



Testbed

Petra Helwig

Senior advisor digital longevity

National Archives of The Netherlands

What can the Testbed do for you?

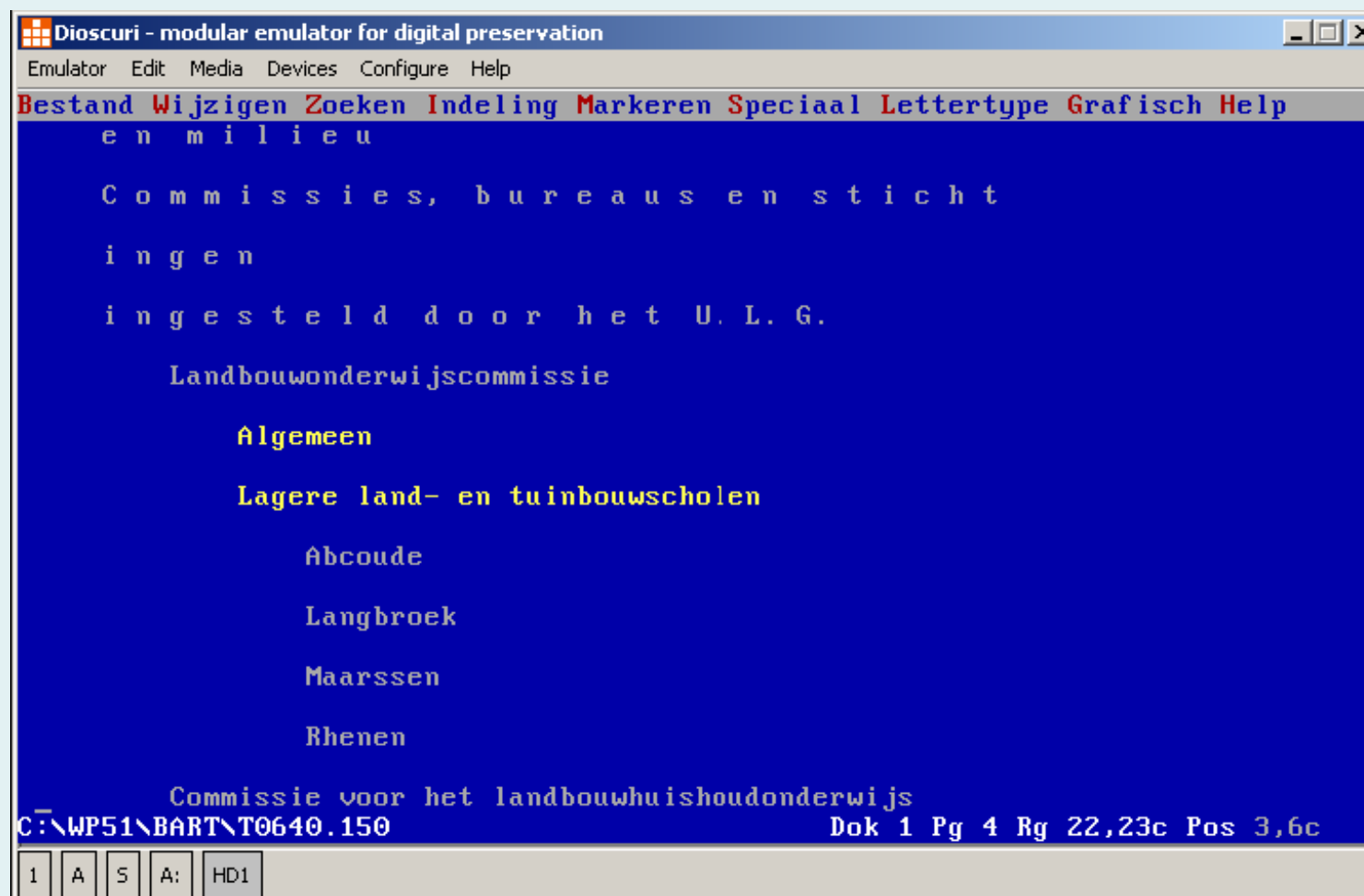


Questions

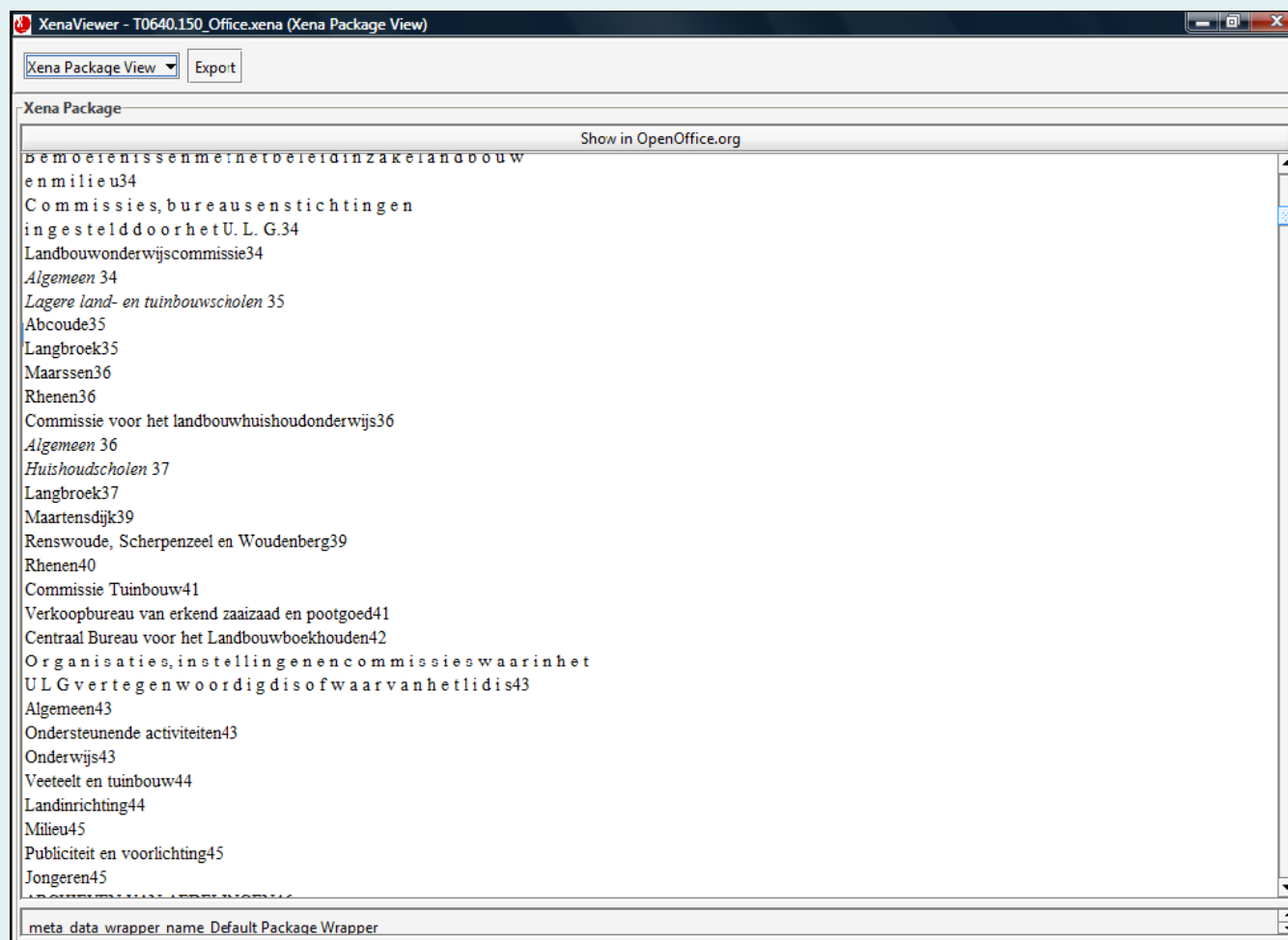
- ❑ What file formats do I have in my collection?
- ❑ Some file formats are becoming obsolete. I want to migrate them. But to what file format? And with which migration tool?
- ❑ Can I preserve the original with an emulation service?



