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

- The British Library
- National Library, Netherlands
- Austrian National Library
- State and University Library, Denmark
- Royal Library, Denmark

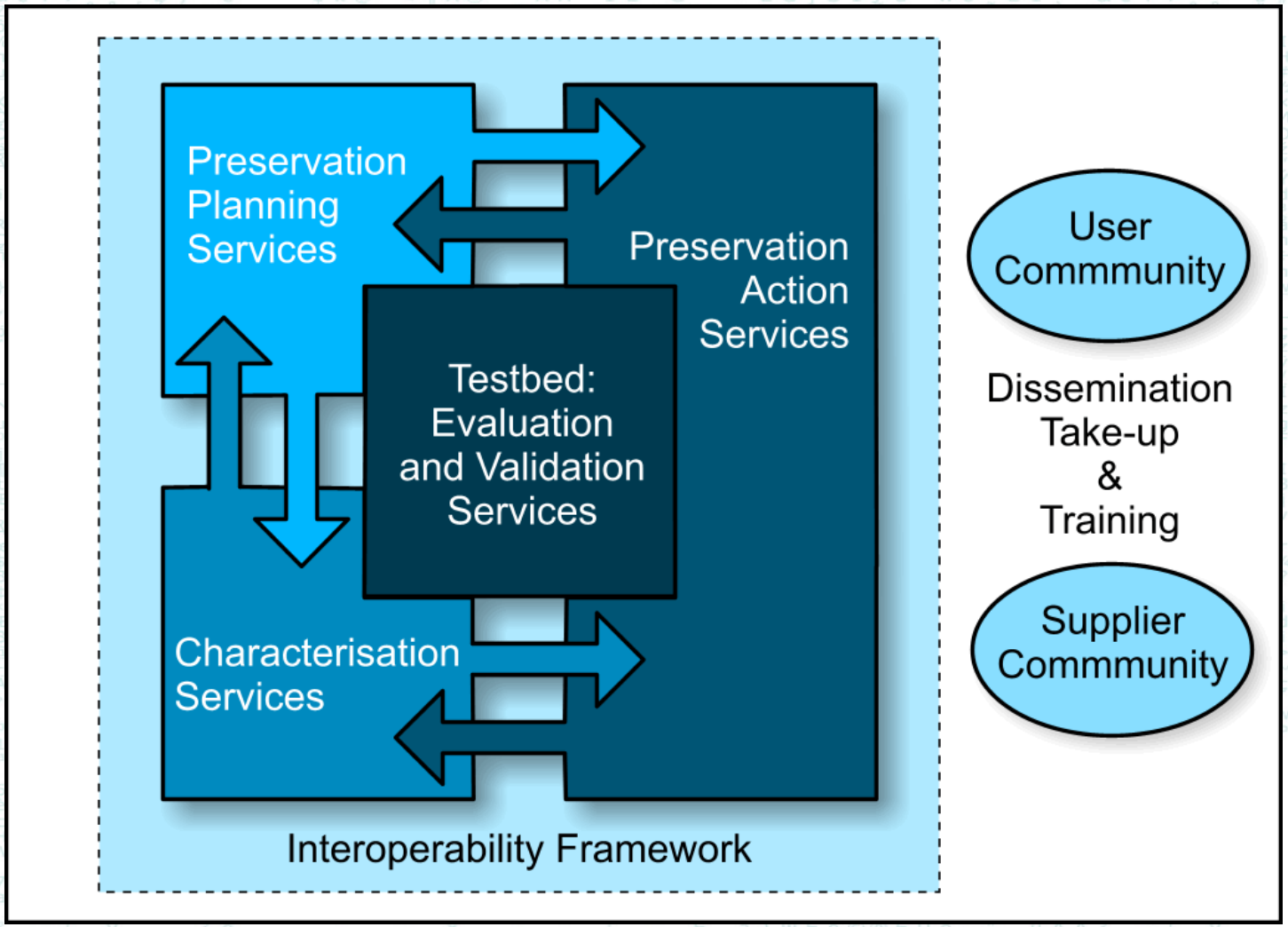
- National Archives, UK
- Swiss Federal Archives
- National Archives, Netherlands

PLANETS – PARTNERS II

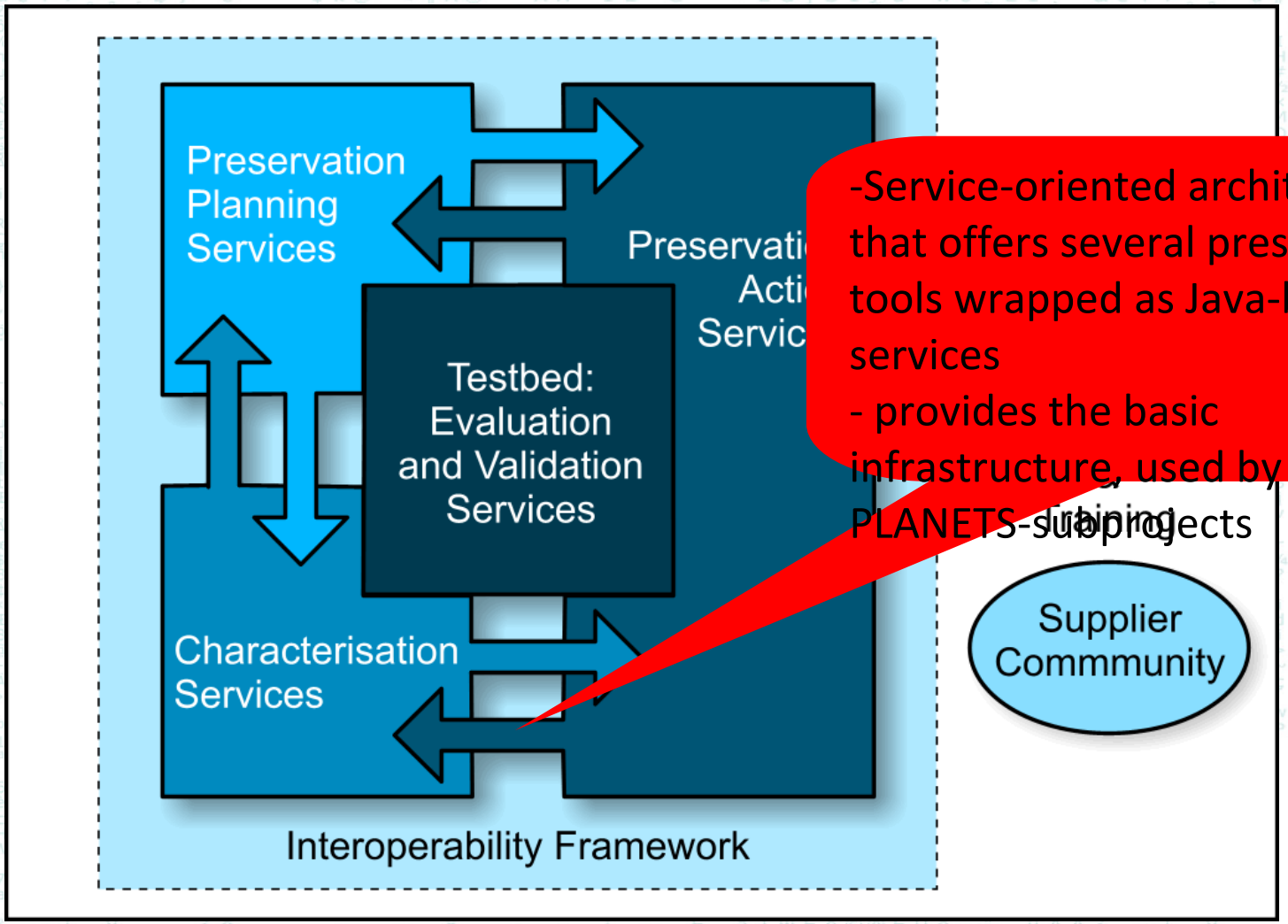
- Tessella Plc
- IBM Netherlands
- Microsoft Research, Cambridge
- Austrian Institute of Technology

