

The PLANETS-Ontology in the context of the PLANETS-Testbed and the XCL-Software

PLANETS (Introduction)

- Preservation and Long-term access via Networked Services

PLANETS (Introduction)

- Preservation and Long-term access via Networked Services
- Funded by the EC

PLANETS (Introduction)

- Preservation and Long-term access via Networked Services
- Funded by the EC
- 4 year project until June 2010

PLANETS – PARTNERS I



Koninklijke Bibliotheek



STATSBIBLIOTEKET

Österreichische
Nationalbibliothek



DET KONGELIGE BIBLIOTEK
NATIONALBIBLIOTEK OG KØBENHAVNS UNIVERSITETSBIBLIOTEK



the national archives



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

nationaal archief

- The British Library
- National Library, Netherlands
- Austrian National Library
- State and University Library, Denmark
- Royal Library, Denmark
- National Archives, UK
- Swiss Federal Archives
- National Archives, Netherlands

PLANETS – PARTNERS II



rechenzentrum
universität freiburg



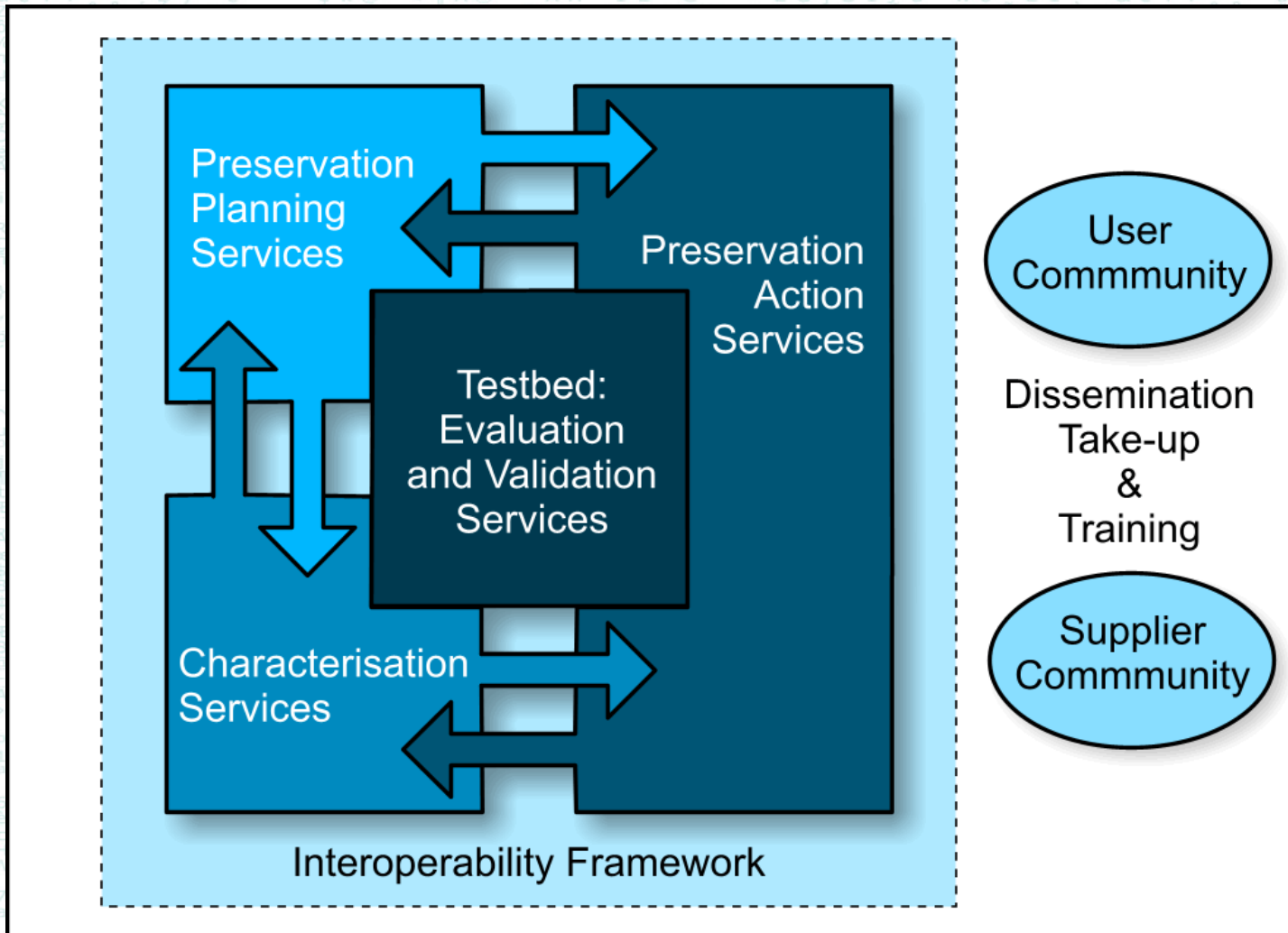
TECHNISCHE
UNIVERSITÄT
WIEN
VIENNA
UNIVERSITY OF
TECHNOLOGY



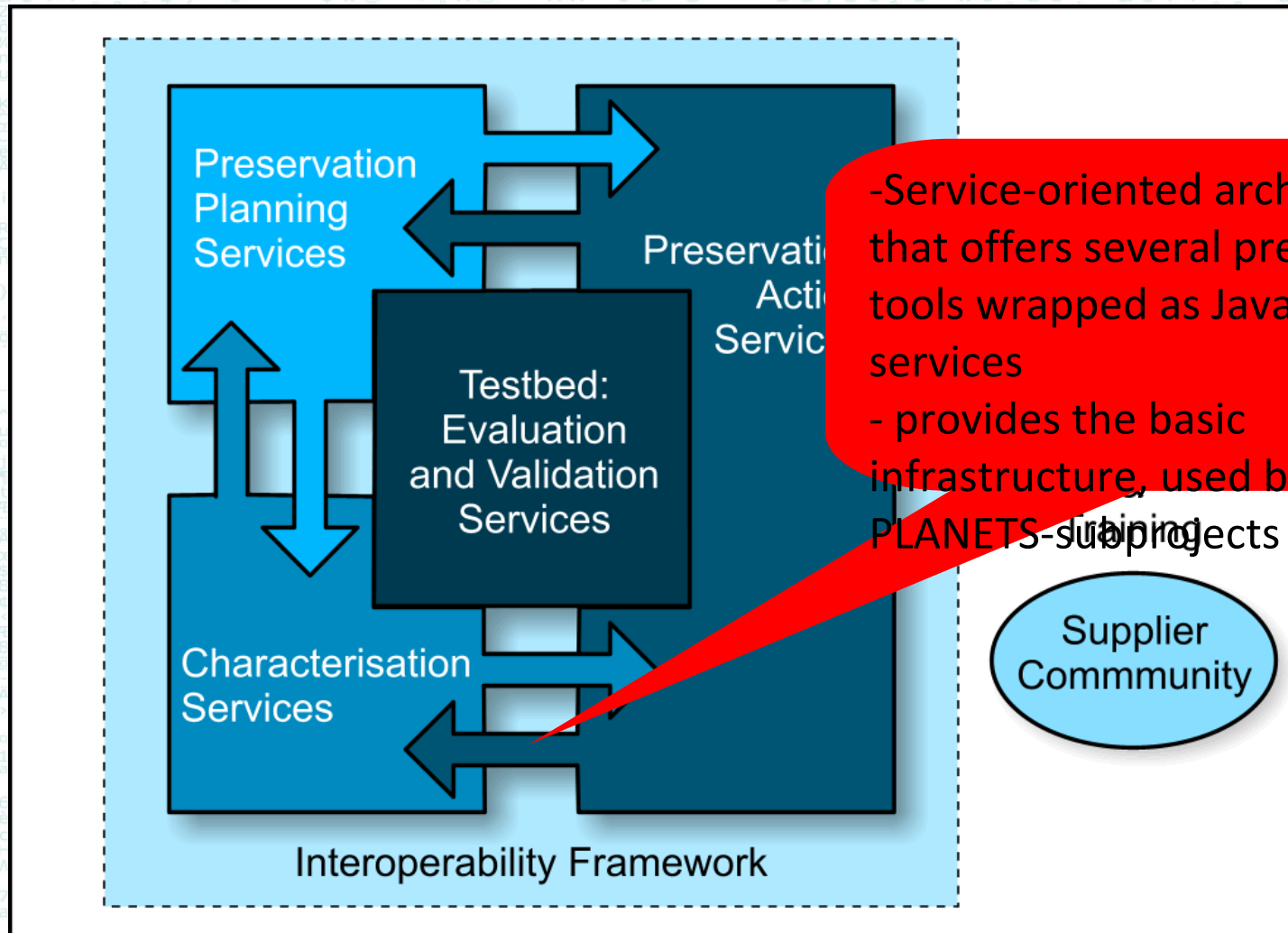
- Tessella Plc
- IBM Netherlands
- Microsoft Research, Cambridge
- Austrian Institute of Technology

