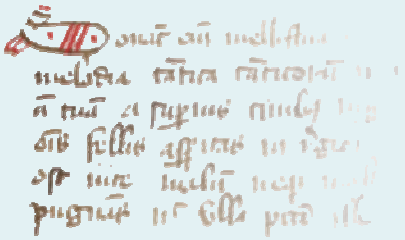


Historisch

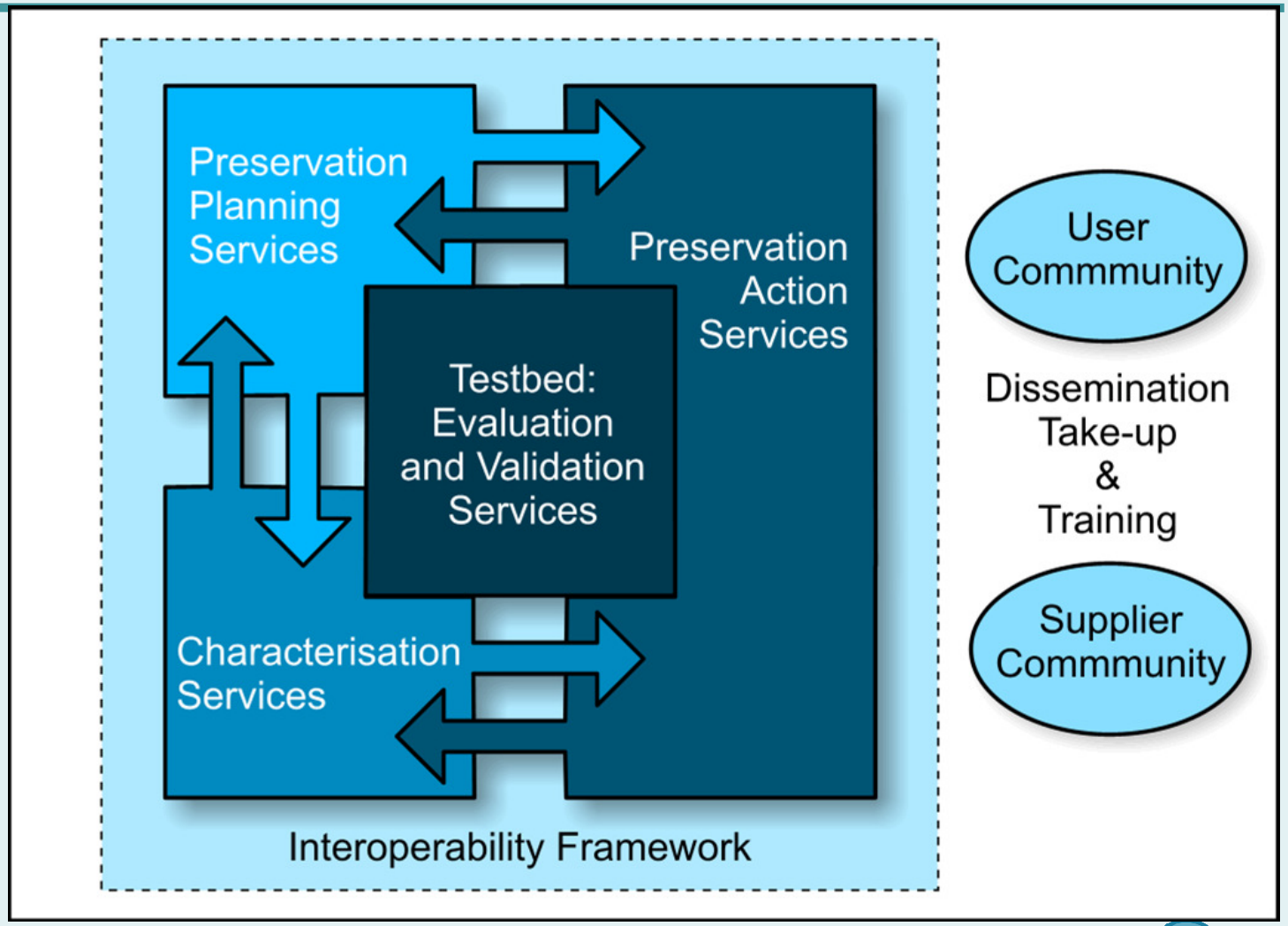
Kulturwissenschaftliche

Informationsverarbeitung

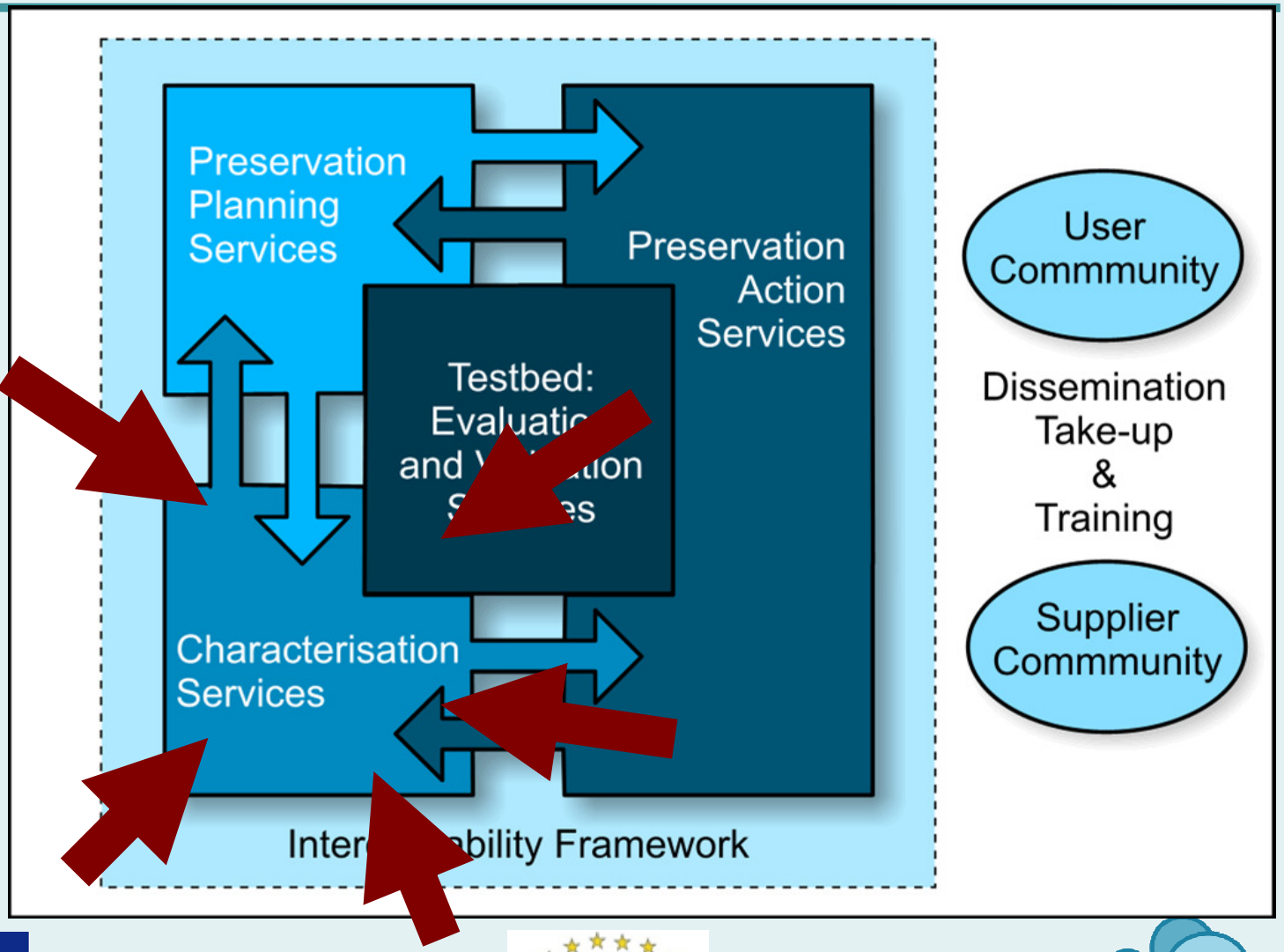
Tools: How to understand files!

Jan Schnasse

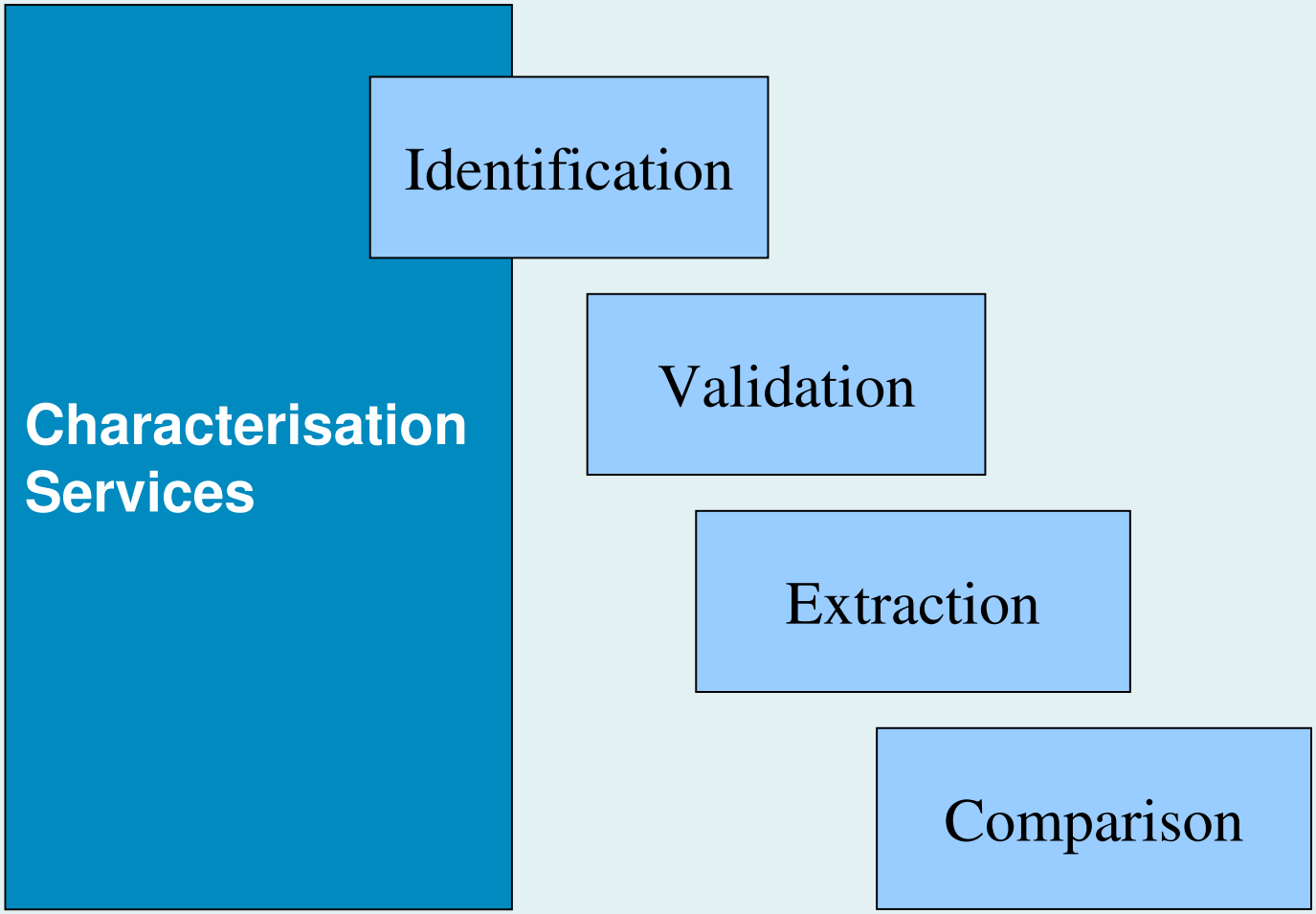
Zoom In



Zoom In


Preservation Planning Services
Preservation Action Services
Testbed: Evaluatio
and V ation
S es
Characterisation Services
Inter bility Framework
User Commmunity
Dissemination
Take-up
&
Training
Supplier Commmunity