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

The PLANETS project-structure

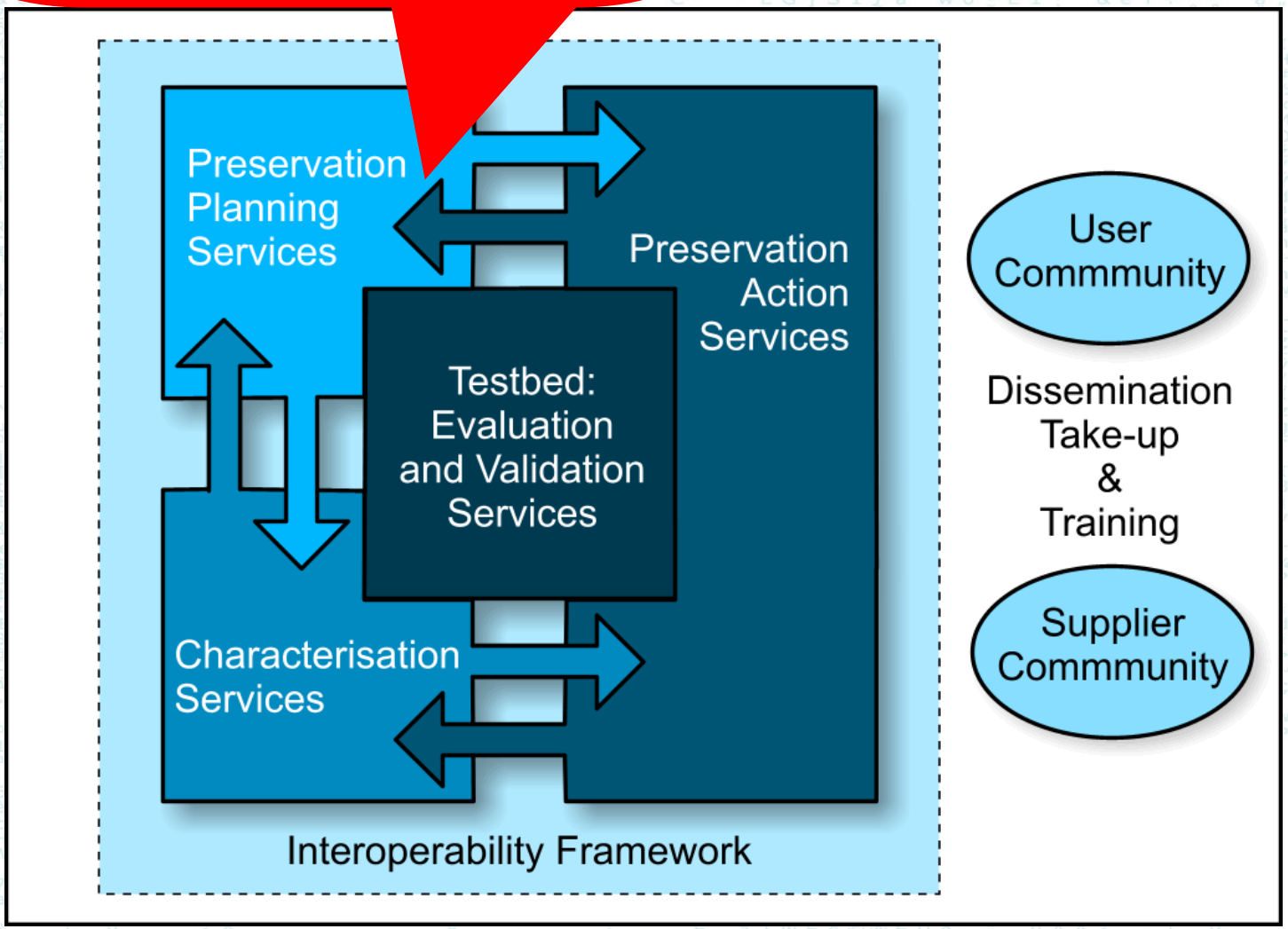


The PLANETS project-structure



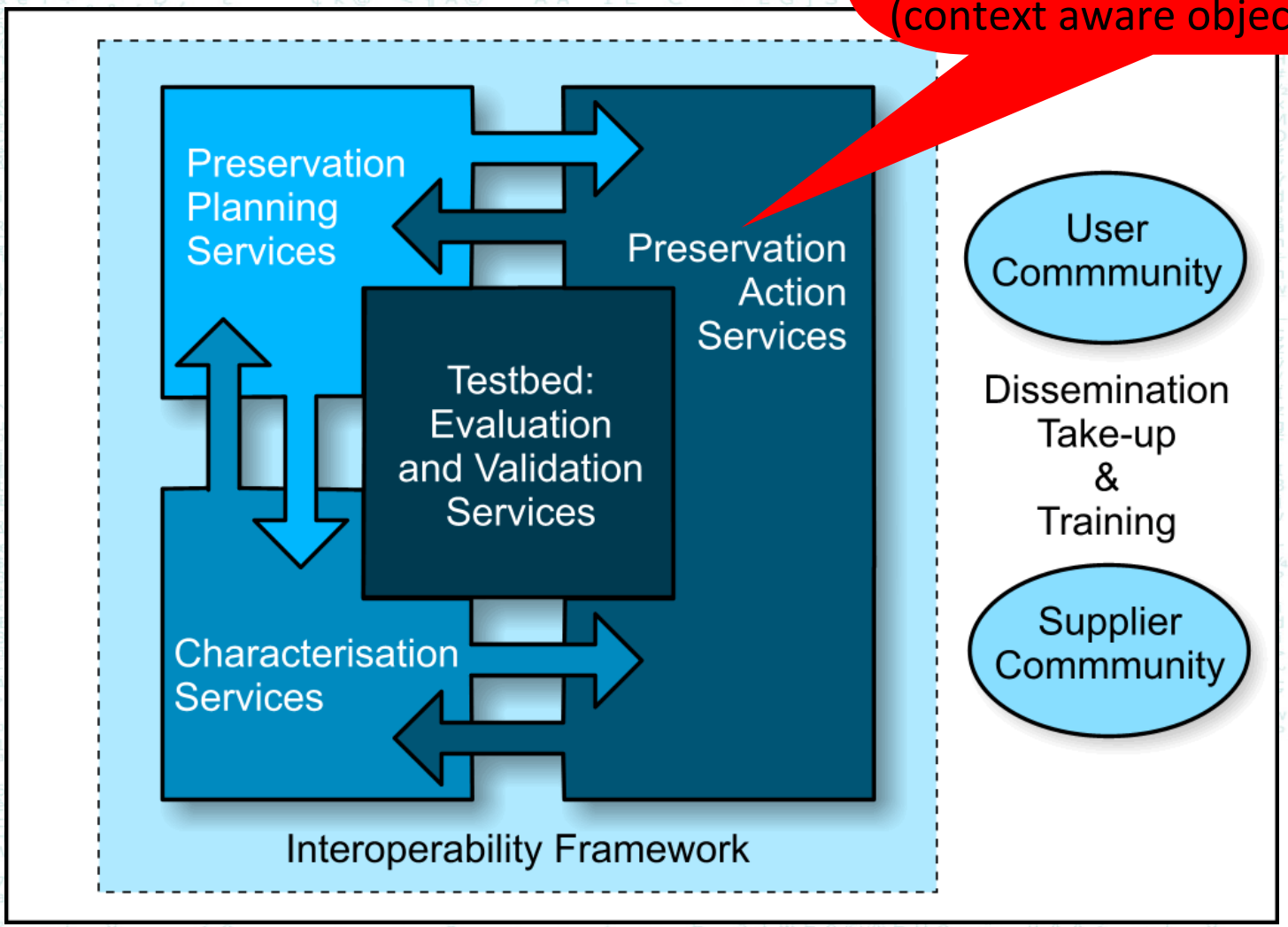
Creating high quality preservation plans to meet the requirements of memory institutions and the kind of their digital resources

