



## Testbed

Maurice van den Dobbелsteen

Senior advisor digital longevity

*National Archives of The Netherlands*



---

**□ What can the Testbed do for you?**

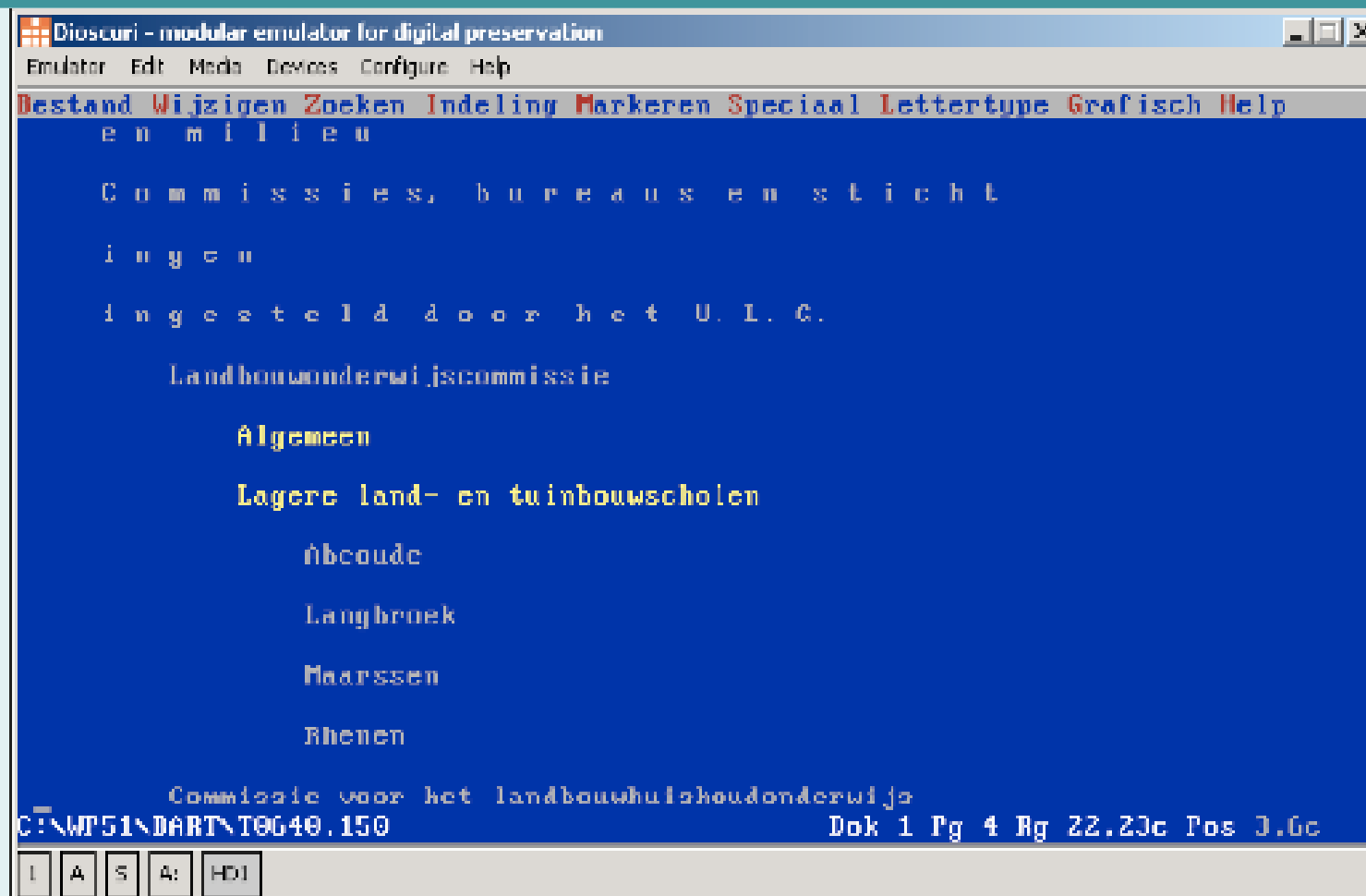


# Questions

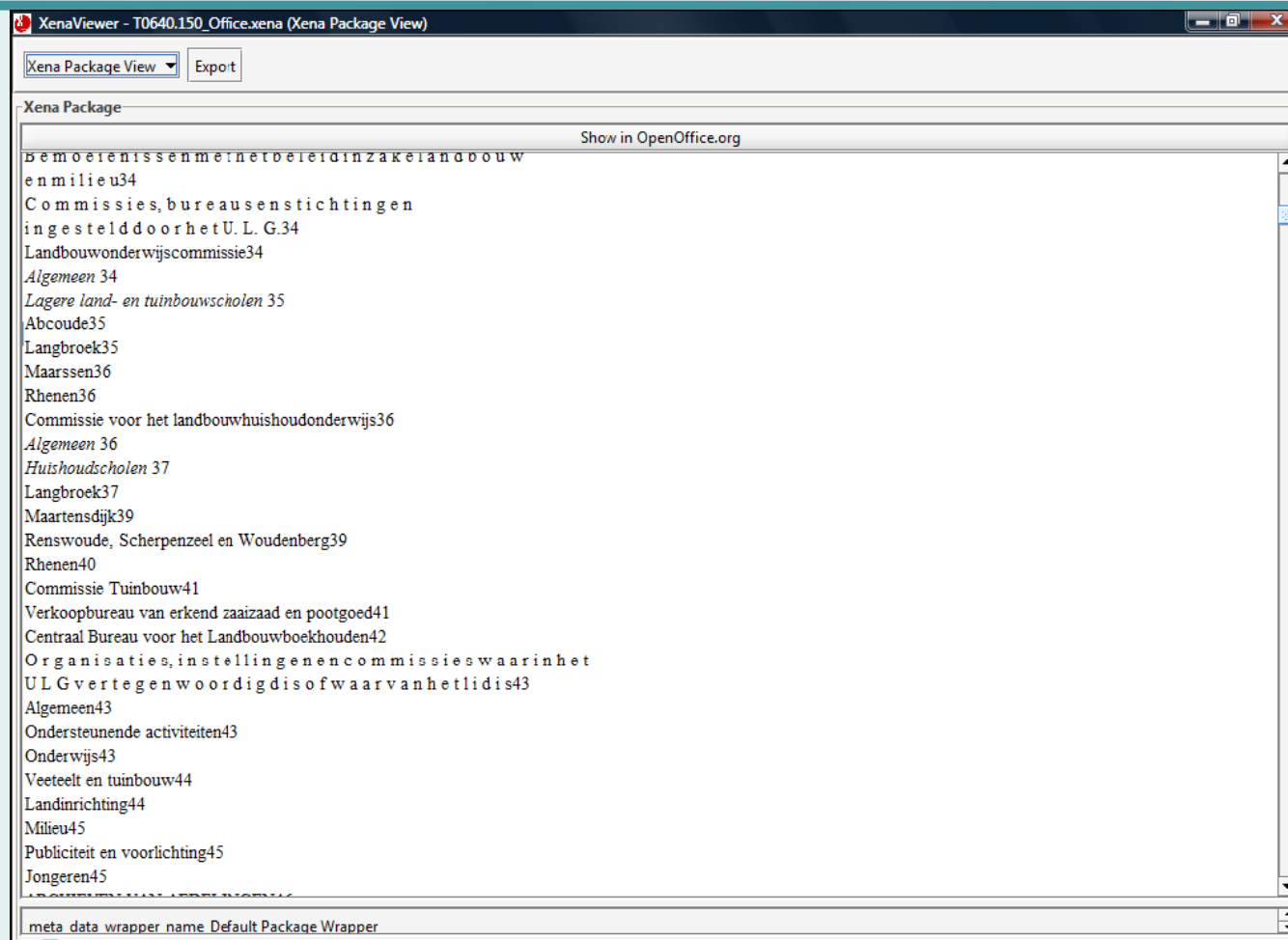
- ❑ What file formats do I have in my collection?
- ❑ Some file formats are becoming obsolete. I want to migrate them. But to what file format? And with which migration tool?
- ❑ Can I preserve the original with an emulation service?