Example 1 – Original document



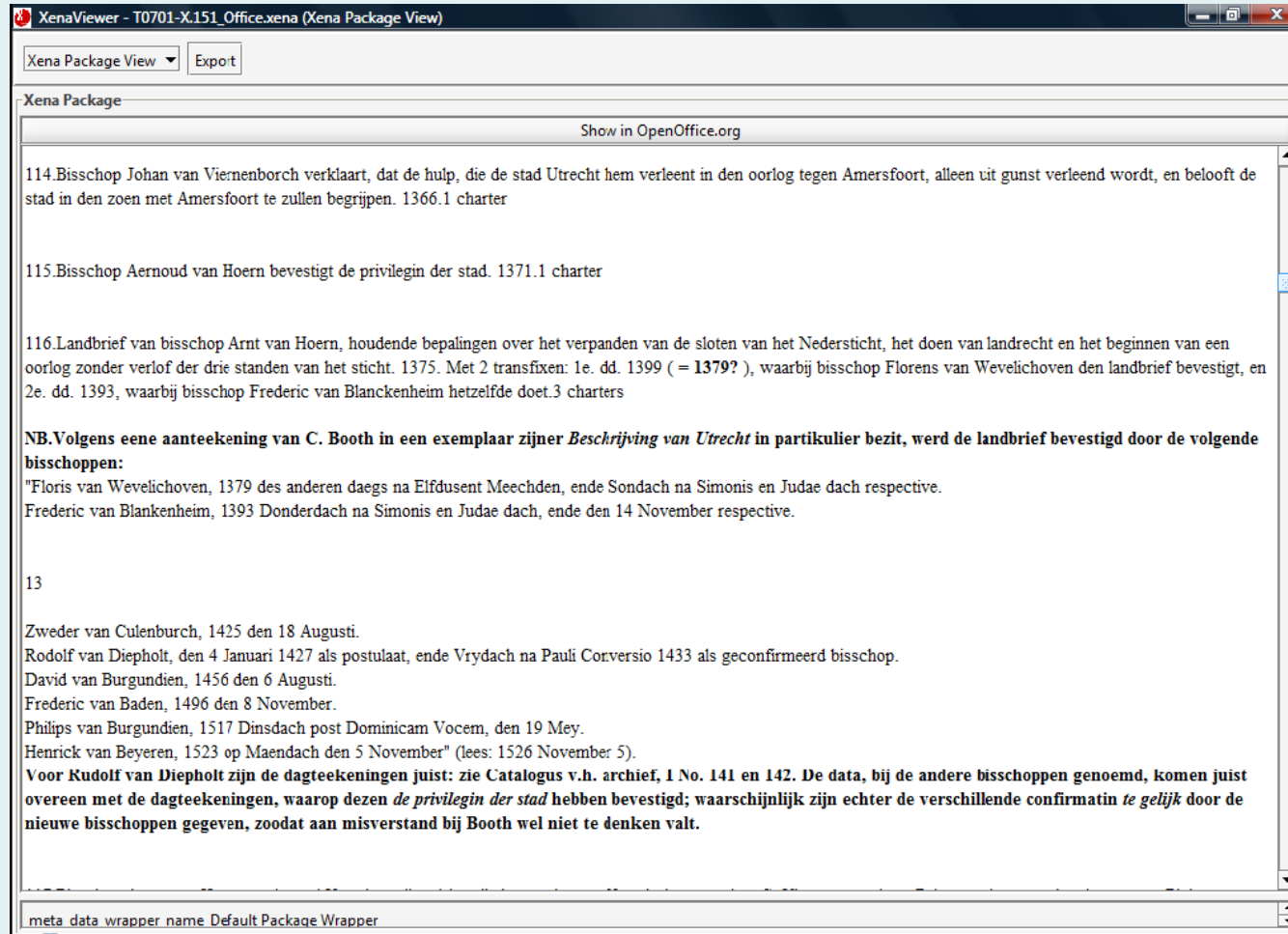
Example 1 – Migrated file



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!
- ❑ So information on tools can be enriched with experimental information



No judgement

- ❑ 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 neighbour's results



What is the Testbed?