- Hatii at University of Glasgow
- University of Freiburg
- Technical University of Vienna
- University at Cologne

The PLANETS project-structure



The PLANETS project-structure

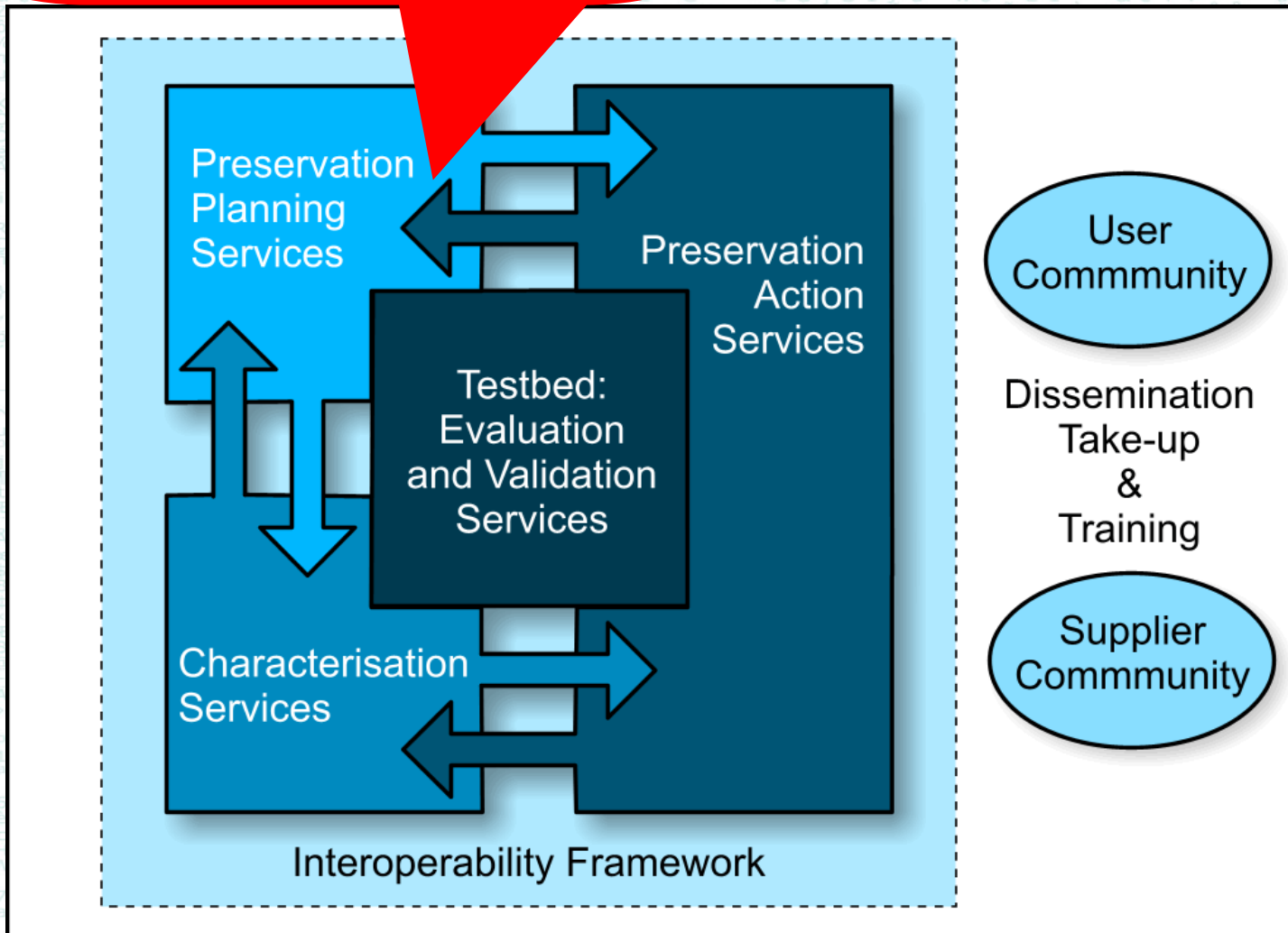


- Service-oriented architecture that offers several preservation-tools wrapped as Java-based services
- provides the basic infrastructure, used by all PLANETS-subprojects

Supplier Community

Creating high quality preservation plans to meet the requirements of memory institutions and the kind of their digital resources