# Example 1 – Original document



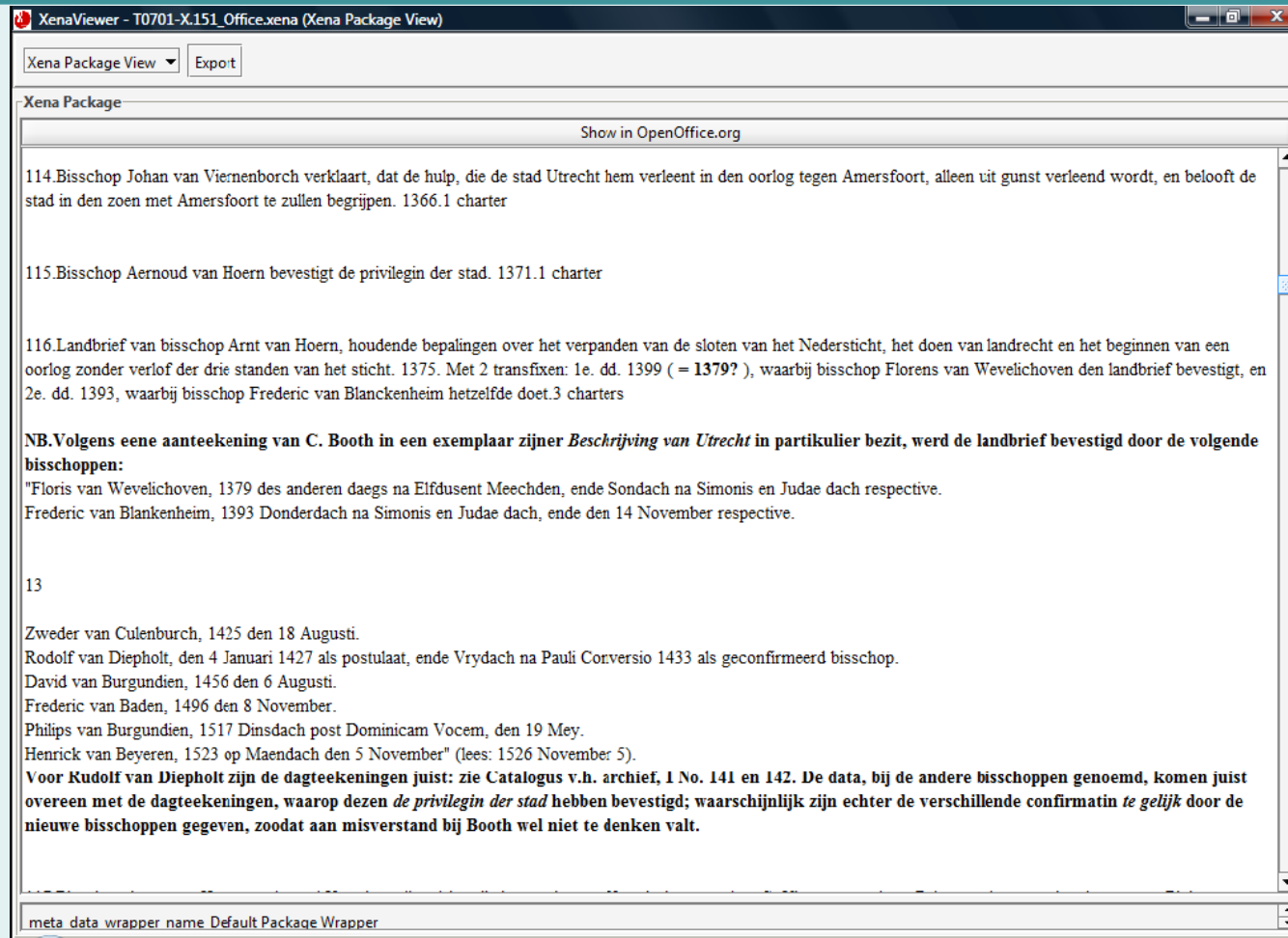
# Example 1 – Migrated file



## Example 2 – Original file



## Example 2 – Migrated file



# Goal of the Testbed

---

- ❑ You will need tools
- ❑ Tools have different performances
- ❑ Find out in the Testbed
- ❑ Many experiments give a good average!
- ❑ Information on tools enriched with experimental information





# Objective - No value judgements

---

- ❑ Whether it is useable for you in your specific situation, is up to you to decide
  - Is it ok to keep the text but lose the layout?
- ❑ In the Testbed, the performance of the tool is captured - "objectively"
  - You can use your neighbor's results



# □ What is the Testbed?



# Testbed components - Application

- ❑ Web application, software
- ❑ PLANETS runs a Central Instance which you can access online (plug & play)

## Testbed MVC architecture

web tier



JSF, CSS, XHTML, AJAX

business tier



Java EE (EJB, JAX-WS,...)

data tier



Java EE, IF Registries,...