Zoom In

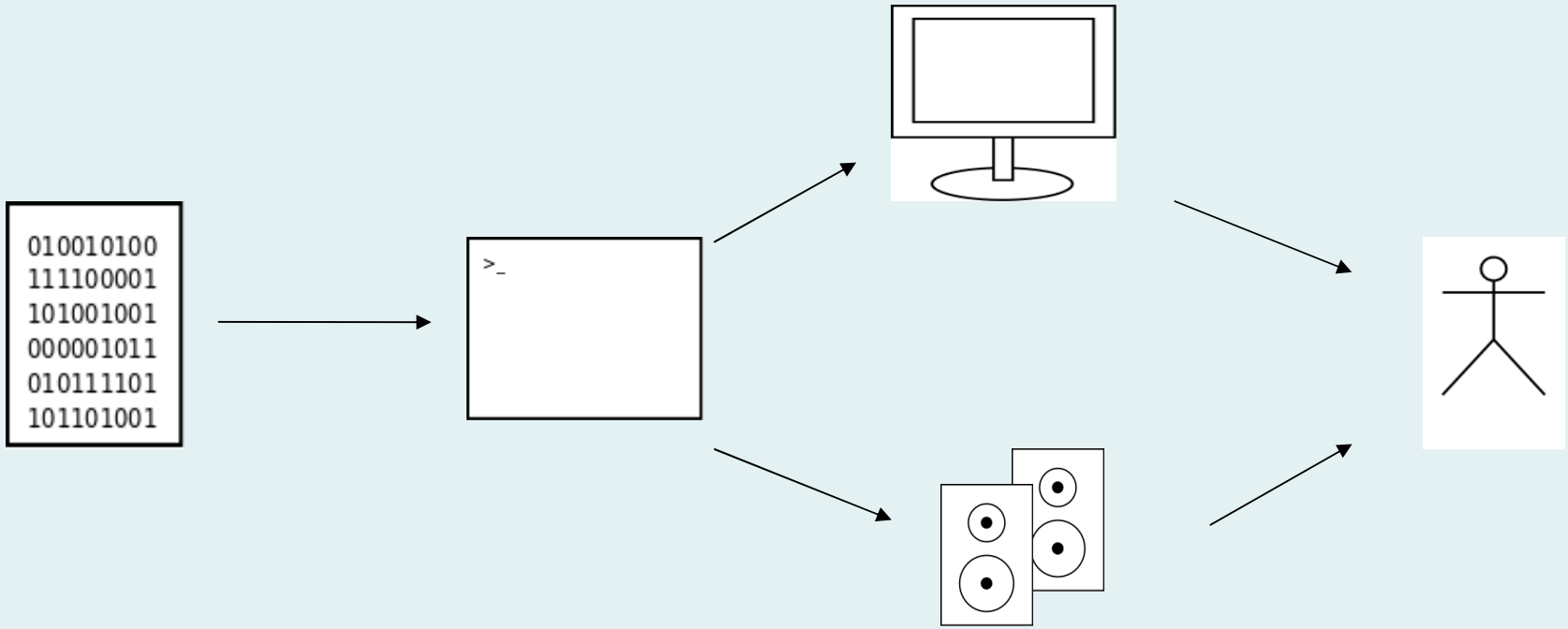

Characterisation Services
Identification
Validation
Extraction
Comparison

Focus

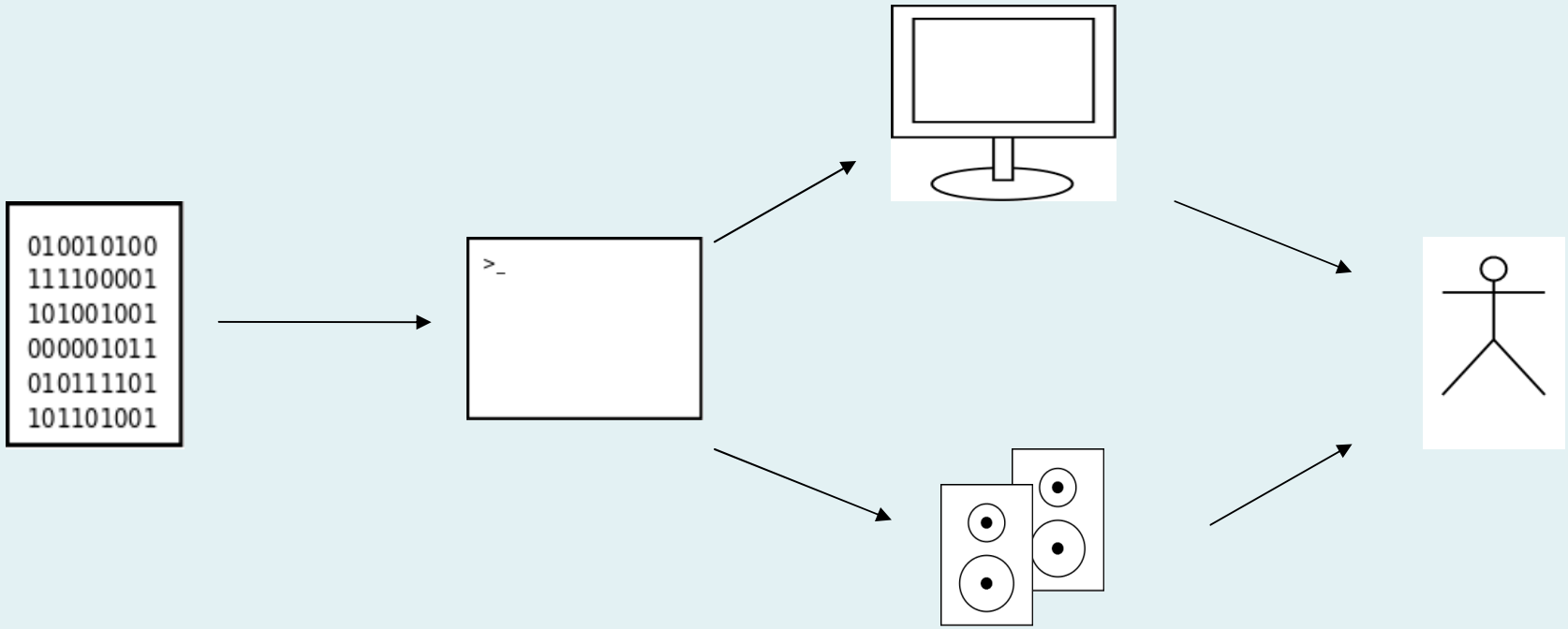
Extraction

Comparison

Data, Perception, Information



Data, Perception, Information



Data Representation

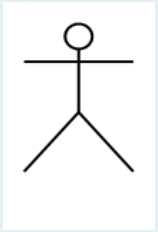
Processing

Presentation

Perception

What to preserve?