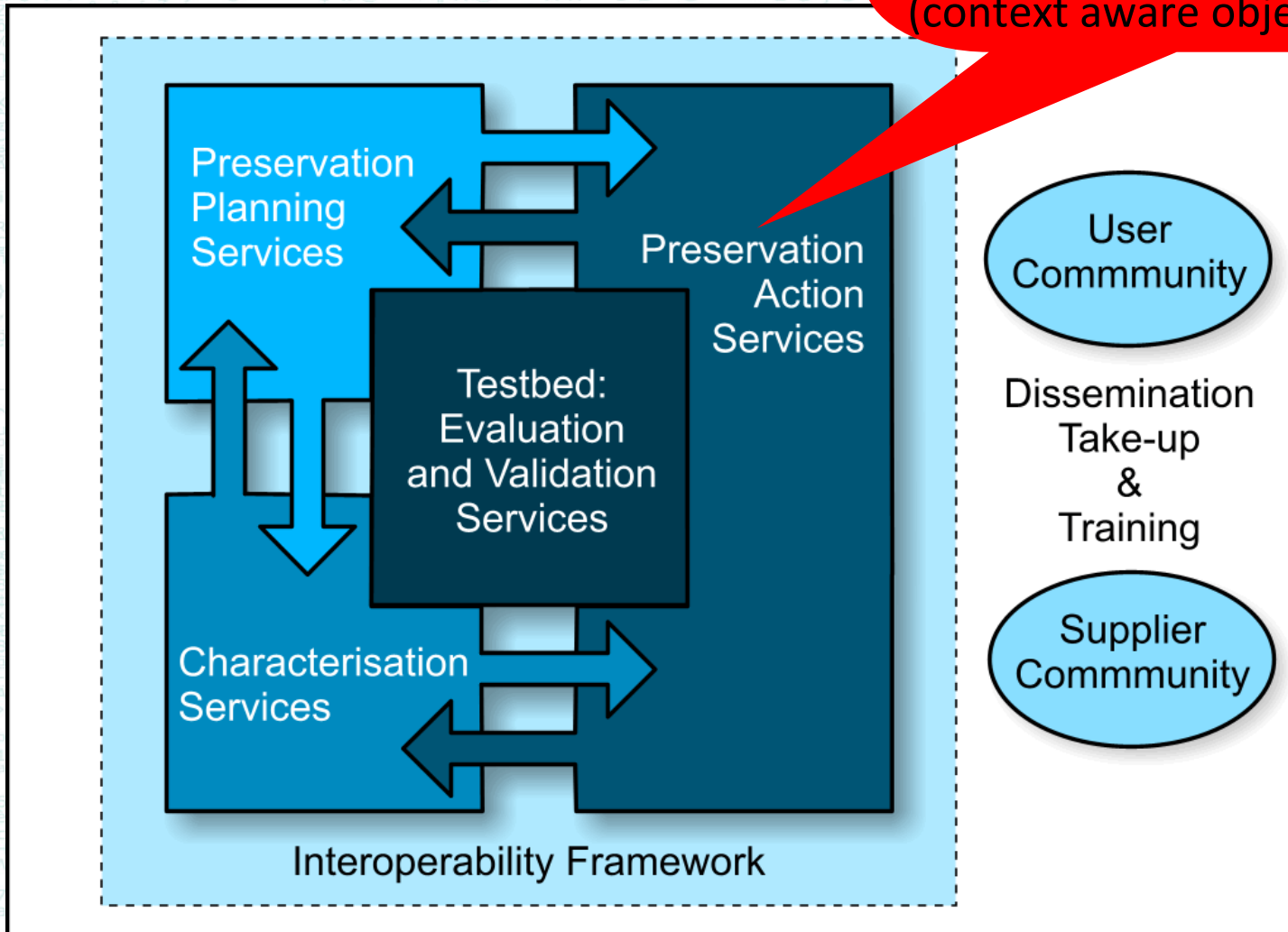
Project-structure



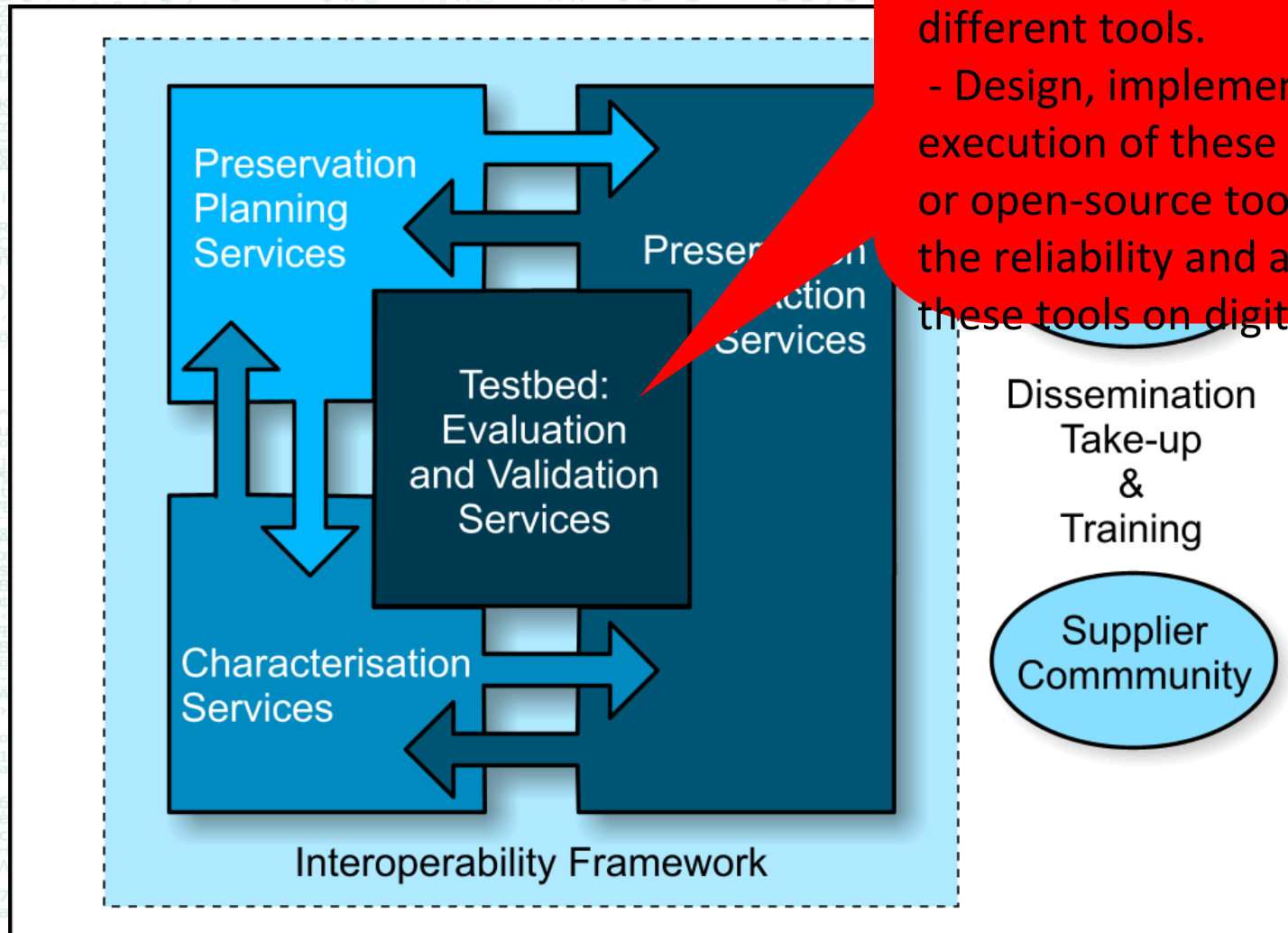
The PLANETS project

Tools & Strategies:

- migration und emulation
- Universal Virtual Computer
- Investigation of existing Solutions and additional development of new Software (context aware objects)

