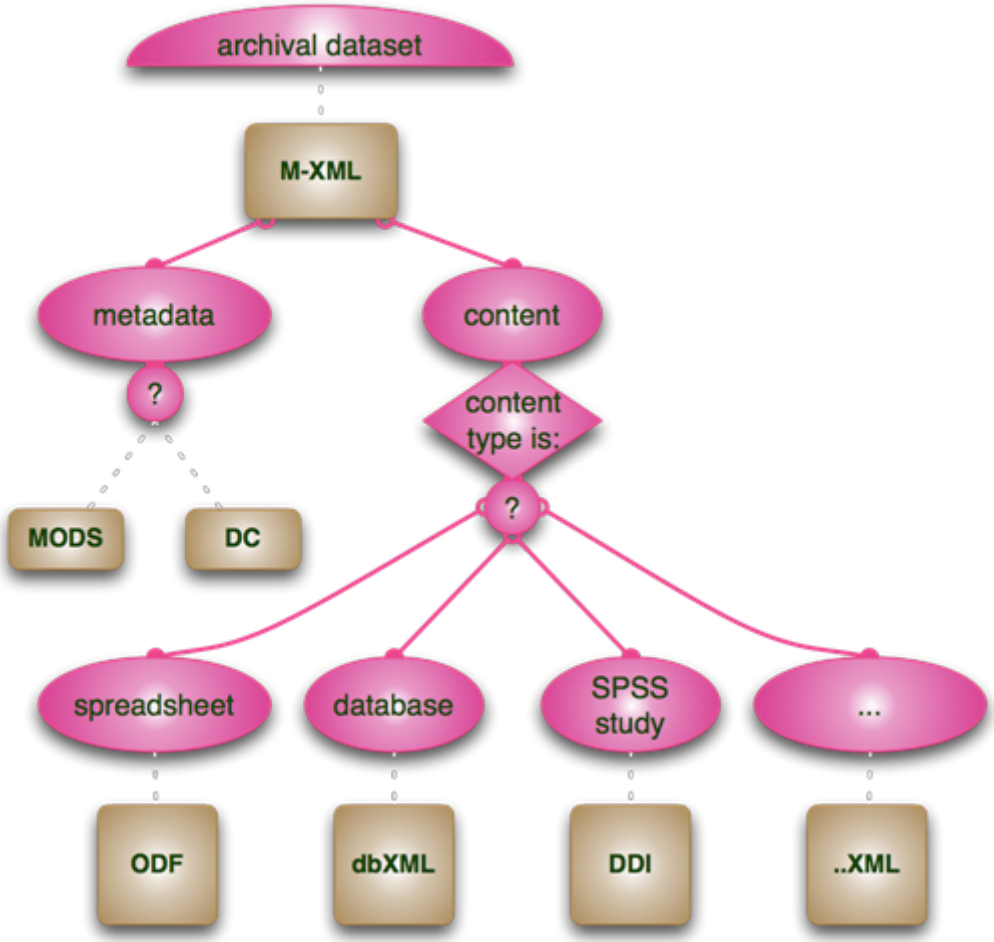
- documents
- spreadsheets
- databases
- statistical data
- images

- word, open(red)office
- excel, openoffice
- access, mysql
- spss, sas
- photoshop, irfanview

Data Archiving and Networked Services
DANS

MIXED
MIGRATION TO INTERMEDIATE XML FOR ELECTRONIC DATA

Umbrella format



Data Archiving and Networked Services
DANS

MIXED
MIGRATION TO INTERMEDIATE XML FOR ELECTRONIC DATA

Making it work

- software
  - = framework
    - + modules
- standard
  - = wrapper
    - + metadata
    - + for each kind of data:
      - selected XML standard for that kind

Data Archiving and Networked Services
DANS

MIXED
MIGRATION TO INTERMEDIATE XML FOR ELECTRONIC DATA

Making software

- make a good product as initial effort
  - generic framework
  - substantial number of conversion plugins
    - for spreadsheets and databases
    - for statistical data
  - connect to repositories, Fedora ready
- integrate efforts of all interested parties by
  - using an open architecture
    - webservices for framework and plugins
    - use third party plugins for SPSS and DDI
  - using an open source paradigm

Data Archiving and Networked Services
DANS

MIXED
MIGRATION TO INTERMEDIATE XML FOR ELECTRONIC DATA

Helping a standard emerge

- finding a name: Preferred Data Formats for Archives (PDFA) suggestions welcome
- using existing auxiliary standards
- connecting to open source software
- seeking a user base in the archiving world

Data Archiving and Networked Services
DANS

MIXED
MIGRATION TO INTERMEDIATE XML FOR ELECTRONIC DATA

Using MIXED

- repositories: preservation planning
- interested in file conversion: web services
- individual users: stand alone

Data Archiving and Networked Services
DANS

MIXED
MIGRATION TO INTERMEDIATE XML FOR ELECTRONIC DATA

Using standards

- XML (Schema, UNICODE)
- OAIS (interface to repositories)
- ODF (spreadsheets)
- SOAP, ESB
- Java, SPRING, OSGI

Data Archiving and Networked Services
DANS

MIXED
MIGRATION TO INTERMEDIATE XML FOR ELECTRONIC DATA

Co-operation

- DExT (Data Exchange Tools) (UKDA)
- ODaF (Open Data Foundation)
- you?

Data Archiving and Networked Services
DANS

MIXED
MIGRATION TO INTERMEDIATE XML FOR ELECTRONIC DATA

Trends

- end of vendor-specific binary formats in sight
- interchange formats more concerned with preservation

Data Archiving and Networked Services
DANS

MIXED
MIGRATION TO INTERMEDIATE XML FOR ELECTRONIC DATA

The future ...

- the major data kinds have a preferred preservation format
- the set of preservation formats is standardized
- easy-to-use software converts between preservation formats and custom formats

Data Archiving and Networked Services
DANS

MIXED
MIGRATION TO INTERMEDIATE XML FOR ELECTRONIC DATA

The end of MIXED?

- if applications can import/export preservation formats directly
- - then no more synchronic conversions
- diachronic conversions keep developing
- - as long as the preservation formats change

Data Archiving and Networked Services
DANS

MIXED
MIGRATION TO INTERMEDIATE XML FOR ELECTRONIC DATA

Questions ...

- can we standardize preservation formats per data kind?
  - is 1 format per kind sufficient?
- is the time right for an XML standard for data preservation?
- what is the most usable form for the MIXED software?