Perception



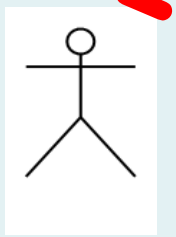
must be preserved

What to preserve?

What do you see?

What to preserve?

Perception

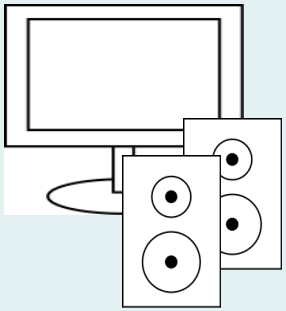


must be preserved

unpreservable

What to preserve?

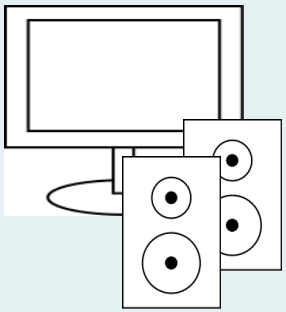
Presentation



must be preserved

What to preserve?

Presentation

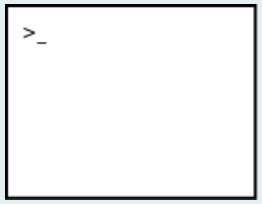


must be preserved

adequate
accurate
authentic
original
significant

What to preserve?

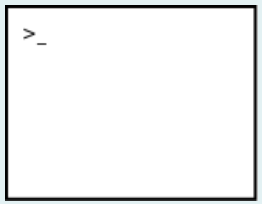
Processing



must be preserved

What to preserve?

Processing

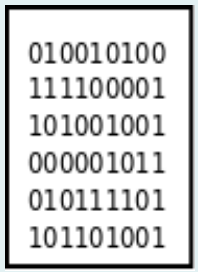


must be preserved →

sufficient

What to preserve?

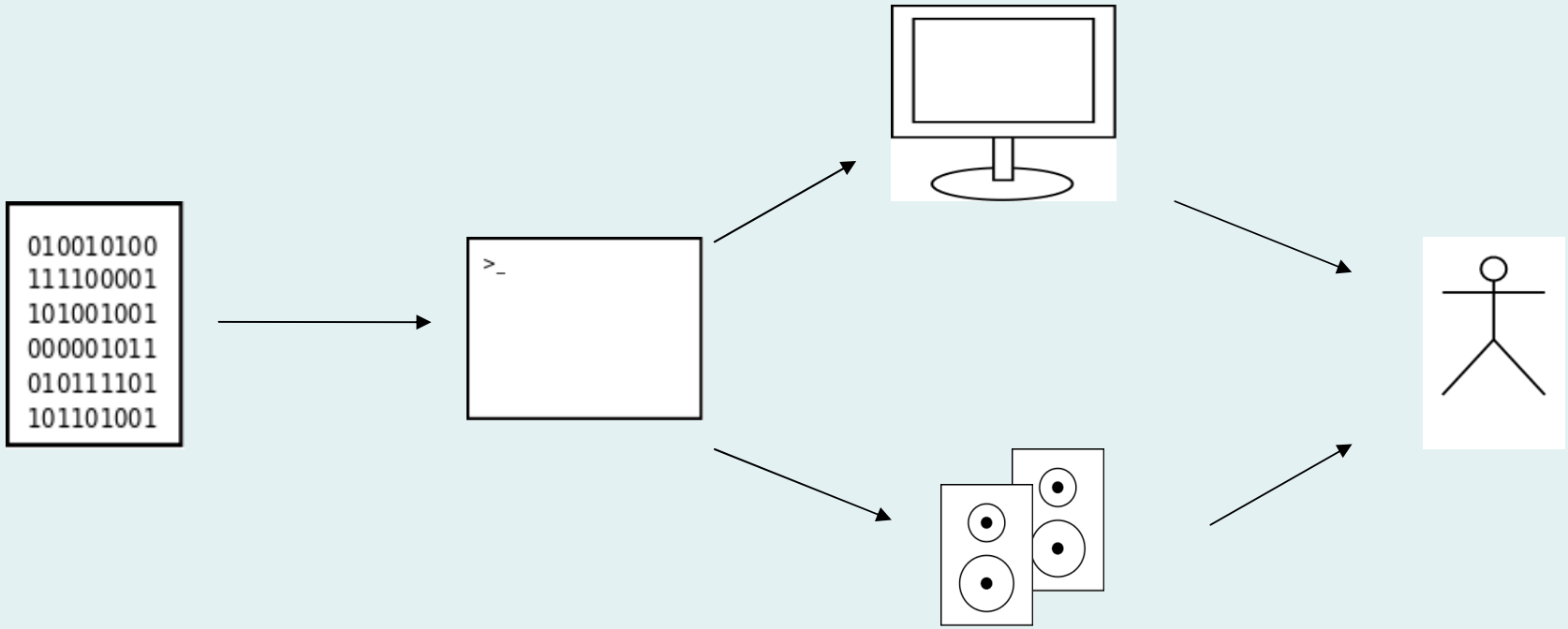
Data Representation



must be preserved →

arbitrary
but
complete!