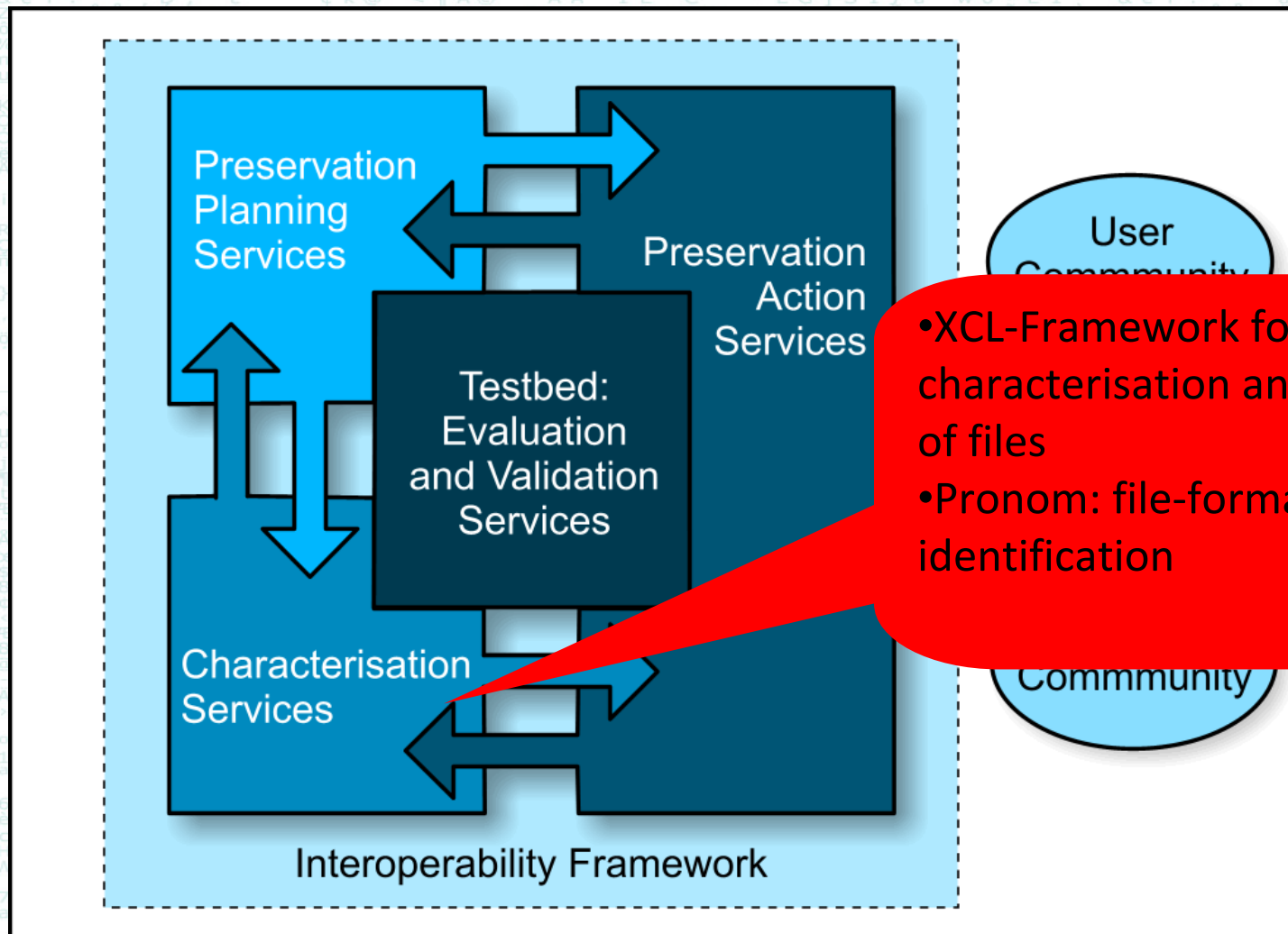


The PLANETS project



- Inspection of different long-term preservation strategies through experiments with different tools.
- Design, implementation und execution of these self-developed or open-source tools to measure the reliability and accuracy of these tools on digital resources.

The PLANETS project-structure



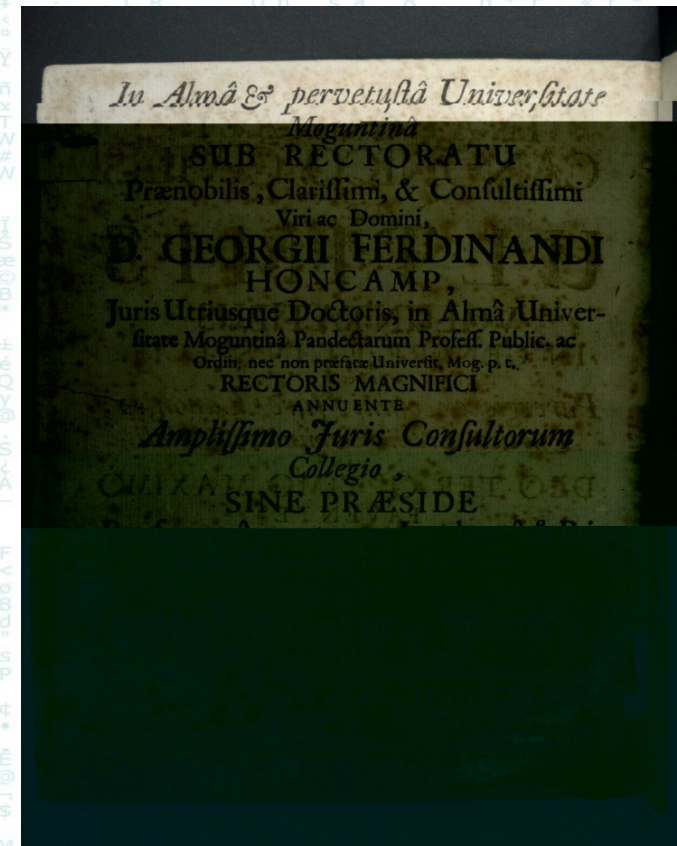
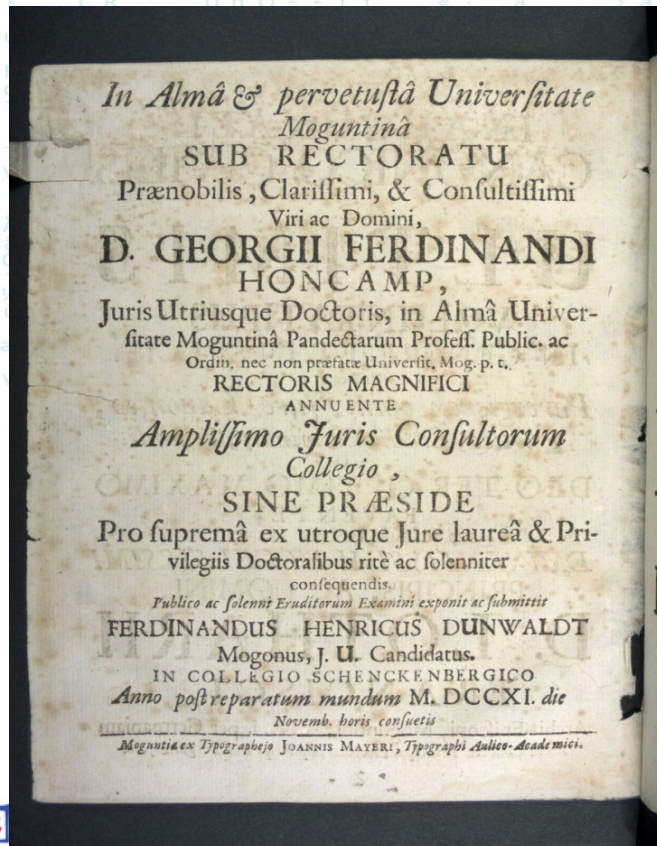
- XCL-Framework for characterisation and comparison of files
- Pronom: file-format identification

Introduction to the XCL-Tools

- **Problem 1:** Memory Institutions (Libraries, Archives, Museums) store a lot of digital material (e.g. images, text-files, emails, databases, audio-video-material).
- How do they find out, that their data has been damaged during a conversion?