roject-structure



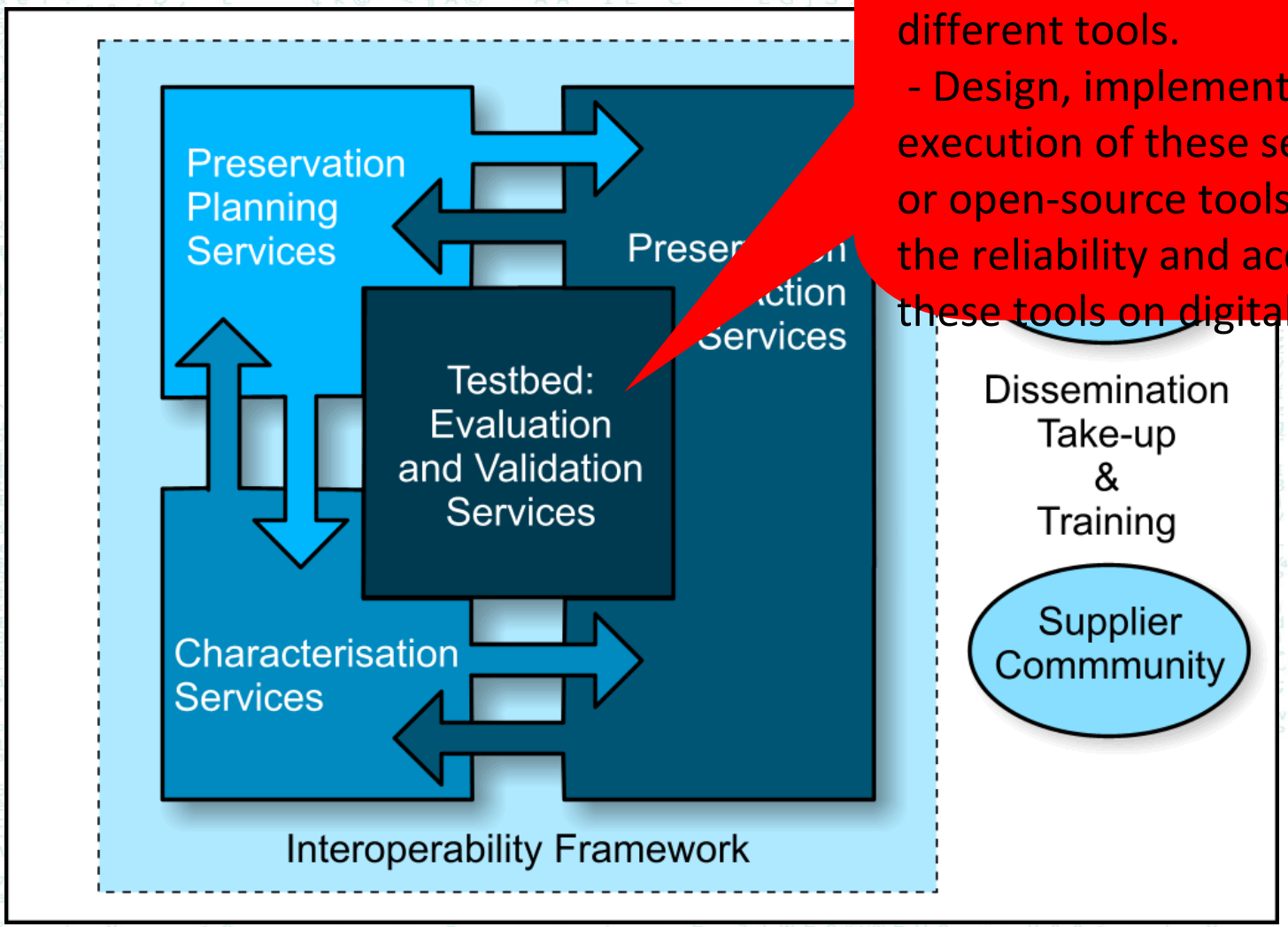
The PLANETS proje

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

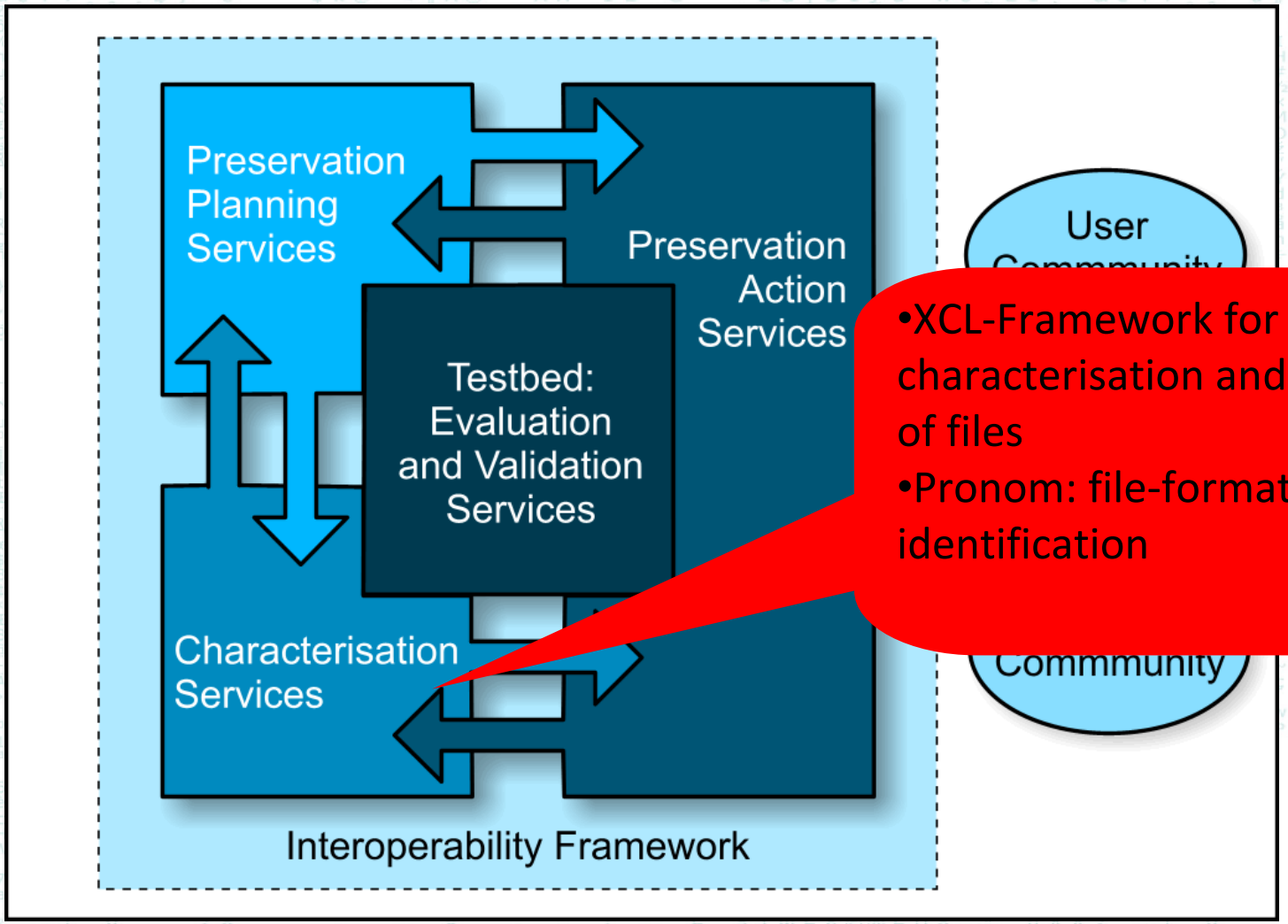


The PLANETS proje

- 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 messure the reliability and accuracy of these tools on digital resources.



The PLANETS project-structure



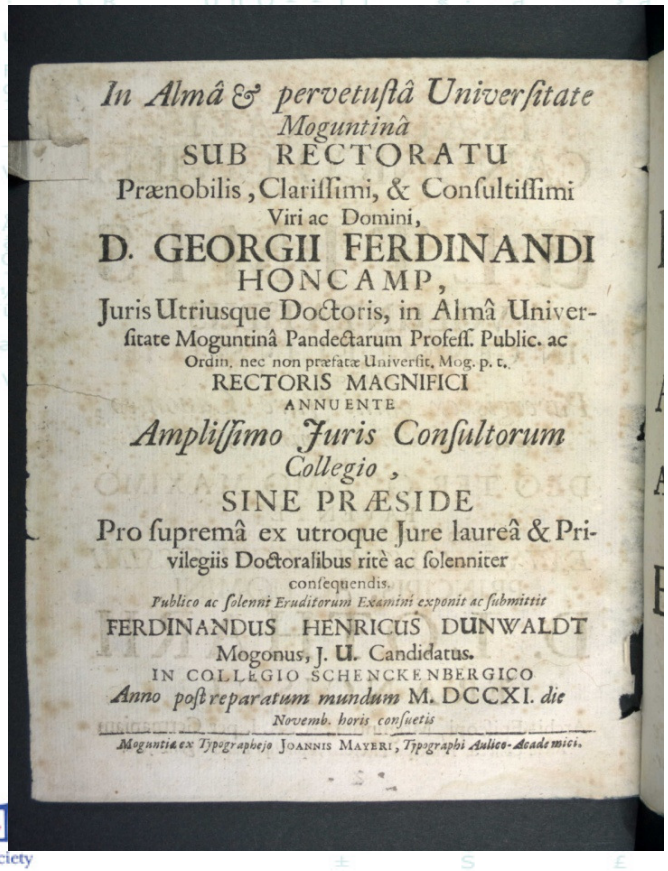
Introduction to the XCL-Tools

- **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

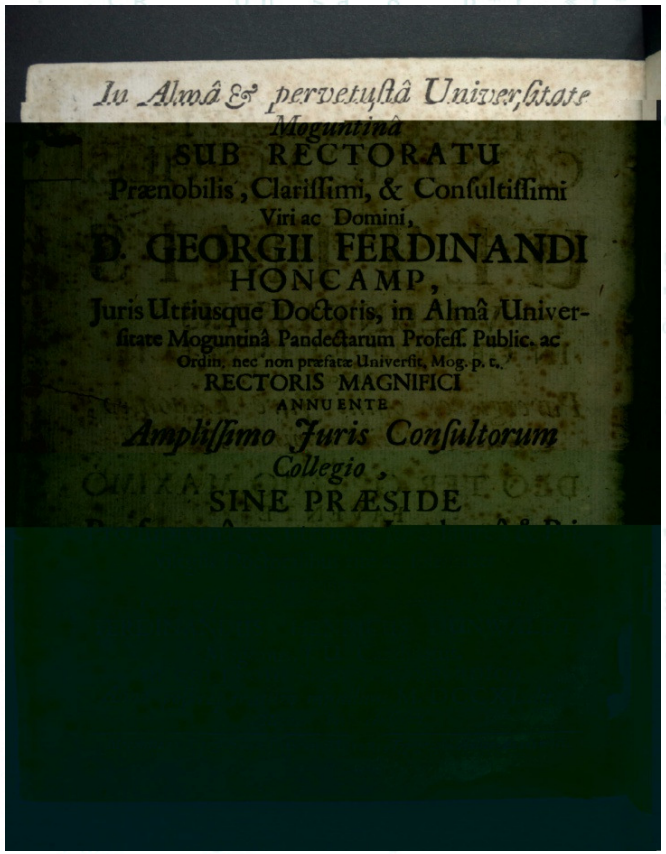
- **Problem 1 - bit-detorriation during conversion:**

before:



In Almâ & pervetuſtâ Univerſitate
Moguntinâ
SUB RECTORATU
Prænobilis, Clariſſimi, & Conſultiſſimi
Viri ac Domini,
D. GEORGII FERDINANDI
HONCAMP,
Juris Utriusque Doctoris, in Almâ Univer-
ſitate Moguntinâ Pandectarum Profeſſ. Public. ac
Ordin. nec non præfatæ Univerſit. Mog. p. t.
RECTORIS MAGNIFICI
ANNUENTE
Ampliſſimo Juris Conſultorum
Collegio,
SINE PRÆSIDE
Pro ſuprema ex utroque Jure laureâ & Pri-
vilegiis Doctoralibus ritè ac ſolenniter
conſequendis,
Publico ac ſolenni Eruditorum Examini exponit ac ſubmittit
FERDINANDUS HENRICUS DUNWALDT
Mogonus, J. U. Candidatus.
IN COLLEGIO SCHENCKENBERGICO
Anno poſt reparatum mundum M. DCCXI. die
Novemb. horis conſuetis
Moguntiæ ex Typographejo JOANNIS MAYERI, Typographi Aulico-Academici.