Data, Perception, Information



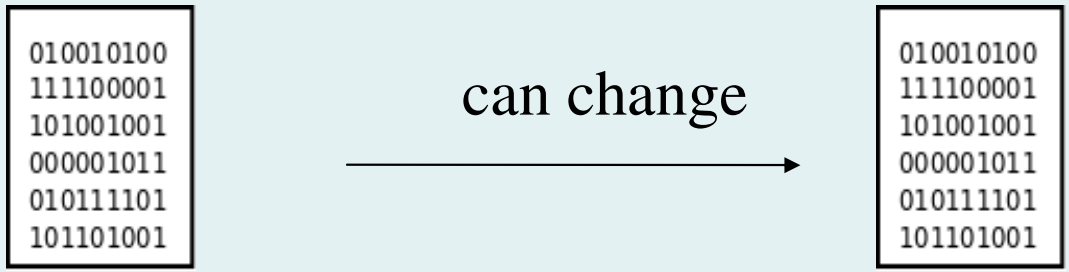
Data Representation

Processing

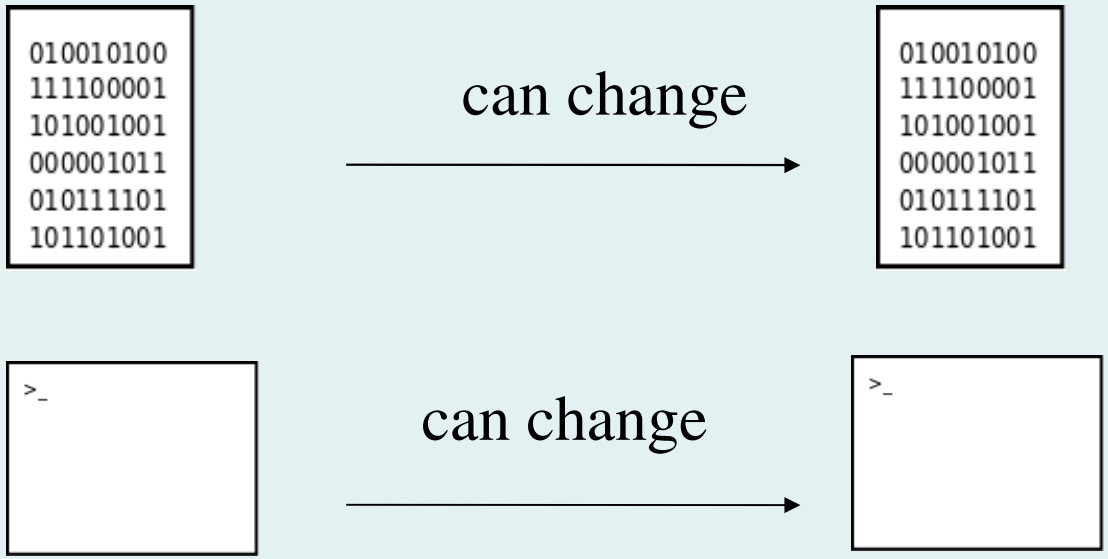
Presentation

Perception

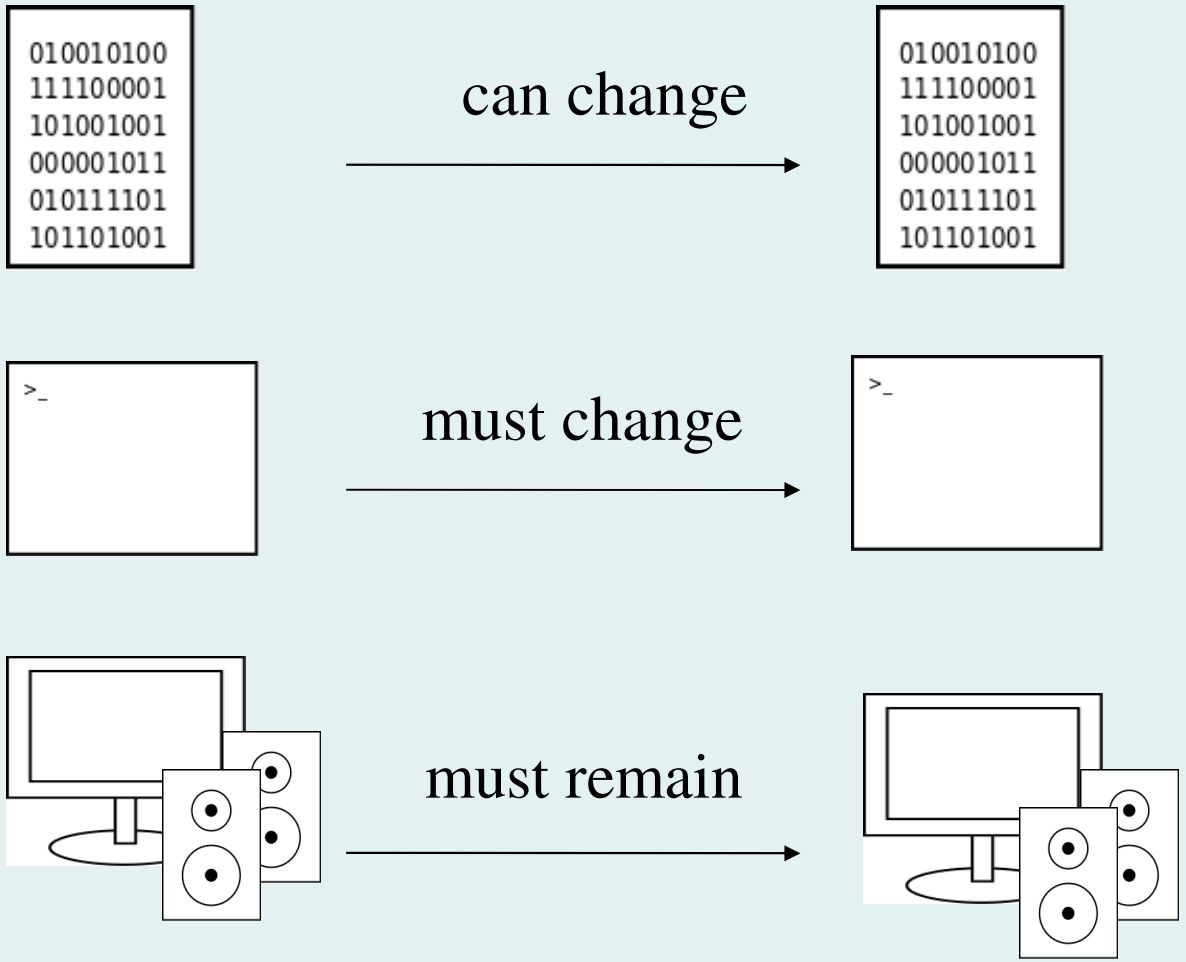
Preservation Scenario: Format Conversion