Introduction to the XCL-Tools

- **Problem 1 - bit-deterioration during conversion:**
before: after:



Introduction to the XCL-Tools

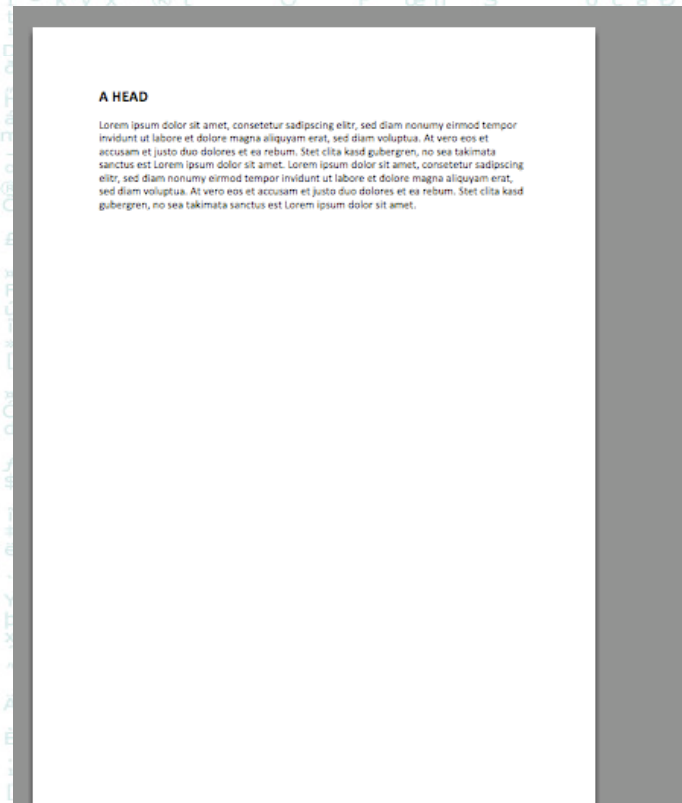
- **Problem 2:**

What if a file-format becomes obsolete? (e.g. *.doc)

➔ **Example-Strategy: file-format migration**

Introduction to the XCL-Tools

- Problem 2: What happened here?
before: after:



Introduction to the XCL-Tools

- Two XML-based languages:
- XCEL (Extensible Characterisation Extraction Language)

Introduction to the XCL-Tools

- Two XML-based languages:
- XCEL (Extensible Characterisation Extraction Language)
- XCDL (Extensible Characterisation Definition Language)

Introduction to the XCL-Tools

- Two Tools:
- **Extractor** uses XCEL-files to extract information from files and produces XCDL-files for each extracted file

Introduction to the XCL-Tools

- Two Tools:
 - **Extractor** uses XCEL-files to extract information from files and produces XCDL-files for each extracted file
 - **Comparator** compares two or more XCDL-files with different statistical methods to express the **equality** of these XCDL files
- ➔ both available as command-line-tools and wrapped as services in the IF-subproject

Introduction to the XCL-Tools

- The XCL-tools know „normdata“

```
<?xml version='1.0' encoding='UTF-8'?>
<xcdl xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns="http://www.planets-
project.eu/xcl/schemas/xcl" xsi:schemaLocation="http://www.planets-
project.eu/xcl/schemas/xcl ../xcl/xcdl/XCDLCore.xsd" id="0" >
  <object id="o1" >
    <normData type="image" id="ndl" >ff ff ff ff 00 00 f7 bd ff ff 00 00 f0 7b ff ff 00 00 e8 39 ff ff 00 00
dd f7 ff ff 00 00 d5 b5 ff ff 00 00 ce 73 ff ff 00 00 c6 31 ff ff 00 00
bd ef ff ff 00 00 b5 ad ff ff 00 00 ae 6b ff ff 00 00 a6 29 ff ff 00 00
9b e7 ff ff 00 00 93 a5 ff ff 00 00 8c 63 ff ff 00 00 84 21 ff ff 00 00
7b de ff ff 00 00 73 9c ff ff 00 00 6c 5a ff ff 00 00 64 18 ff ff 00 00
59 d6 ff ff 00 00 51 94 ff ff 00 00 4a 52 ff ff 00 00 42 10 ff ff 00 00
39 ce ff ff 00 00 31 8c ff ff 00 00 2a 4a ff ff 00 00 22 08 ff ff 00 00
17 c6 ff ff 00 00 0f 84 ff ff 00 00 08 42 ff ff 00 00 00 00 ff ff 00 00
ff ff f7 bd 00 00 f7 bd f7 bd 00 00 f0 7b f7 bd 00 00 e8 39 f7 bd 00 00
dd f7 f7 bd 00 00 d5 b5 f7 bd 00 00 ce 73 f7 bd 00 00 c6 31 f7 bd 00 00
bd ef f7 bd 00 00 b5 ad f7 bd 00 00 ae 6b f7 bd 00 00 a6 29 f7 bd 00 00
9b e7 f7 bd 00 00 93 a5 f7 bd 00 00 8c 63 f7 bd 00 00 84 21 f7 bd 00 00
7b de f7 bd 00 00 73 9c f7 bd 00 00 6c 5a f7 bd 00 00 64 18 f7 bd 00 00
59 d6 f7 bd 00 00 51 94 f7 bd 00 00 4a 52 f7 bd 00 00 42 10 f7 bd 00 00
39 ce f7 bd 00 00 31 8c f7 bd 00 00 2a 4a f7 bd 00 00 22 08 f7 bd 00 00
17 c6 f7 bd 00 00 0f 84 f7 bd 00 00 08 42 f7 bd 00 00 00 00 f7 bd 08 42
ff ff f0 7b 00 00 f7 bd f0 7b 00 00 f0 7b f0 7b 00 00 e8 39 f0 7b 00 00
dd f7 f0 7b 00 00 d5 b5 f0 7b 00 00 ce 73 f0 7b 00 00 c6 31 f0 7b 00 00
bd ef f0 7b 00 00 b5 ad f0 7b 00 00 ae 6b f0 7b 00 00 a6 29 f0 7b 00 00
9b e7 f0 7b 00 00 93 a5 f0 7b 00 00 8c 63 f0 7b 00 00 84 21 f0 7b 00 00
7b de f0 7b 00 00 73 9c f0 7b 00 00 6c 5a f0 7b 00 00 64 18 f0 7b 00 00
59 d6 f0 7b 00 00 51 94 f0 7b 00 00 4a 52 f0 7b 00 00 42 10 f0 7b 00 00
39 ce f0 7b 00 00 31 8c f0 7b 00 00 2a 4a f0 7b 00 00 22 08 f0 7b 00 00
17 c6 f0 7b 00 00 0f 84 f0 7b 00 00 08 42 f0 7b 08 42 00 00 f0 7b 0f 84
ff ff e8 39 00 00 f7 bd e8 39 00 00 f0 7b e8 39 00 00 e8 39 e8 39 00 00
dd f7 e8 39 00 00 d5 b5 e8 39 00 00 ce 73 e8 39 00 00 c6 31 e8 39 00 00
bd ef e8 39 00 00 b5 ad e8 39 00 00 ae 6b e8 39 00 00 a6 29 e8 39 00 00
9b e7 e8 39 00 00 93 a5 e8 39 00 00 8c 63 e8 39 00 00 84 21 e8 39 00 00
7b de e8 39 00 00 73 9c e8 39 00 00 6c 5a e8 39 00 00 64 18 e8 39 00 00
59 d6 e8 39 00 00 51 94 e8 39 00 00 4a 52 e8 39 00 00 42 10 e8 39 00 00
39 ce e8 39 00 00 31 8c e8 39 00 00 2a 4a e8 39 00 00 22 08 e8 39 00 00
17 c6 e8 39 00 00 0f 84 e8 39 08 42 08 42 e8 39 0f 84 00 00 e8 39 17 c6
ff ff dd f7 00 00 f7 bd dd f7 00 00 f0 7b dd f7 00 00 e8 39 dd f7 00 00
```

Properties in the XCL-Tools

XCLSuite

File		
basi0g01	✗	✗
basi0g02	✓	✗
basi0g04	✓	✗
basi0g08	✓	✗
basi0g16	✗	✗
basi2c08	✓	✗
basi2c16	✗	✗
basi3p01	✓	✗
basi3p02	✓	✗
basi3p04	✓	✗
basi3p08	✓	✗
basi4a08	✗	✗
basi6a08	✗	✗
basn0g01	✓	✗
basn0g02	✓	✗
basn0g04	✓	✗
basn0g08	✓	✗
basn0g16	✗	✗
basn2c08	✓	✗
basn2c16	✗	✗
basn3p01	✓	✗
basn3p02	✓	✗
basn3p04	✓	✗
basn3p08	✓	✗
basn4a08	✗	✗
basn6a08	✗	✗
bgai4a08	✗	✗
bgan6a08	✗	✗
bgbn4a08	✗	✗
bgwn6a08	✗	✗