Testbed components - Application

- ❑ Web application
- ❑ Runs at Humanities Advanced Technology And Information Institute (HATII), Glasgow

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



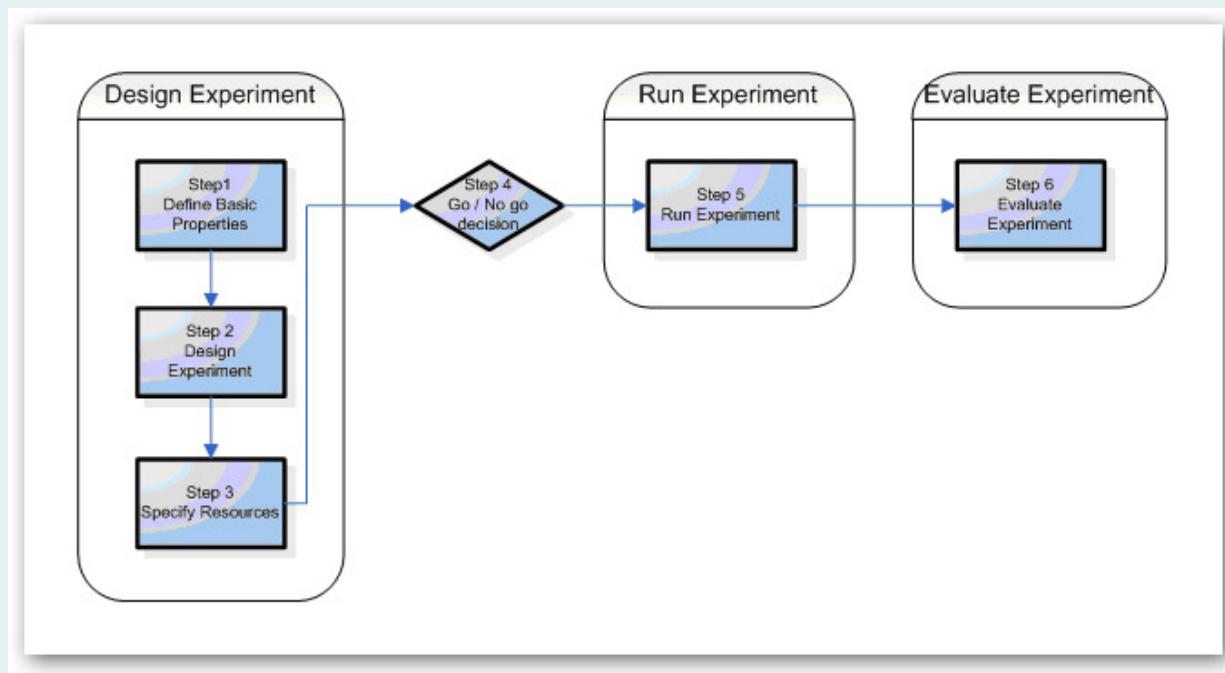
Testbed components - Services

- ❑ Services (“wrapped” tools) for
 - Identification
 - Validation
 - Characterisation
 - Migration
 - Emulation
- ❑ e.g.
 - JHOVE
 - PS2PDF
 - ImageMagik
 - Sanselan
 - MSWord migration
 - HTMLCleaner