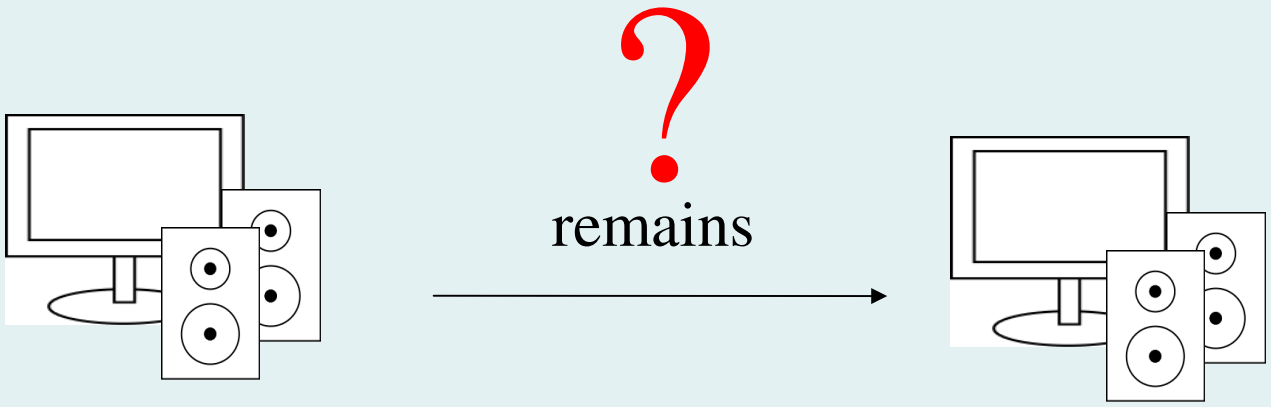
Preservation Scenario: Format Conversion



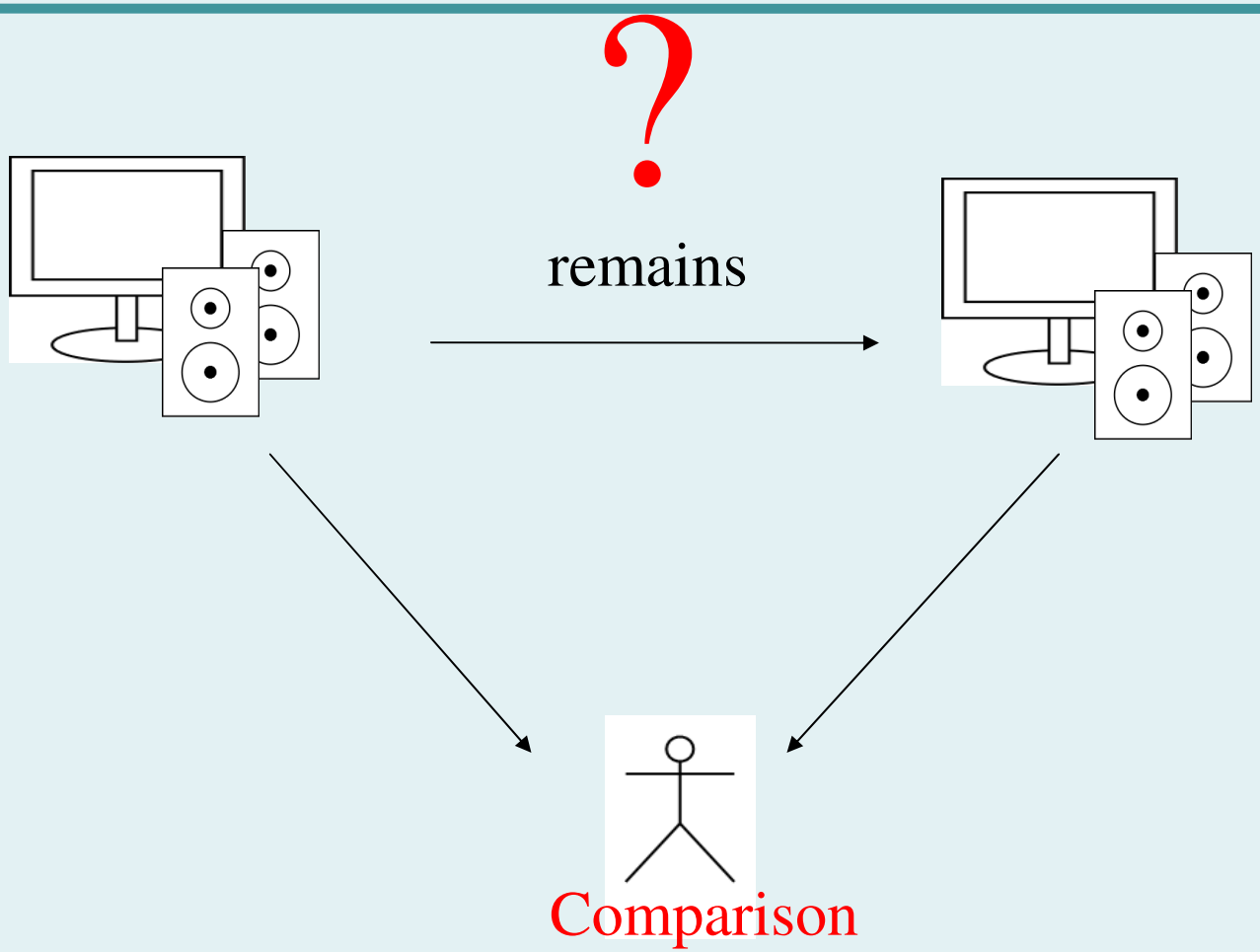
Preservation Scenario: Format Conversion