after:



In Almâ & pervetuſtâ Univerſitate
Moguntinâ
SUB RECTORATU
Prænobilis, Clariſſimi, & Conſultiſſimi
Viri ac Domini,
D. GEORGII FERDINANDI
HONCAMP,
Juris Utriusque Doctoris, in Almâ Univer-
ſitate Moguntinâ Pandectarum Profeſſ. Public. ac
Ordin. nec non præfatæ Univerſit. Mog. p. t.
RECTORIS MAGNIFICI
ANNUENTE
Ampliſſimo Juris Conſultorum
Collegio,
SINE PRÆSIDE

Introduction to the XCL-Tools

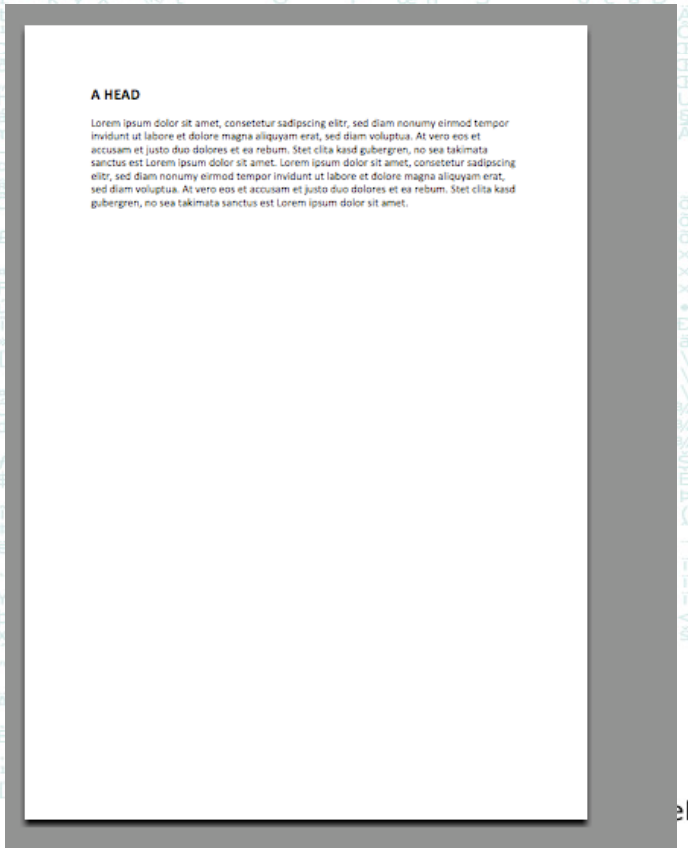
- **Problem 2:**
  What if a file-format becomes obsolete? (e.g. *.doc)
  ➔ **Example-Strategy:** file-format **migration**

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

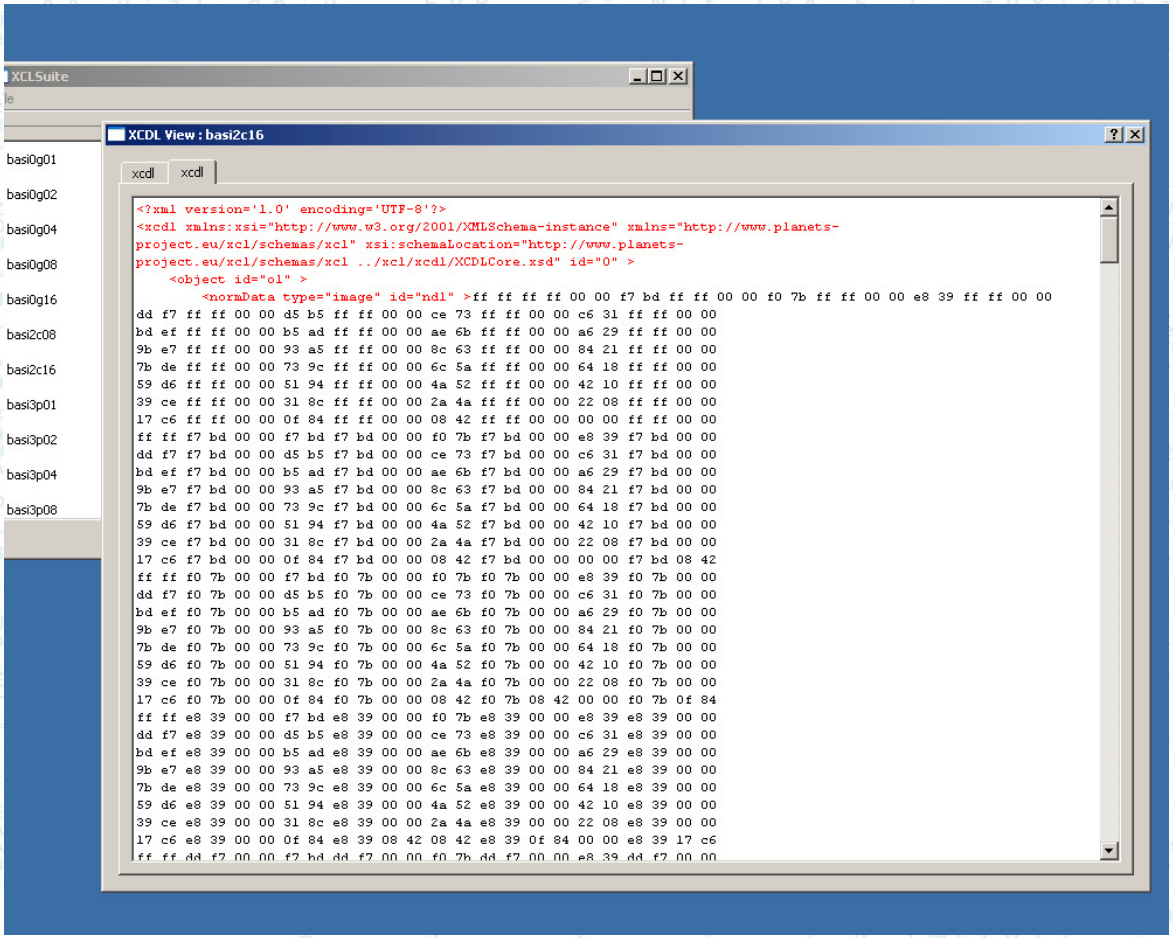
Introduction to the XCL-Tools

- 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

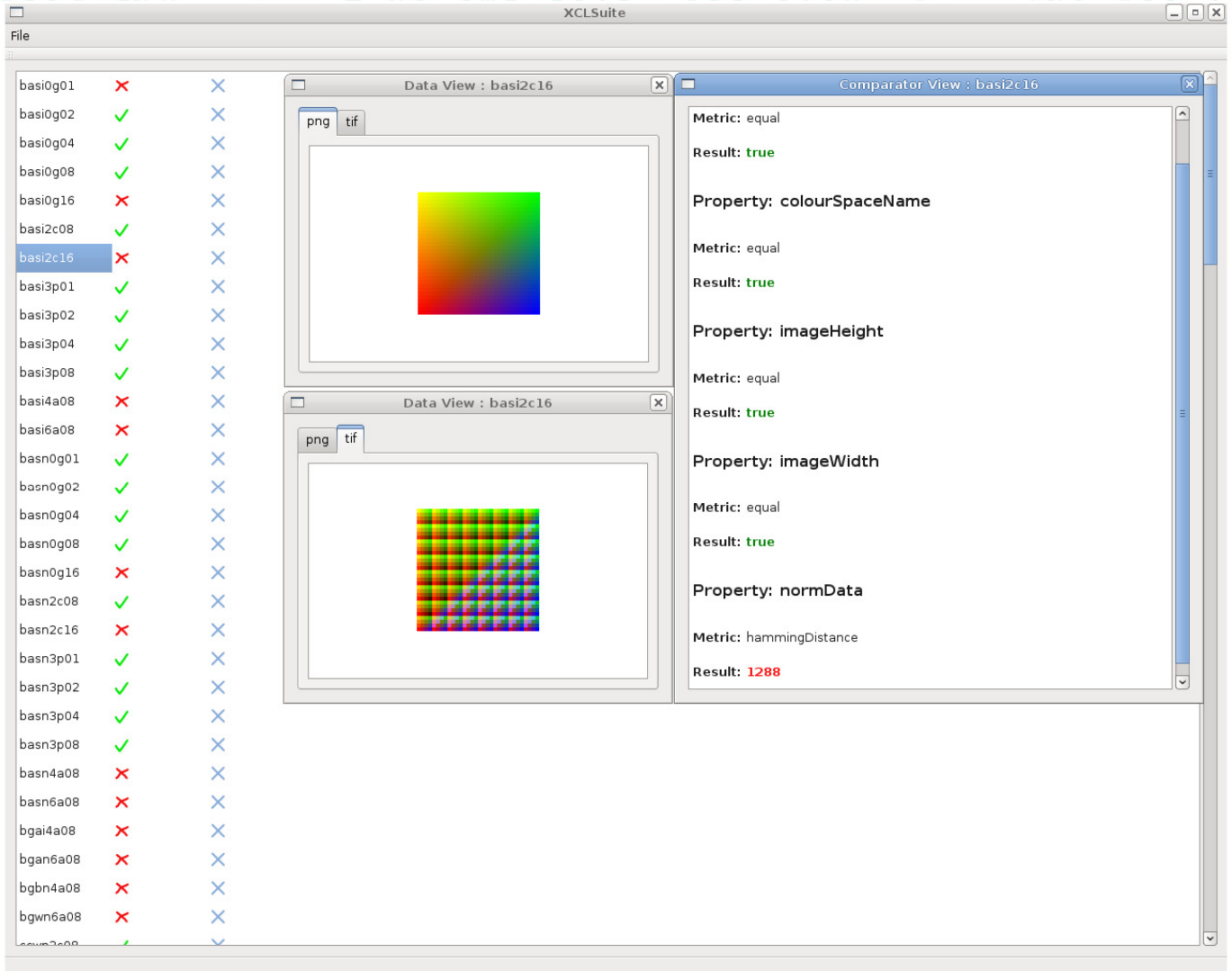
➔both available as command-line-tools and wrapped as services in the IF-subproject