Data View : basi2c16

Comparator View : basi2c16

Metric: equal
Result: true

Property: colourSpaceName

Metric: equal
Result: true

Property: imageHeight

Metric: equal
Result: true

Property: imageWidth

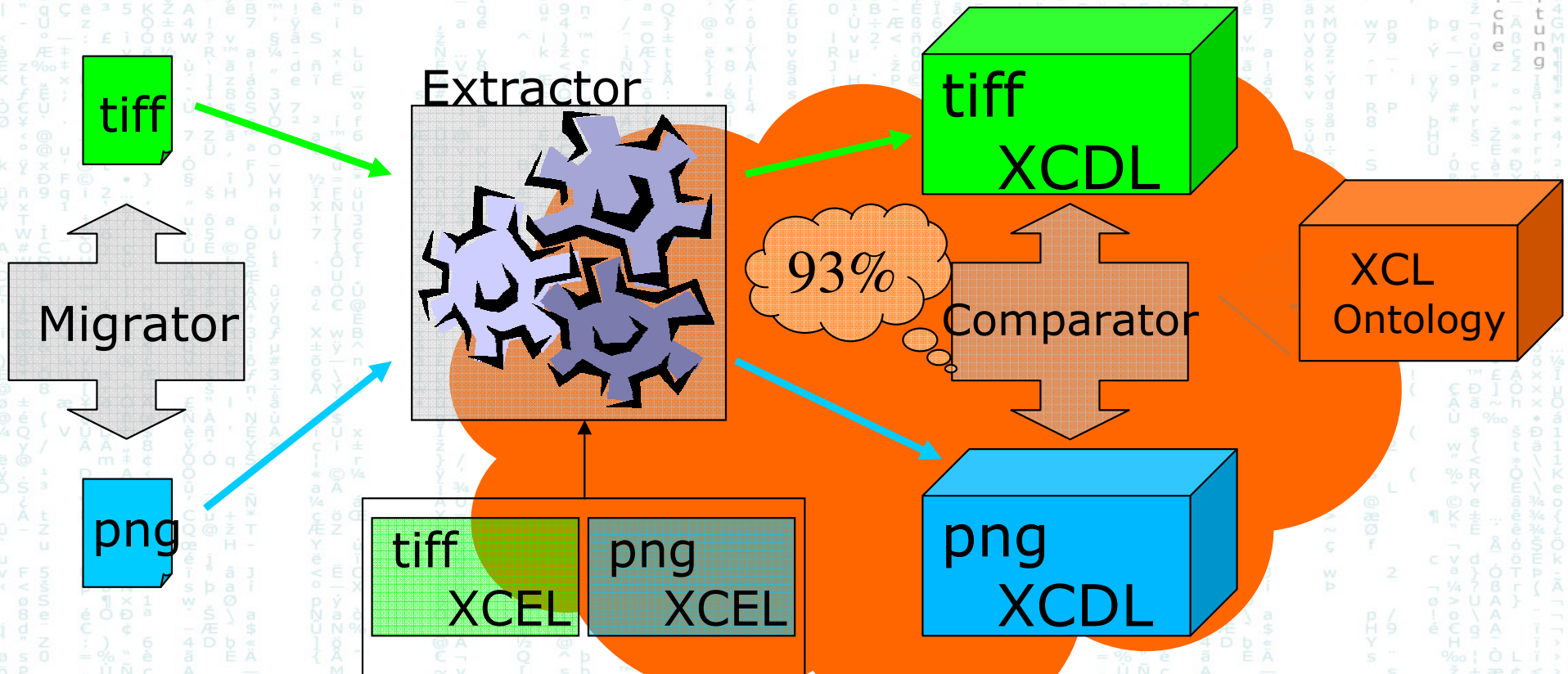
Metric: equal
Result: true

Property: normData

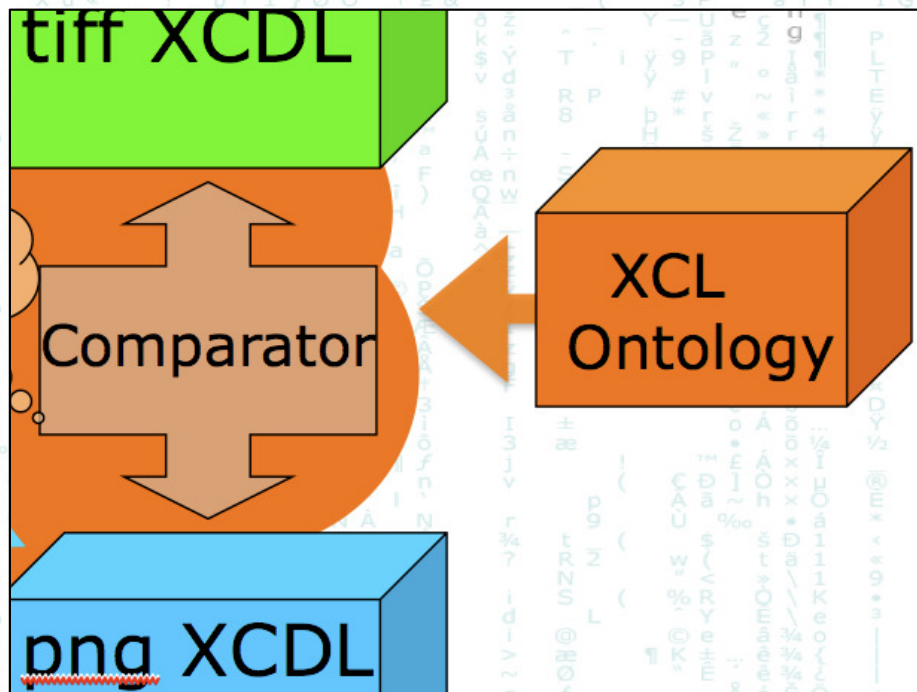
Metric: hammingDistance
Result: 1288

Data View : basi2c16

The XCL-Ontology

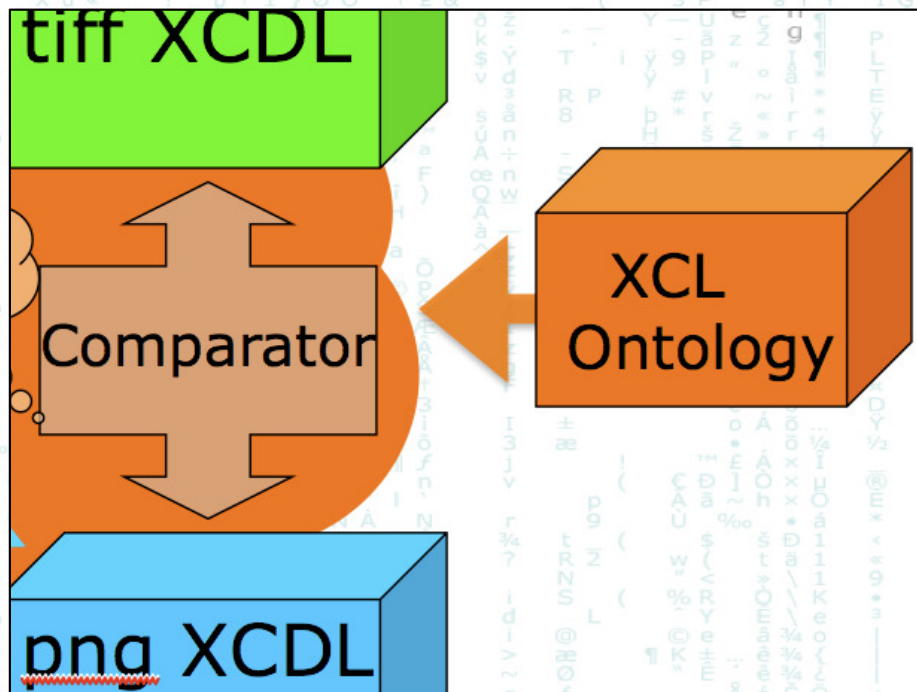


The XCL-Ontology



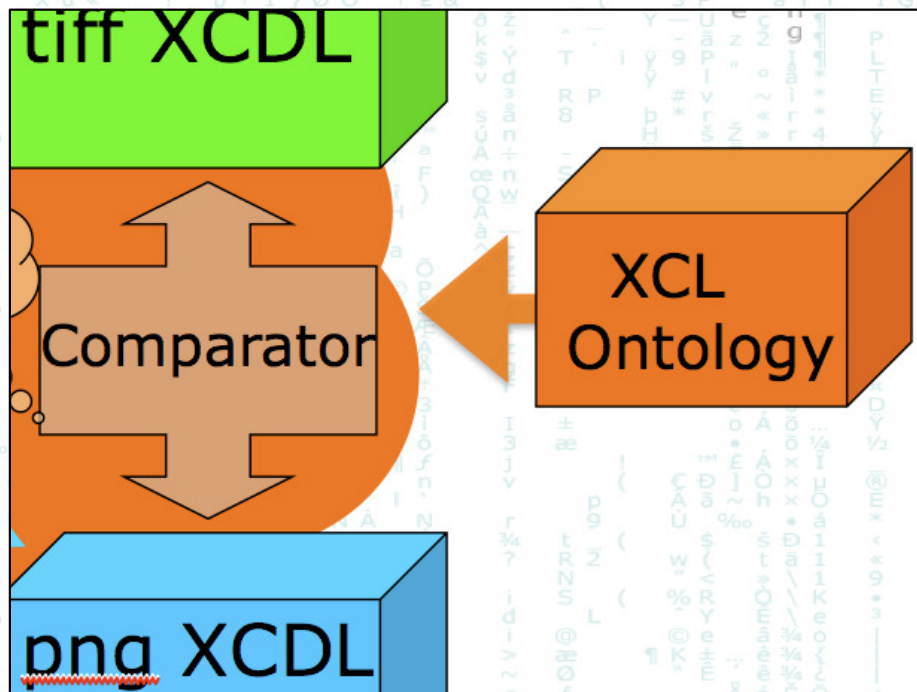
- Lists all available property-names for file-formats, the extractor can process

The XCL-Ontology



- Lists all available property-names for file-formats, the extractor can process
- Maps these terms to a normalised naming, which is then used in the XDL-languages

The XCL-Ontology



- Lists all available property-names for file-formats, the extractor can process
- Maps these terms to a normalised naming, which is then used in the XDL-languages
- Defines datatypes and units for these properties

Ontologies – what?

Definition:

“formal, shared conceptualization of a particular domain of interest”

(T. Gruber, “ Translation Approach to Portable Ontology Specifications”, Knowledge Acquisition 5, No. 2, 199-220 (1993).)

Ontologies – why?

Advantages:

- XML-based (rdf-owl)

Ontologies – why?

Advantages:

- XML-based (rdf-owl)
- machine-readable

Ontologies – why?

Advantages:

- XML-based (rdf-owl)
- machine-readable
- Clear relationships between different kinds of „things“

Ontologies – why?

Advantages:

- XML-based (rdf-owl)
- machine-readable
- Clear relationships between different kinds of „things“
- Enable inference

Ontologies – why?

Advantages:

- XML-based (rdf-owl)
- machine-readable
- Clear relationships between different kinds of „things“
- Enable inference
- are extensible

What concepts do we need?

Example

- A colour map is called 'PLTE' in PNG
- Is the same as 'lookup' in pdf

What concepts do we need?

Example

- A colour map is called 'PLTE' in PNG
- Is the same as 'lookup' in pdf
- 'PLTE' is modelled as an instance of PNG_Properties
- 'lookup' is modelled as an instance of PDF_Properties

What concepts do we need?

In OWL terminology:

- **Classes** (= abstract)
- **Individuals** (=concrete instances of these)
- **Object-Properties** (relationships between individuals of different classes)

And what about datatypes and units?

Abstract concepts as Classes:

- datatypes and units
- Instances as Individuals (cm, inch / int, char)

How does it look like?

The screenshot displays a software interface with several panels:

- Asserted class hierarchy / Inferred class hierarchy:** A tree view showing the class hierarchy for 'fmt_13_png'. The root is 'fmt_13_png', which includes sub-classes like 'fmt_142_CGM', 'fmt_17_pdf', 'fmt_385_mpeg1', 'fmt_386_mpeg2', 'fmt_392_jpeg2000', 'fmt_40_msword', 'fmt_42_jpeg_1.0', 'fmt_4_gif89a', 'fmt_91_svg', 'fmt_xxx_docx', 'fmt_xxx_imagemagick', 'fmt_xxx_JHOVE', 'fmt_xxx_mpeg4', 'fmt_xxx_rtf1.5', 'fmt_xxx_rtf1.9', 'fmt_xxx_tiffinfo', 'fmt_xxx_videoSpec', 'NISO_Property', and 'XCL_Properties'. 'XCL_Properties' further branches into 'audioInformation', 'generalInformation', 'imageInformation' (with sub-items 'rasterInformation', 'vectorInformation', and 'otherInformation'), and 'otherInformation'.
- Class Annotations: fmt_13_png:** A panel showing annotations for the selected class, including 'PronomID' with the value 'fmt/13'.
- Class Description: fmt_13_png:** A list of properties for the class, including 'gAMA', 'hIST', 'iCCP', 'IHDR', 'IHDR_bitDepth', 'IHDR_ColorType', 'IHDR_CompressionMethod', 'IHDR_CRC', 'IHDR_height', 'IHDR_interlace', 'IHDR_width', 'iTXT', 'pHYs_resolutionUnit', 'pHYs_resolutionX', 'pHYs_resolutionY', and 'PLTE'.
- Object property hierarchy / Data property hierarchy / Individuals:** A panel at the bottom for viewing object and data properties.

How does it look like?

The screenshot displays a software interface with two main panels. The top panel, titled "Individual Annotations: PLTE", contains three sections: "Annotations" with a plus sign, "Datatype" with a diamond icon and the value "int", and "Unit" with a diamond icon and the value "NULL". The bottom panel, titled "Description: PLTE", contains two sections: "Types" with a plus sign and a yellow circle icon next to the text "fmt_13_png", and "Same individuals" with a plus sign. To the right, a partially visible panel titled "Property assertions: PLTE" shows "Object property assertions" with a plus sign and a blue bar icon next to the text "convertTo colourPalette". At the top of the interface, there are two tabs: "Individual Annotations" (selected) and "Individual Usage".

How does it look like?

The screenshot displays a web-based ontology editor interface. On the left, a 'Class Hierarchy' tree shows a list of classes under 'XCL_Properties', with 'rasterInformation' selected. The main area is divided into several panels:

- Individual Annotations: colourPalette:** Shows annotations for the class 'colourPalette'.
 - comment:** "A colour map for palette colour images. Usually for RGB coloured pictures. Defined as array of hex-colour-values, appearing in the order left to right and top to bottom"
 - Datatype:** int
 - ID:** "id25"
 - Unit:** NULL
- Description: colourPalette:** Shows the type 'rasterInformation'.
- Property assertions: colourPalette:** Lists object property assertions: 'has_alternative_filespecific_name colourMapTiff', 'has_alternative_filespecific_name lookup', and 'has_alternative_filespecific_name PLTE'.