Preservation Scenario: Format Conversion



Evaluation of Format Conversion



Evaluation of Format Conversion

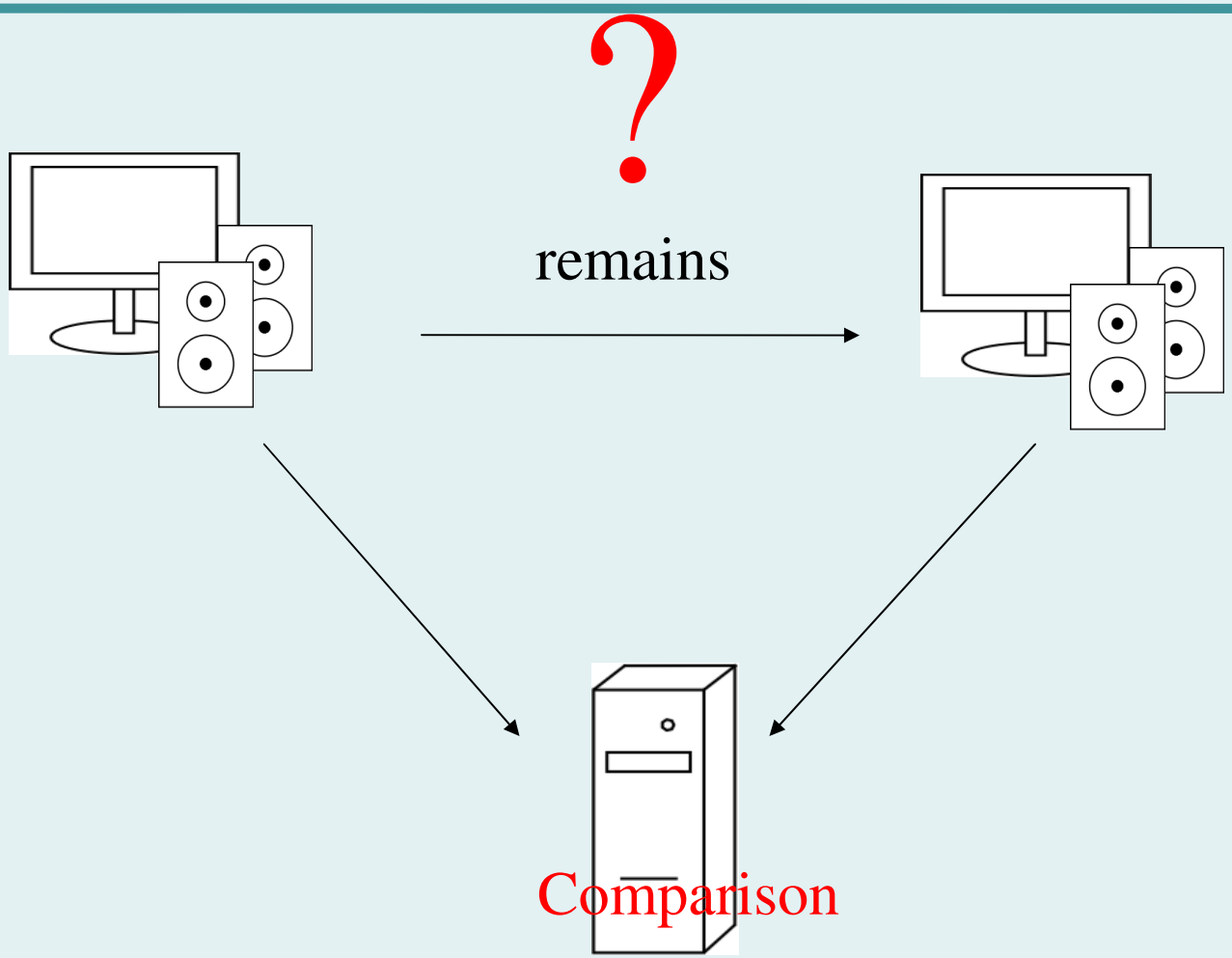
Why automate?

1 million objects: use five minutes for each.