Introduction to the XCL-Tools

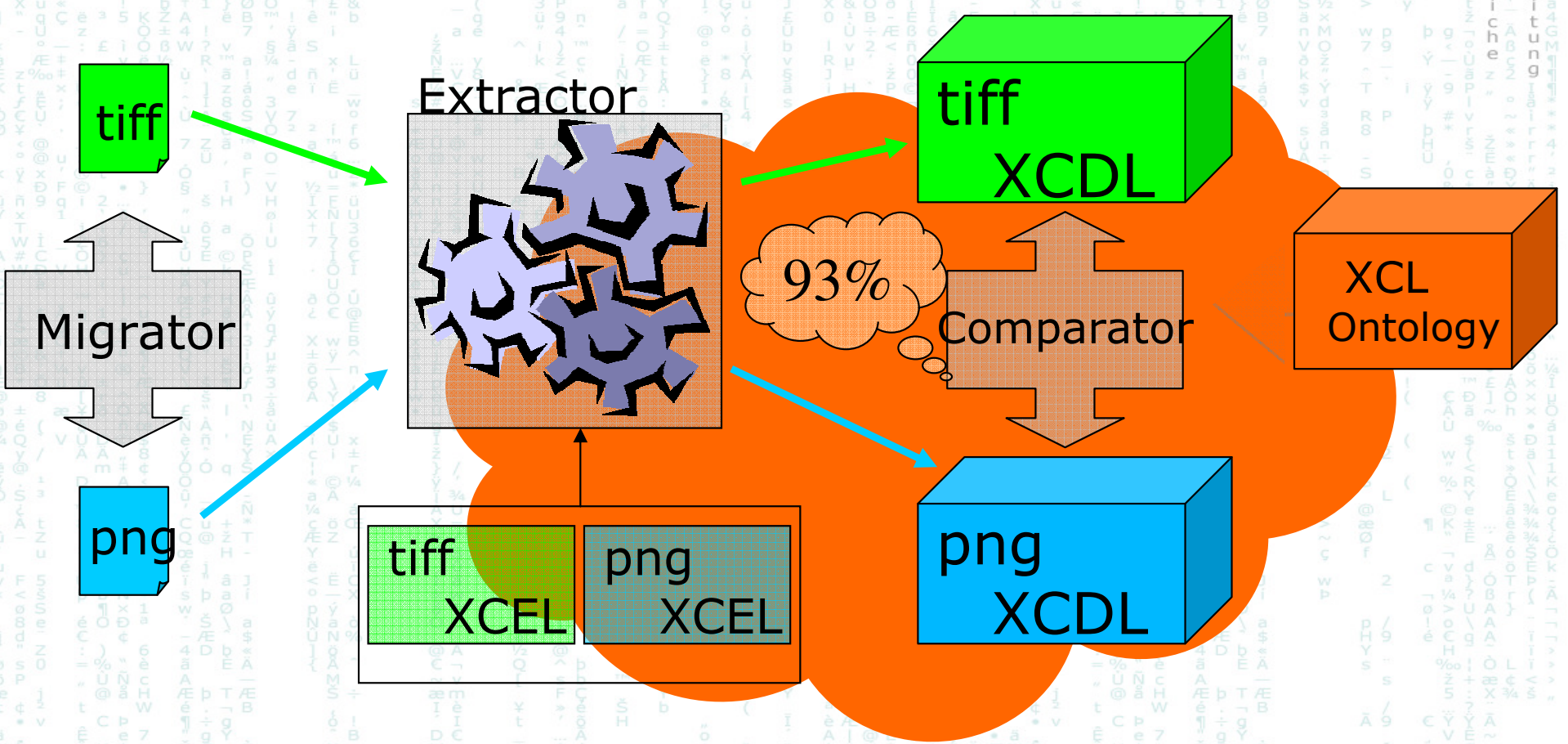
- The XCL-tools know „normdata“



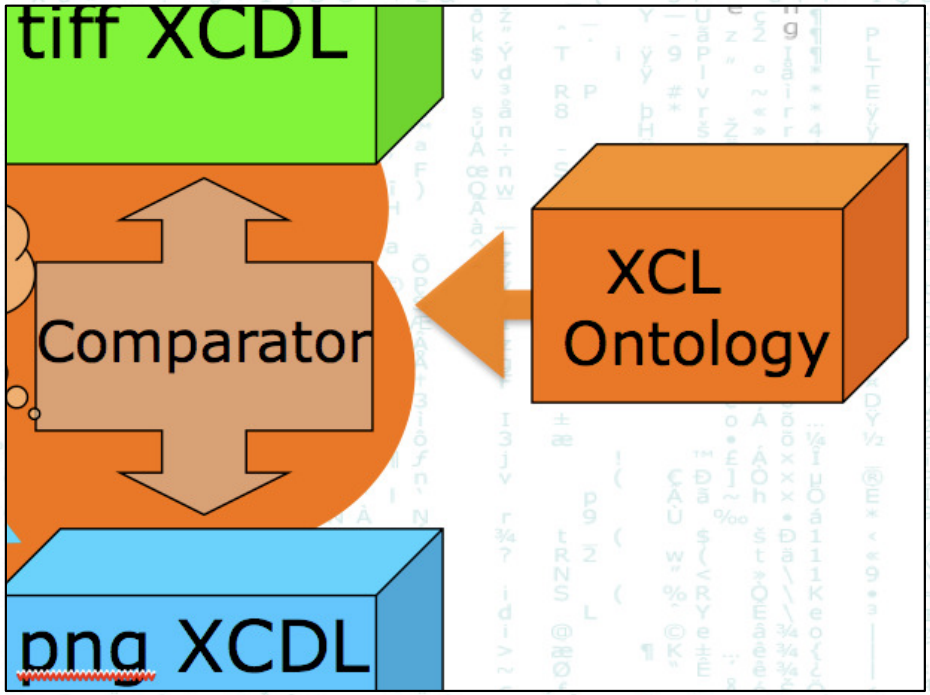
Properties in the XCL-Tools



The XCL-Ontology

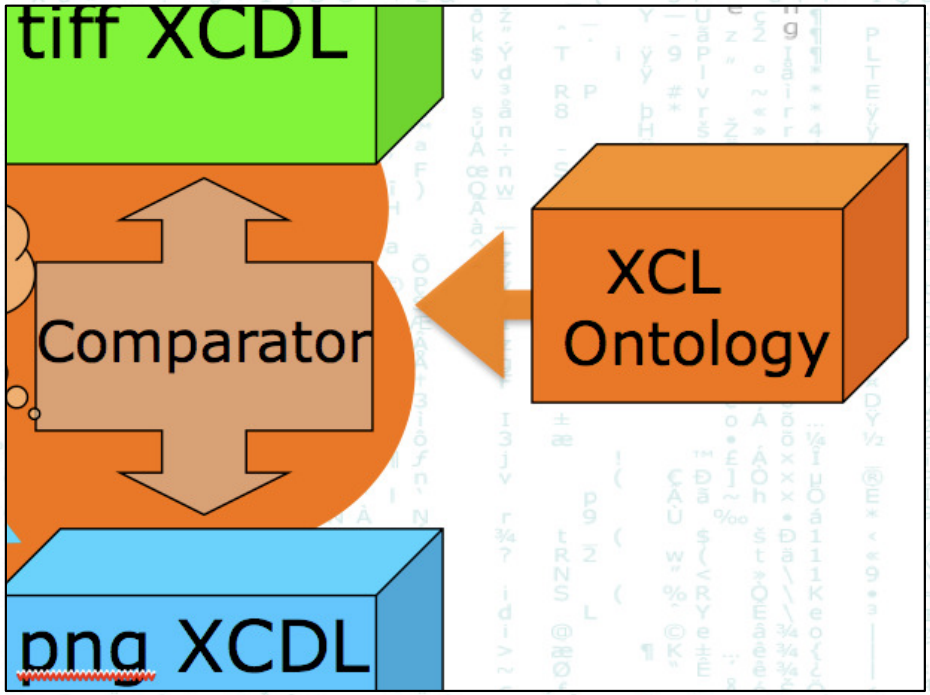


The XCL-Ontology



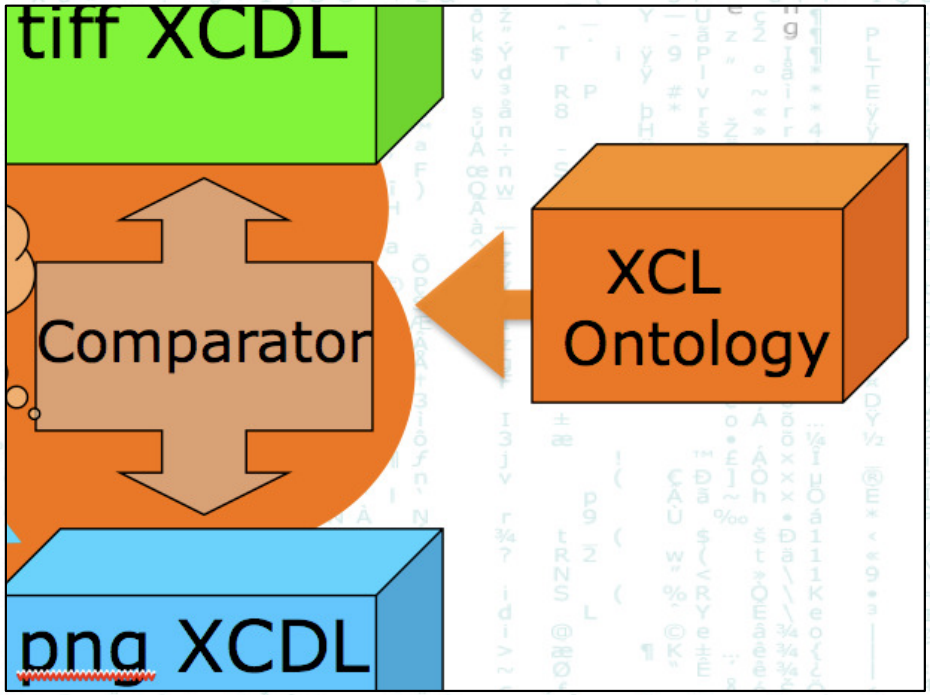
- Lists all available property-names for file-formats, the extractor can process