Testbed components - Workflow

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



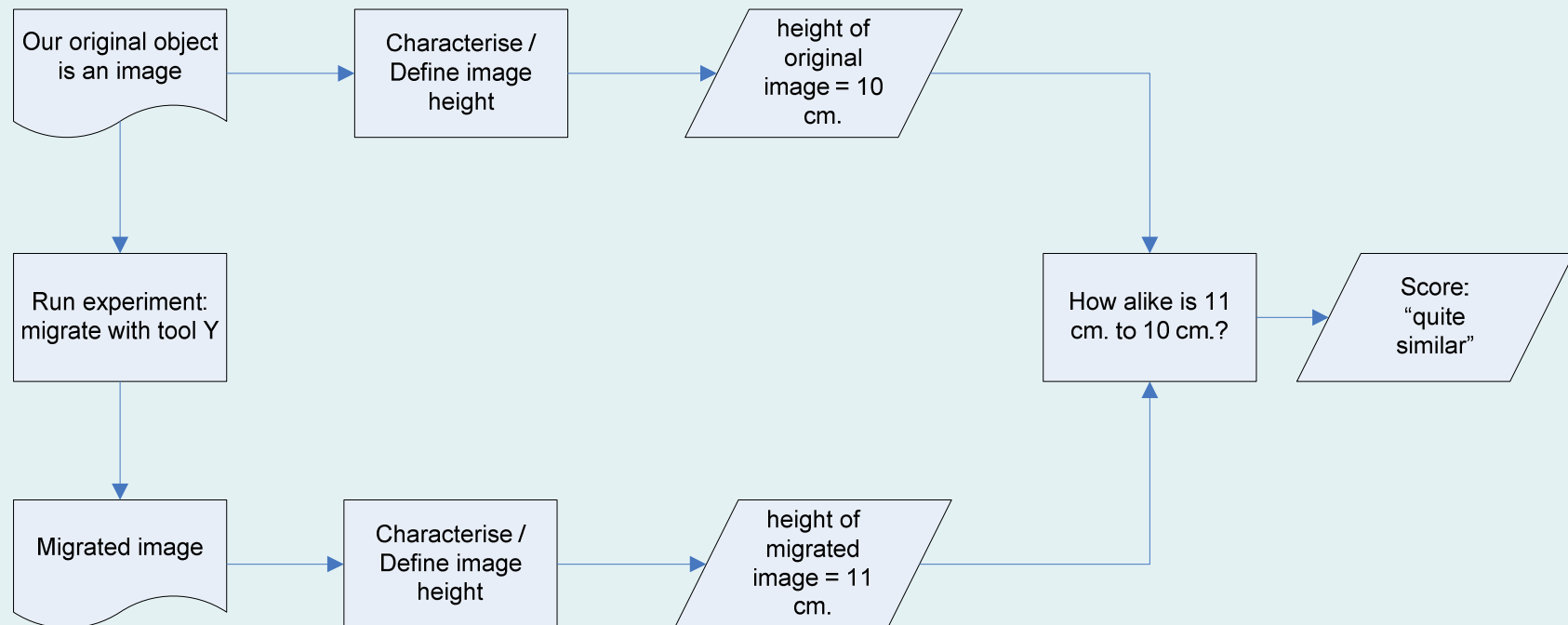
Testbed components - Ontology

- ❑ Collection of properties that might be of interest, e.g.
 - Image
 - Height
 - Width
 - Color space
 - Text
 - Font
 - ...
 - Database
 - ...
- ❑ To be able to compare particular objects before/after treatment (evaluation criteria)



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 advantages

- ❑ Controlled environment
- ❑ A consistent methodology
 - Structured documentation of tests and results
- ❑ Structured processes for arrangement and evaluation of experiments
- ❑ Structured test data (corpora)
- ❑ Experiments reproducible



A bit of marketing



The Testbed and ...

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



Testing – Time scale

- ❑ Summer 2009: Go public for external institutions
 - Early summer 2009: Testing by a small number of external institutions
 - Autumn 2009: Mass experimentation by selected partners
- ❑ 2010: Full external release



Interested?

- ❑ Interested in Testbed?

- > Register without any obligations on Planets website

- www.planets-project.eu/community

- ❑ Carrying out experiments?

- > contact Testbed Helpdesk

- helpdesktb@planets-project.eu

- * information on becoming an active partner

- * further information on training events, community,...

- ❑ Further questions: max.kaiser@onb.ac.at