How does it look like?

The screenshot displays a software interface with several panels:

- Asserted class hierarchy:** A tree view showing a hierarchy starting with 'Thing', followed by 'XCL_Properties', and then 'rasterInformation' (highlighted in blue).
- Class Annotations: rasterInformation:** A panel titled 'Annotations' with a plus sign, currently empty.
- Class Description: rasterInformation:** A list of properties for 'rasterInformation', including CCITT3, CCITT4, cellHeight, cellLengthBilevel, cellWidth, cellWidthBilevel, clipPath, colourPalette, exifPhotoMetadata, extraSamples, fillOrder, gammaValueGray, gammaValueRGB, grayResponseCurve, grayResponseUnit, and histogram.
- Object property hierarchy:** A panel with a right-pointing arrow.
- Object Properties:** A panel with icons for adding, removing, and refreshing properties.

How does it look like?

[!— http://planetarium.hki.uni-koeln.de/public/XCL/ontology/XCLOntology.owl#colourPalette](http://planetarium.hki.uni-koeln.de/public/XCL/ontology/XCLOntology.owl#colourPalette) →

```
<rasterInformation rdf:about="#colourPalette">
```

```
<rdfs:comment rdf:datatype="&xsd:string"
```

```
>A colour map for palette colour images.
```

Usually for RGB coloured pictures.

Defined as array of hex-colour-values, appearing in the order left to right and top to bottom</rdfs:comment>

```
<ID rdf:datatype="&xsd:string">id25</ID>
```

```
<Unit rdf:resource="#NULL"/>
```

```
<has_alternative_filespecific_name rdf:resource="#PLTE"/>
```

```
<has_alternative_filespecific_name rdf:resource="#colourMapTiff"/>
```

```
<Datatype rdf:resource="#int"/>
```

```
<has_alternative_filespecific_name rdf:resource="#lookup"/>
```

```
</rasterInformation>
```

How does it look like?

```
<!-- http://planetarium.hki.uni-koeln.de/public/XCL/ontology/XCLOntology.owl#PLTE -->
```

```
<fnt_13_png rdf:about="#PLTE">
  <rdfs:comment xml:lang="en"
```

>Chunk that contains the Palette for truecolour-PNGs. Use is optional.

If the Image contains also an alpha-channel there should also be a bKGD Chunk.

Required for colour_type3-PNGs</rdfs:comment>

```
<Unit rdf:resource="#NULL"/>
```

```
<convertTo rdf:resource="#colourPalette"/>
```

```
<Datatype rdf:resource="#int"/>
```

```
</fnt_13_png>
```

How is it used?

- As a SPARQL-based webservice to produce XML-Schema-files for each format – all adhering to the same (XCL-) naming convention

Generate Nameslib | View Testfiles | Add Testfile

Generate nameslib from data/xcl.rdf

Base URI: <http://planetarium.hki.uni-koeln.de/public/XCL/ontology/XCLOntology.owl#>

Select format: **fmt_91_svg** | Generate Nameslib | Generate Comparator-XML

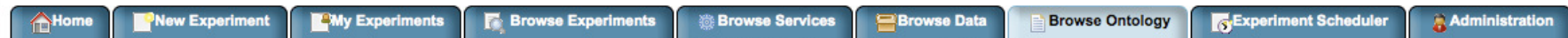
5 properties for

- NISO_Property
- fmt_100_html4.1
- fmt_10_tiff
- fmt_115_bmp
- fmt_136_odt
- fmt_137_ods
- fmt_138_otp
- fmt_13_png
- fmt_142_CGM
- fmt_17_pdf
- fmt_385_mpeg1
- fmt_386_mpeg2
- fmt_392_jpeg2000
- fmt_40_msword
- fmt_42_jpeg_1.0
- fmt_4_gif89a

Download Name (xsd)

How is it used?

- In the current version of the PLANETS-Testbed



PLANETS Testbed - Browse Ontology

PLANETS ONTOLOGY BROWSER

Explore all **digital object properties** the Testbed can measure for you.

Multiple **tree views** of the ontology are available.

Filter properties using the **full-text search**.

Drag and drop properties you're interested in to the **clipboard**.

INFORMATION

For further information regarding properties or the process of adding new properties to the Testbed please contact The Planets Properties Working Group (DOPWG)

all available Planets properties:

apply settings: full-text filtering and different tree views

filter properties:

tree-context menu (slow!): select tree view:

- specificationPropertyNames
 - NISO_Property (158)
 - fmt_100_html4.1 (123)
 - fmt_10_tiff (83)
 - fmt_115_bmp (3)
 - fmt_136_odt
 - fmt_137_ods
 - fmt_138_odp
 - fmt_13_png (47)
 - fmt_17_pdf (241)
 - fmt_385_mpeg1
 - fmt_386_mpeg2
 - fmt_392_jpeg2000 (11)
 - fmt_40_msword
 - fmt_42_jpeg_1.0 (12)
 - fmt_4_gif89a (21)
 - fmt_91_svg (17)

selected property information panel

selected node:
human readable name:
uri:

clipboard

Next Steps „PLANETS-wide“ Ontology

- Include more properties (*non-extractable* or *observational*) from all PLANETS-Subprojects and external projects
 1. Preservation Planning / PLATO
 2. PLANETS Testbed (benchmark-goals)
 3. INSPECT
- Unify and map these to each other...

Next Steps „PLANETS-wide“ Ontology

- Include more properties
(non-extractable or observational)

The screenshot displays the Protege ontology editor interface. The main window is titled "Active Ontology" and shows the "Class Annotations" tab for the class "PreservationPlanningProperties".

Class Annotations: PreservationPlanningProperties

Annotations: +

Class Description: PreservationPlanningProperties

- ◆ 'ObservationalPersonOrGroupProperty(Inventory)'
- ◆ 'ObservationalPersonOrGroupProperty(Registry)'
- ◆ 'ObservationalProductProperty(Inventory)'
- ◆ 'ObservationalProductProperty(Registry)'
- ◆ 'ObservationalSoftwareProperty(Registry)'
- ◆ 'ObservationalStandardProperty(Registry)'
- ◆ 'ObservationalStorageMediumProperty(Registry)'
- ◆ ObservationalAbstract
- ◆ ObservationalAcquisitionBudget
- ◆ ObservationalAdministrativeSupport
- ◆ ObservationalAlternativeText
- ◆ ObservationalAnimation
- ◆ ObservationalAppearanceProperty
- ◆ ObservationalAppendixGroup
- ▲ ObservationalAppendixStructure

Asserted Class Hierarchy: PreservationPlanningProperties

- Thing
 - ObservationProperties
 - PlatoProperties
 - PreservationPlanningProperties
 - testbedBenchmarkGoalsProperties
 - observationalAudioProperties
 - observationalImageProperties
 - observationalTextProperties
 - observationalVideoProperties
 - XCL_Staff

Object Properties:

Next Steps „PLANETS-wide“ Ontology

- Display the „observational“ Properties as list in the testbed-webservice
- Let the user add values
- ...work in progress (until May 2010)

Summary

- Characterisation tool for different file-formats = Extractor (XCDL-output)
- Evaluation tool = Comparator (compare XCDLs)
- The XCL Ontology (soon extended)
- Embedded into the PLANETS Interoperability Framework and the PLANETS Testbed

Thank you

Find out more:

<http://planetarium.hki.uni-koeln.de/>

<http://planets-project.eu/>