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:

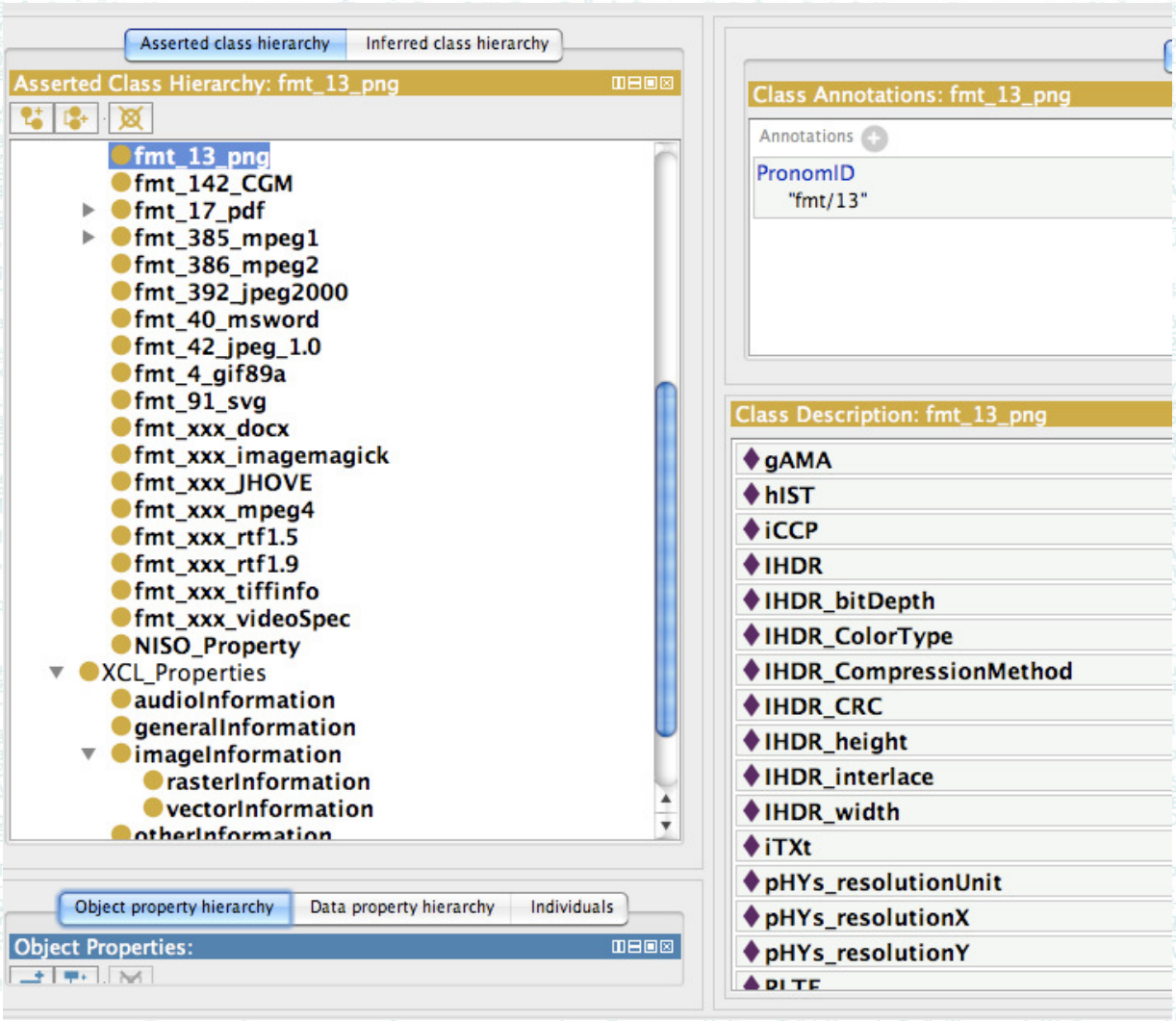
- Classes (= abstract)
- Individuals (=concrete instances of these)
- Object-Properties (relationships between individuals of different classes)

And what about datatypes and units?

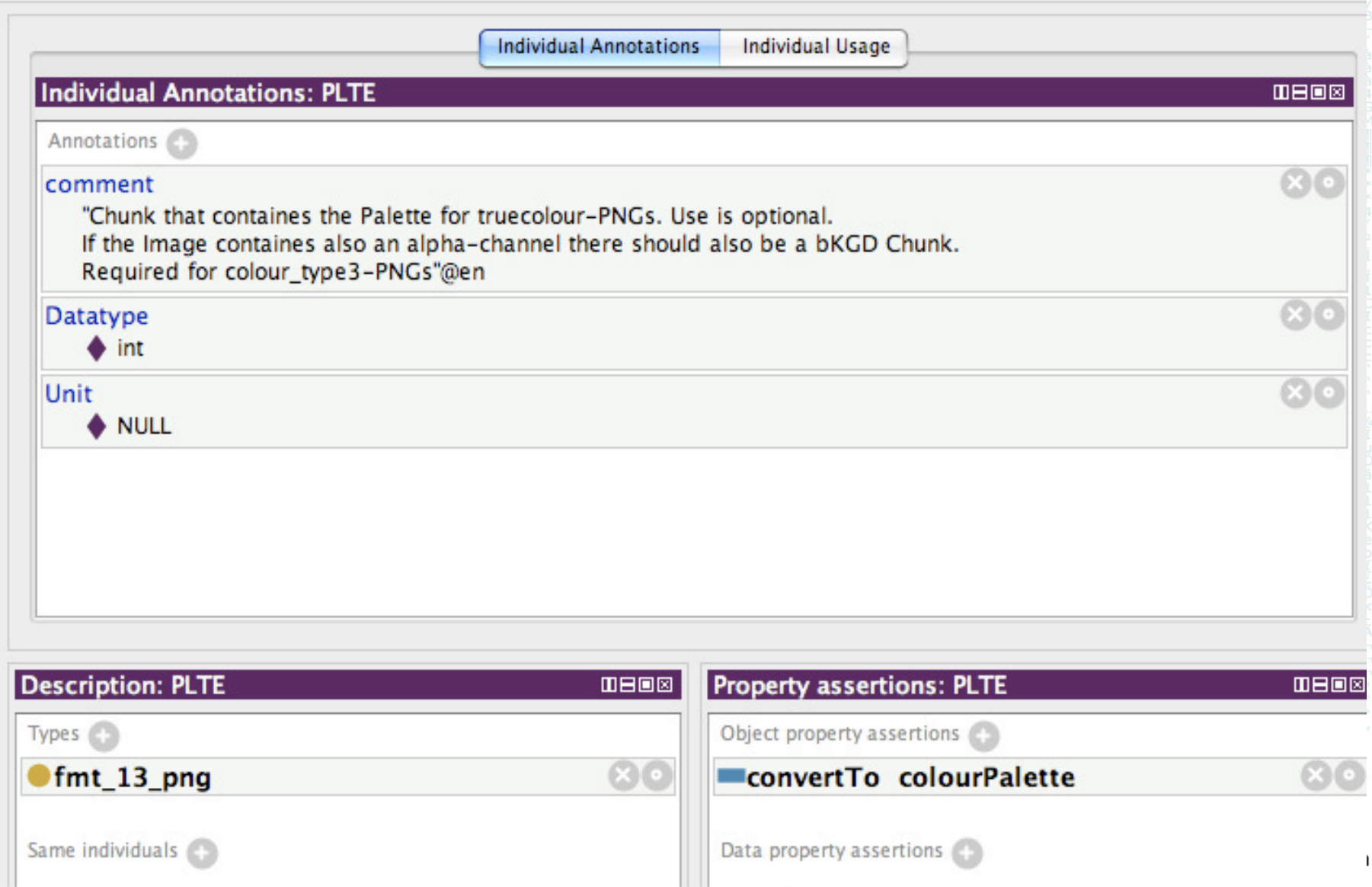
Abstract concepts as Classes:

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

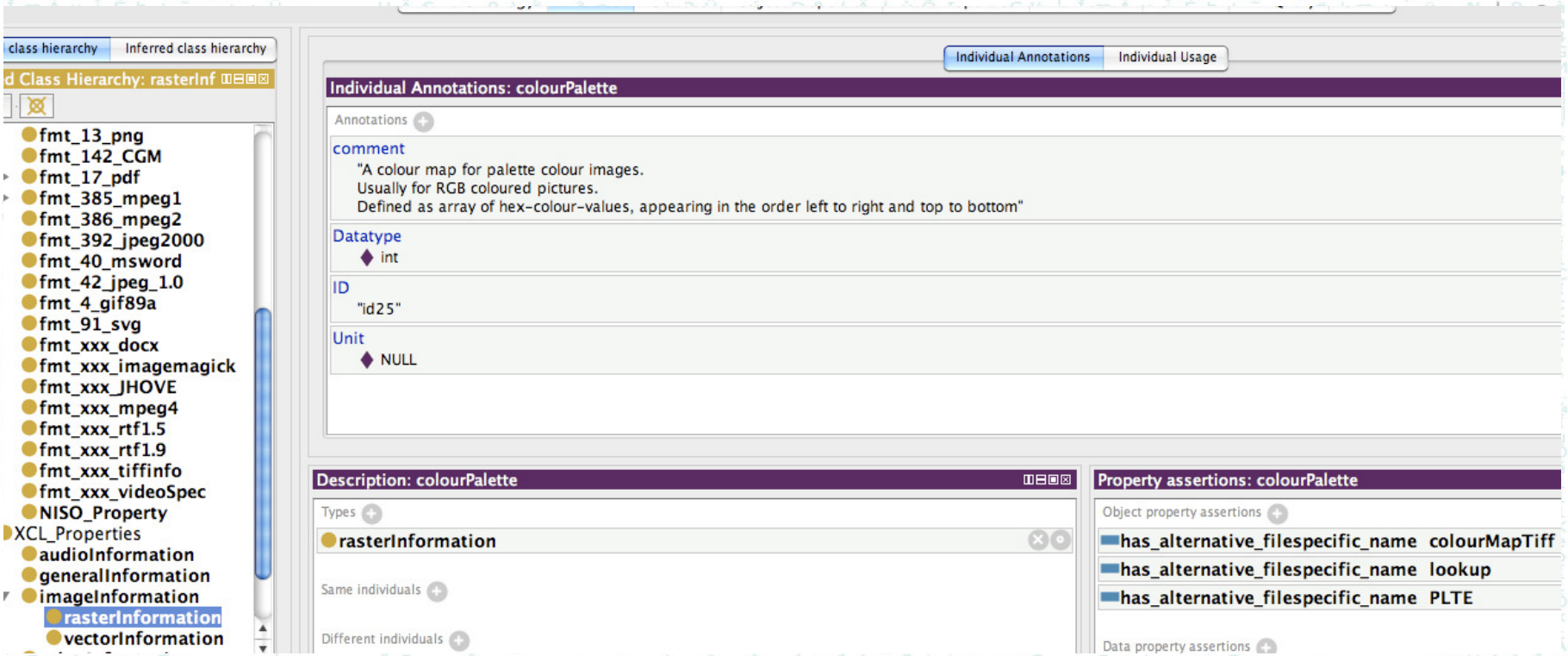
How does it look like?



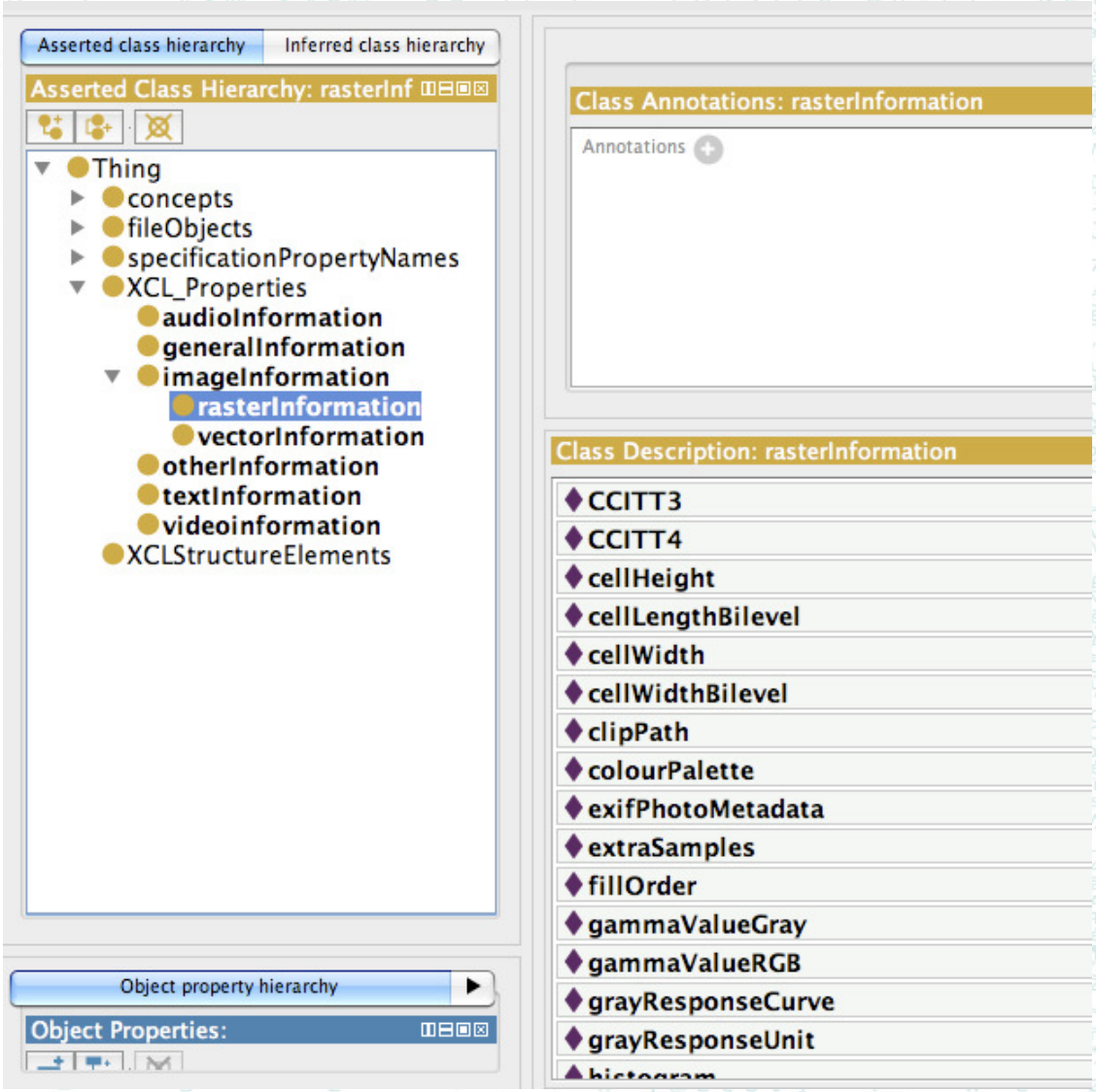
How does it look like?



How does it look like?



How does it look like?



How does it look like?

```
<!— 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>
```

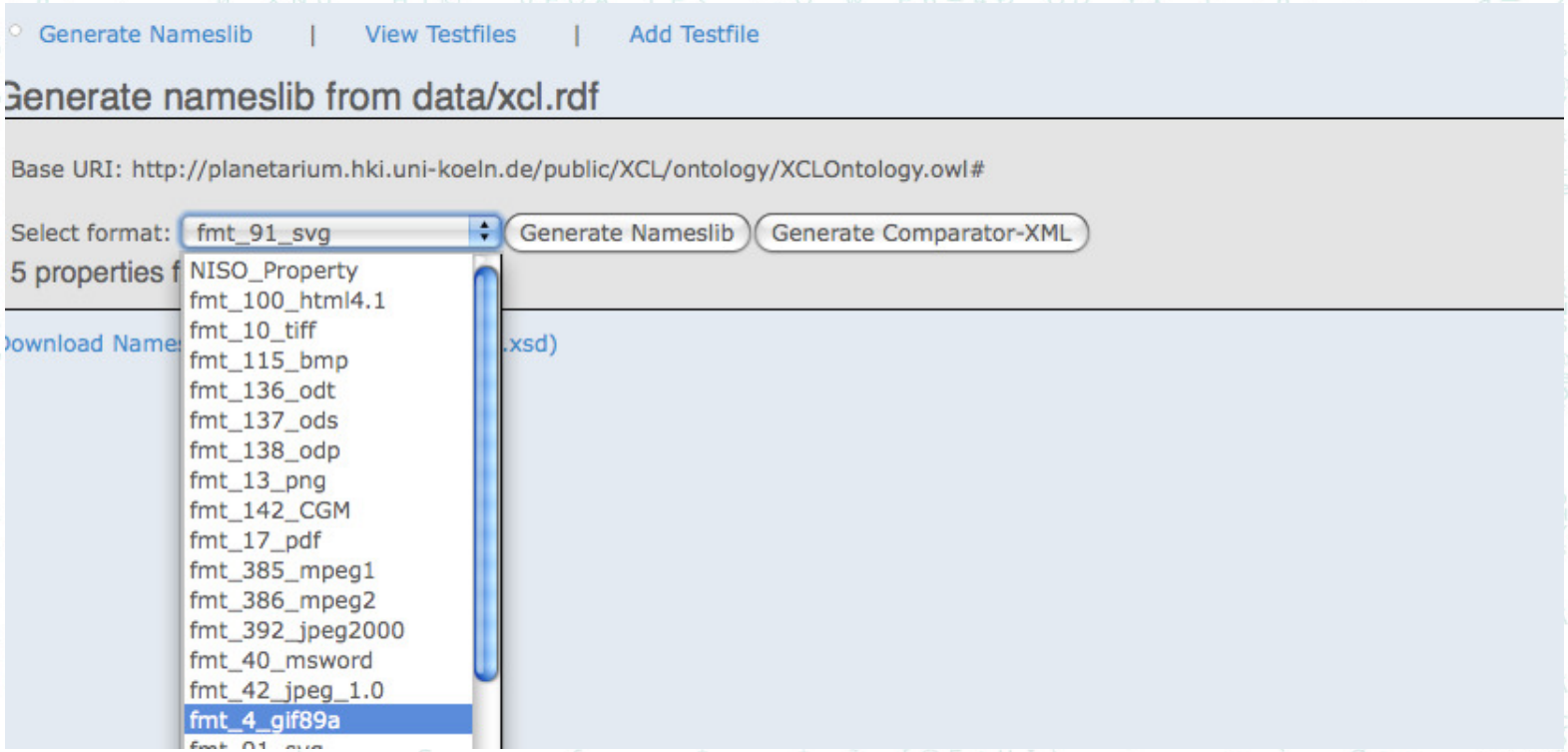
How does it look like?