# Preservation questions and services

---

- ❑ What file format is this file?
  - pdf, .xls etc.
  - **Identification** service
- ❑ Is this file well-formed?
  - yes, no
  - **Validation** service
- ❑ What are the values of certain properties?
  - font size = 11, image height = 135
  - **Characterisation** service
- ❑ Migrate this file for me into another file format
  - .doc >> .pdf/a
  - **Migration** service
- ❑ I want to see this file; emulate an old environment for me
  - **Emulation** service



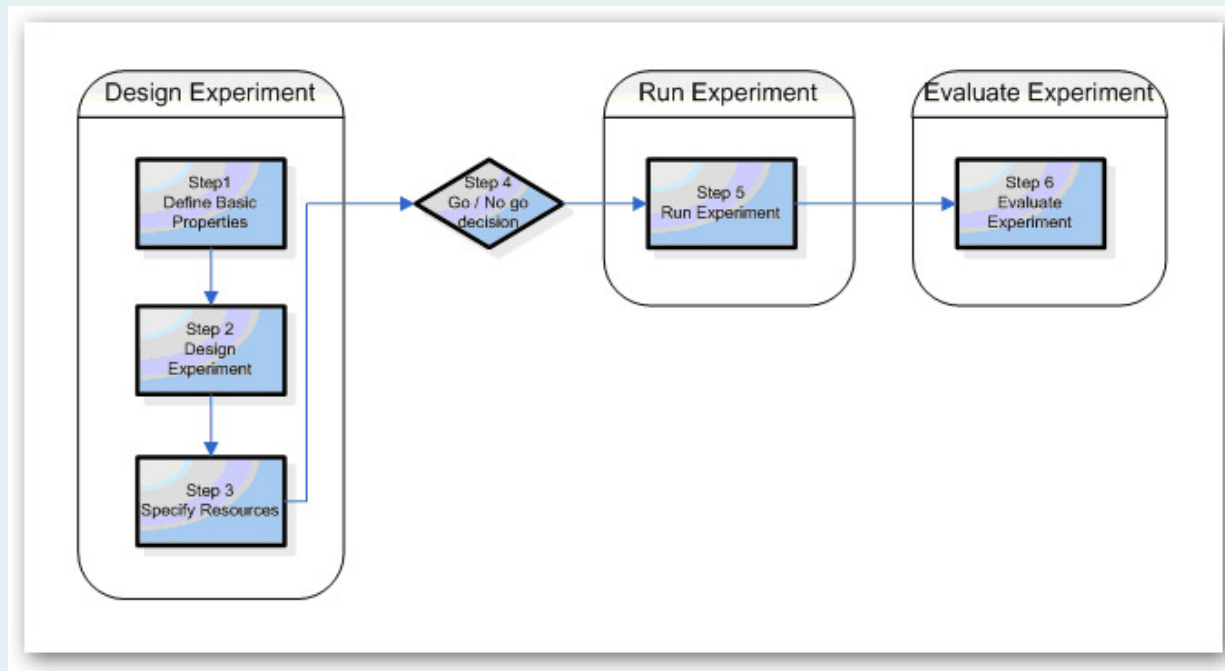
# Testbed components - Services

- ❑ Wrap tools into services  
Services, put them up in  
the IF and use them in  
the Testbed
- ❑ Examples:
  - JHOVE
  - PS2PDF
  - ImageMagik
  - Sanselan
  - MSWord migration
  - HTMLCleaner



# Testbed components - Workflow

- ❑ Every experiment follows a fixed experimentation workflow that consists of six steps:



# Testbed components - Ontology

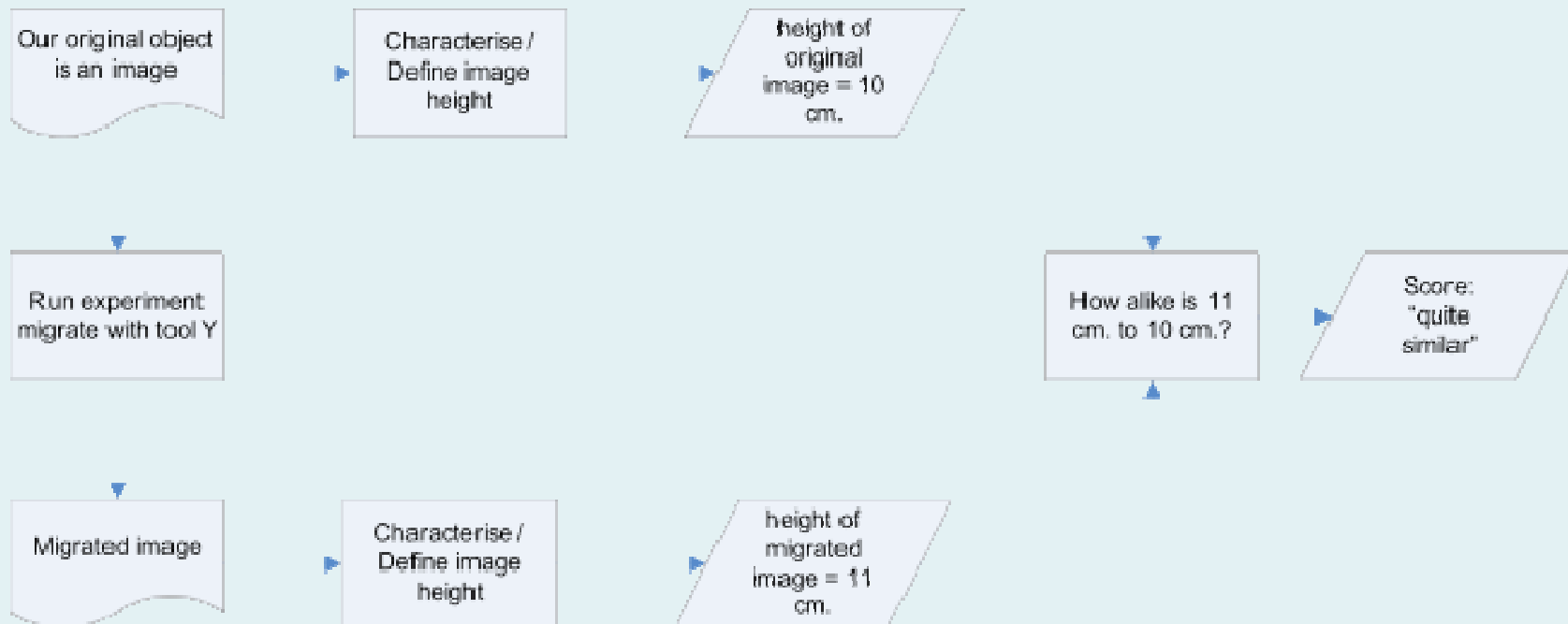
---

- ❑ Collection of digital object properties that might be of interest (making them 'significant' properties):
  - Image
    - Height
    - Width
    - Color space
  - Text
    - Font
    - ...
- ❑ Compare particular objects before/after treatment (evaluation criteria, benchmark goals)



# Example – Use of properties

## Example: test of migration tool Y





# Testbed components - Extractor

---

- ❑ Automatic extraction of properties of input and output files
- ❑ Using Planets XCDL language
- ❑ Developed by University of Cologne



# Testbed components - Corpora

---

- ❑ Constructed „known“ files (annotated)
  - Supported and maintained by UzK Cologne
  - Wide range of different files and formats
- ❑ „Real“ files provided by all partners



# Testbed benefits

---

- ❑ Controlled hardware & software environment
- ❑ A consistent methodology
- ❑ Structured processes for arrangement and evaluation of experiments
- ❑ Structured properties ontology
- ❑ Structured test data - corpora
- ❑ Experiments reproducible



# The Testbed and who?

---

- ❑ Through its extensible structure and tight connection to practitioners and research Testbed is a valuable application for a wide range of partners,
- ❑ such as...



# The Testbed and ...

---

- ❑ ... heritage institutions and content holders
- ❑ Have a look into representative simulation of problems without exposing own material by using Corpora
- ❑ Get information on behaviour of tools at a „one-stop-shop“
- ❑ May use a set of readily available digital content for experiments
- ❑ Upload own files for special needs



# The Testbed and ...

---

- ❑ ... tool and service providers
- ❑ Helps to understand needs of user community.
- ❑ Test tools and see if they behave the way expected



# The Testbed and ...

---

- ❑ ... the digital community
- ❑ Find knowledge, expertise and people involved in digital preservation
- ❑ Become a practical and active part of the preservation community



# Testbed time scale

---

- ❑ 2nd half 2009: Involve external institutions
- ❑ Peer2peer system of experiments
- ❑ Together evaluate new functionality
- ❑ Incorporate feedback into the Testbed
- ❑ Q1 2010: Testbed external release





# Who to contact?

---

- ☐ Register without any obligations on Planets website
- ☐ [www.planets-project.eu/community](http://www.planets-project.eu/community)
- ☐
- ☐ Central contact: The Testbed Helpdesk
- ☐ [helpdesk@planets-project.eu](mailto:helpdesk@planets-project.eu)

Maurice van den Dobbelsteen  
Nationaal Archief of The Netherlands  
[maurice.van.den.dobbelsteen@nationaalarchief.nl](mailto:maurice.van.den.dobbelsteen@nationaalarchief.nl)