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

  <fmt_13_png rdf:about="#PLTE">
    <rdfs:comment xml:lang="en"
      >Chunk that containes the Palette for truecolour-PNGs. Use is optional.
If the Image containes 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"/>
  </fmt_13_png>
```

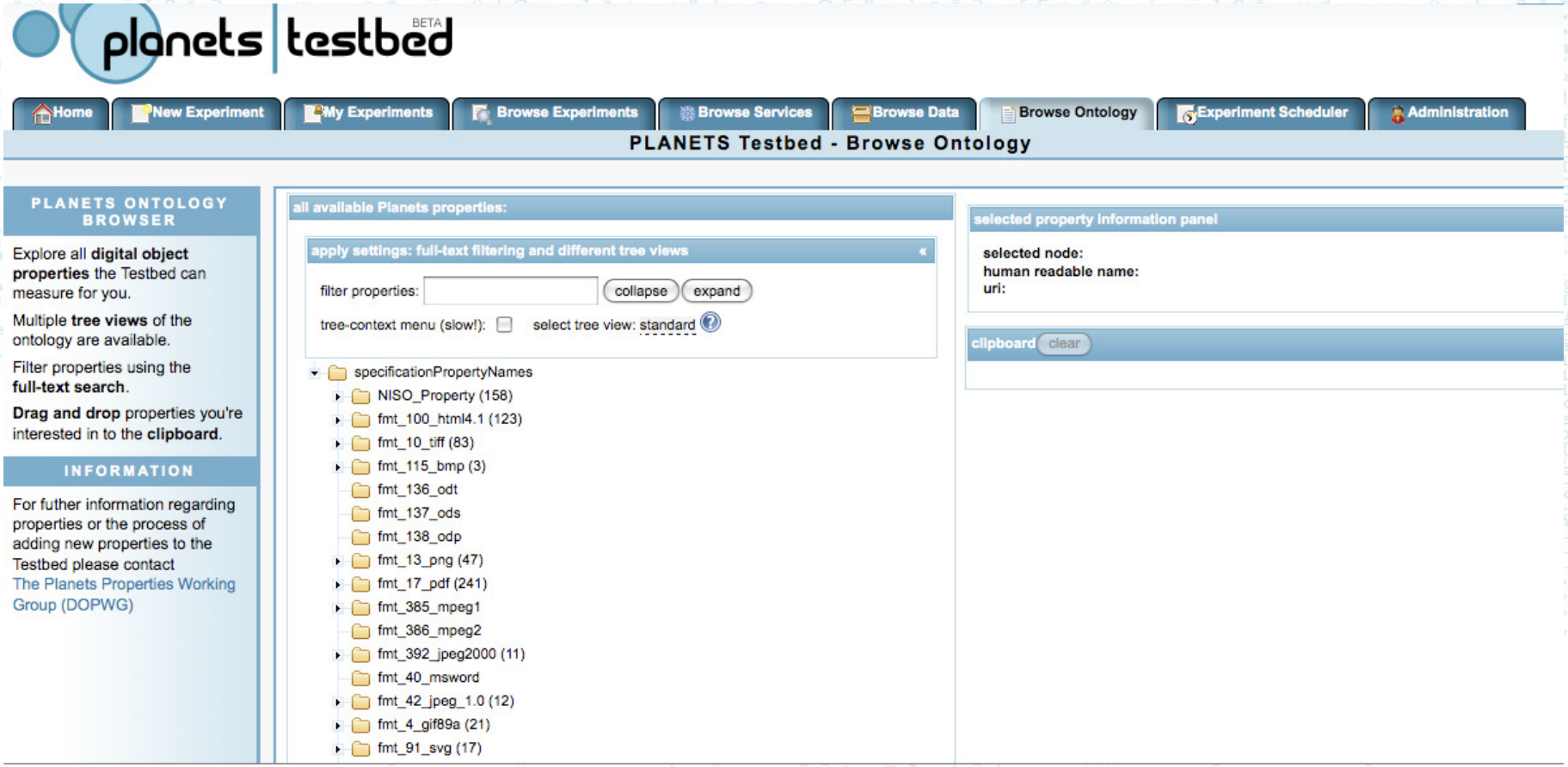
How is it used?

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



How is it used?

- In the current version of the PLANETS-Testbed



Next Steps „PLANETS-wide“ Ontology

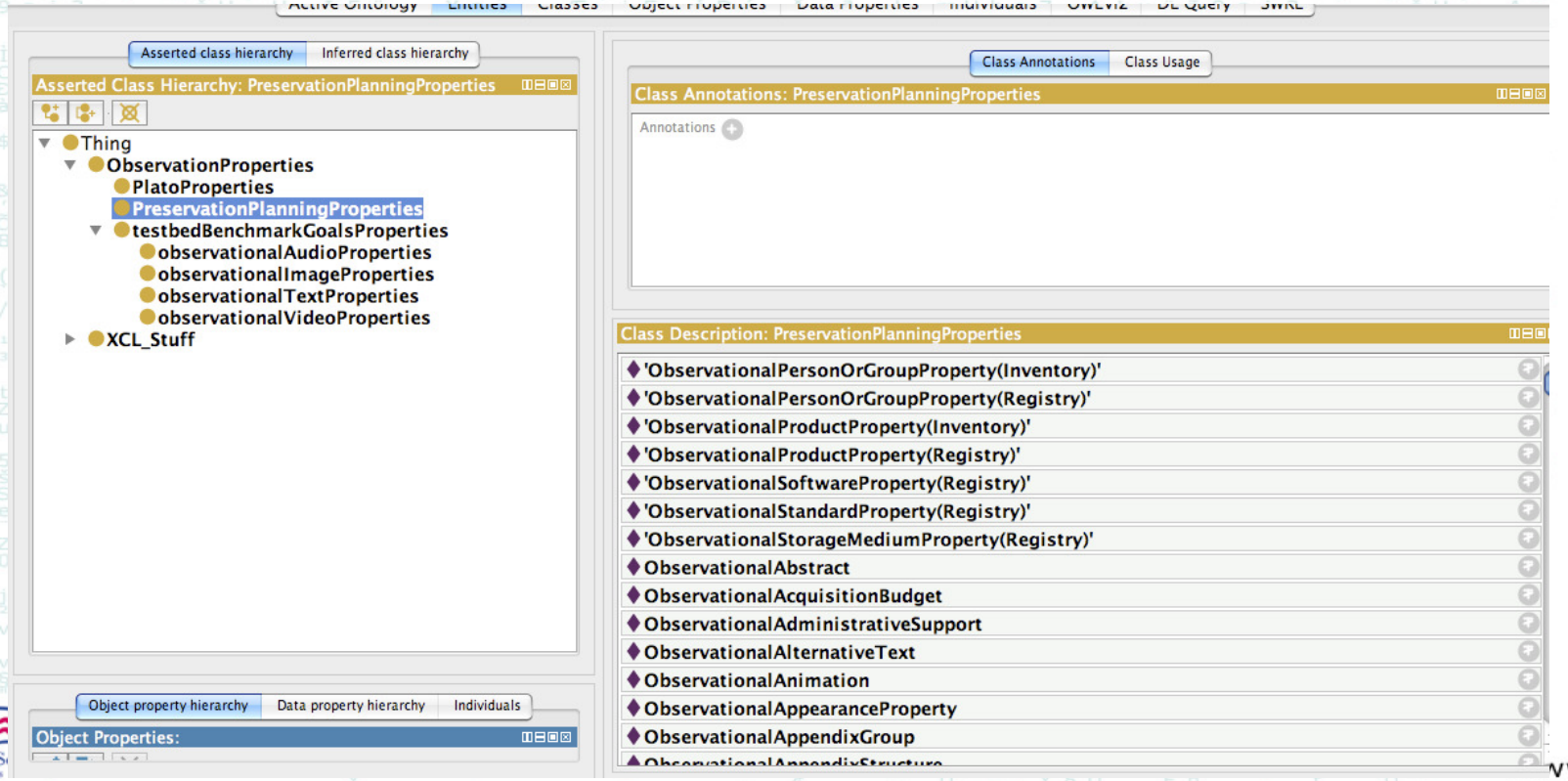
- Include more properties (*non-extractable* or *observational*) from all PLANETS-Subprojects and external projects

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*)



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/