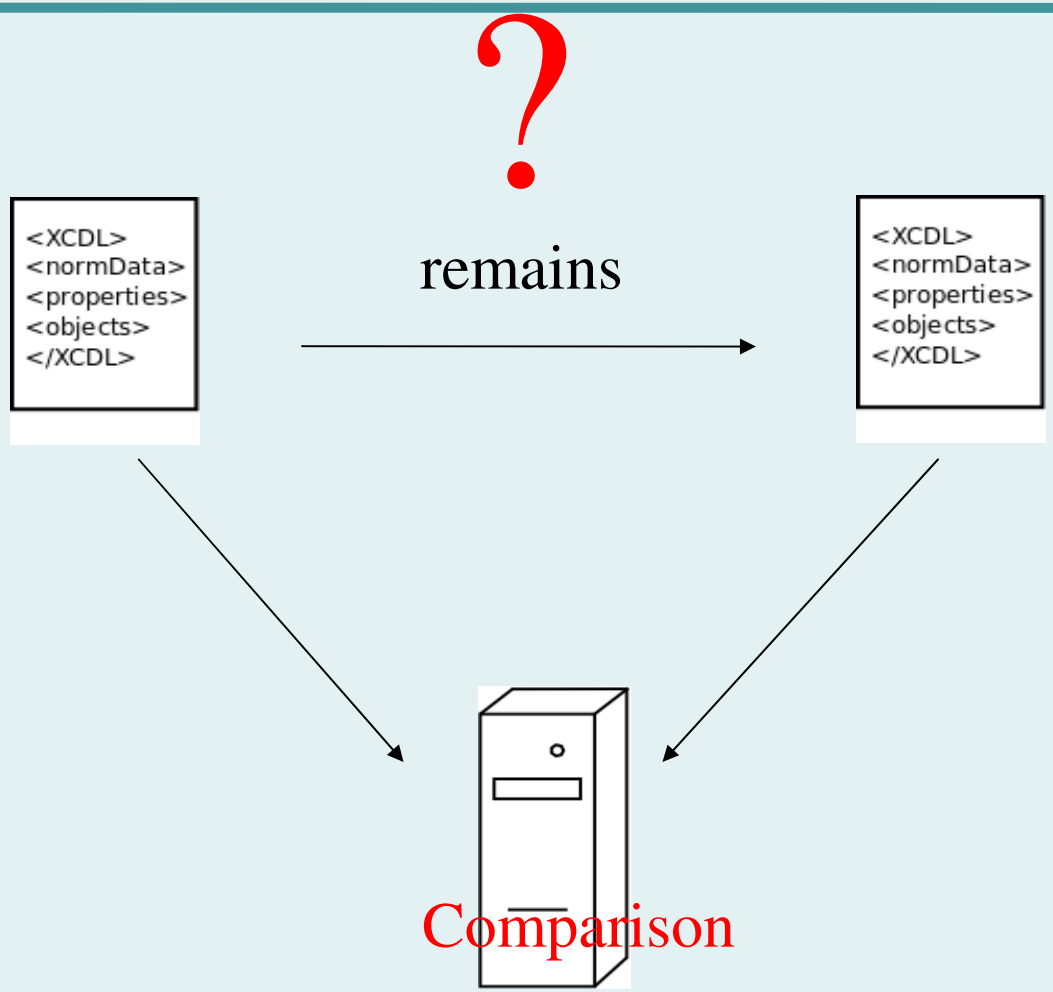
== 416 666.7 hours

== 52 803.4 8-hour days for a Human
== way too much for anything

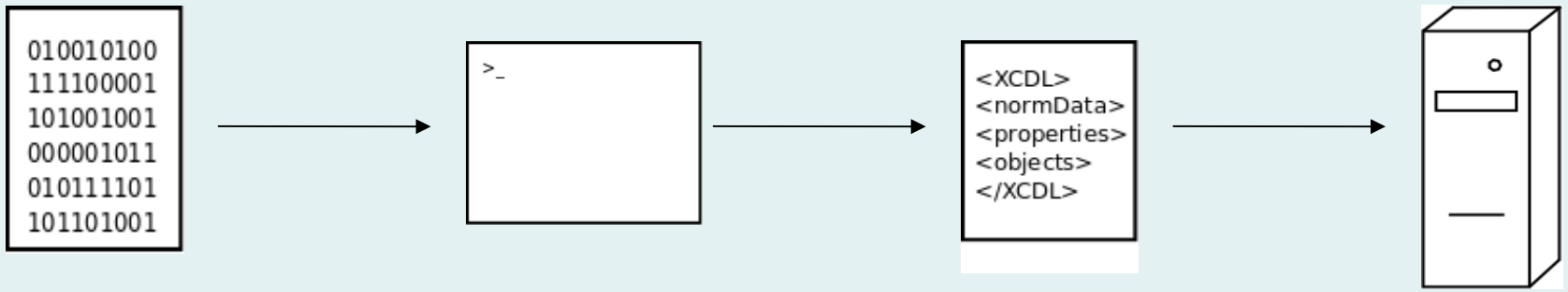
Evaluation of Migration



Evaluation of Format Conversion



Evaluation of Format Conversion



Data Representation

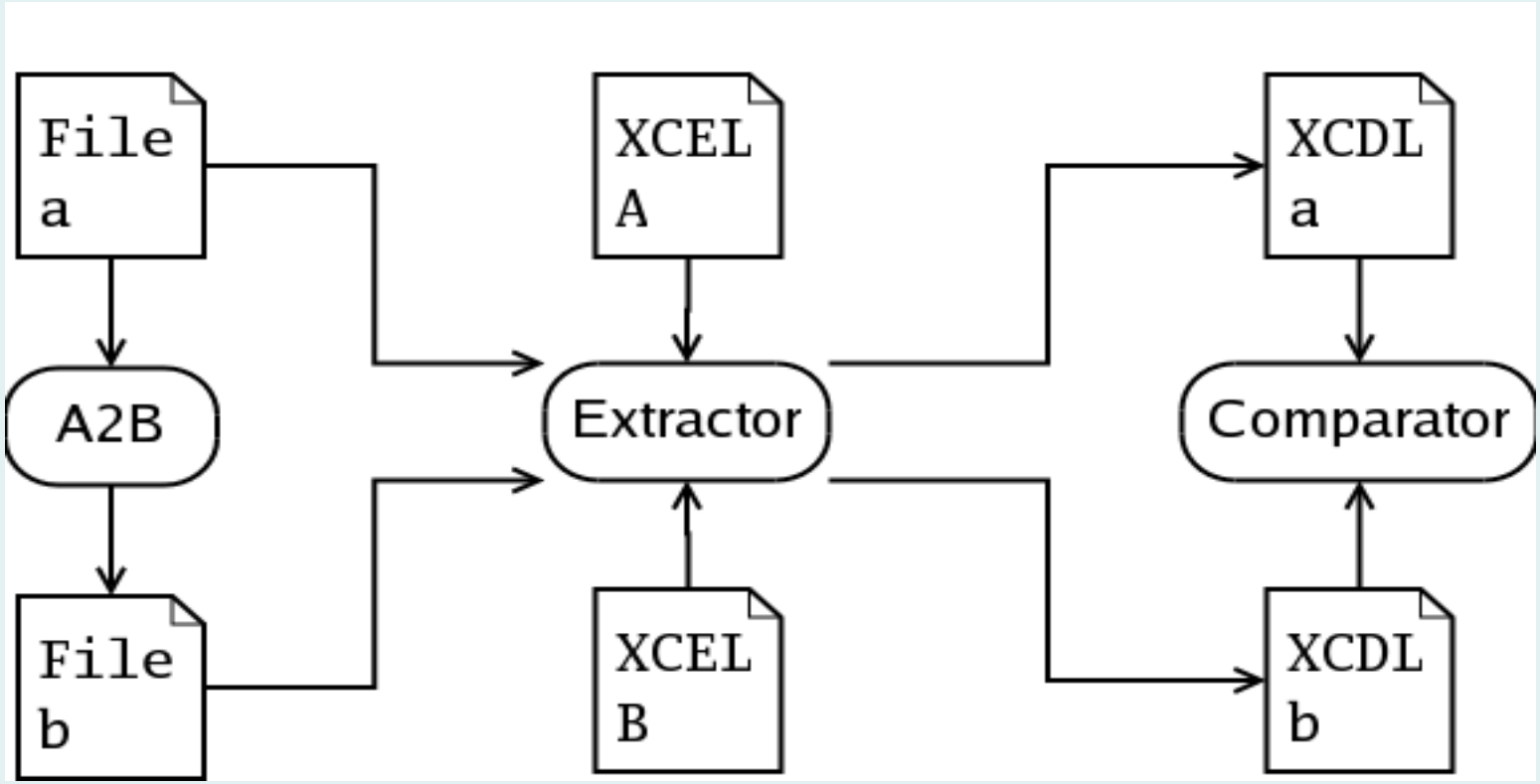
Processing

Machine-readable Presentation

Calculation

The Planets XCL Approach - Overview

Evaluation of Format Conversion



The Planets XCL Approach – Describing File Formats

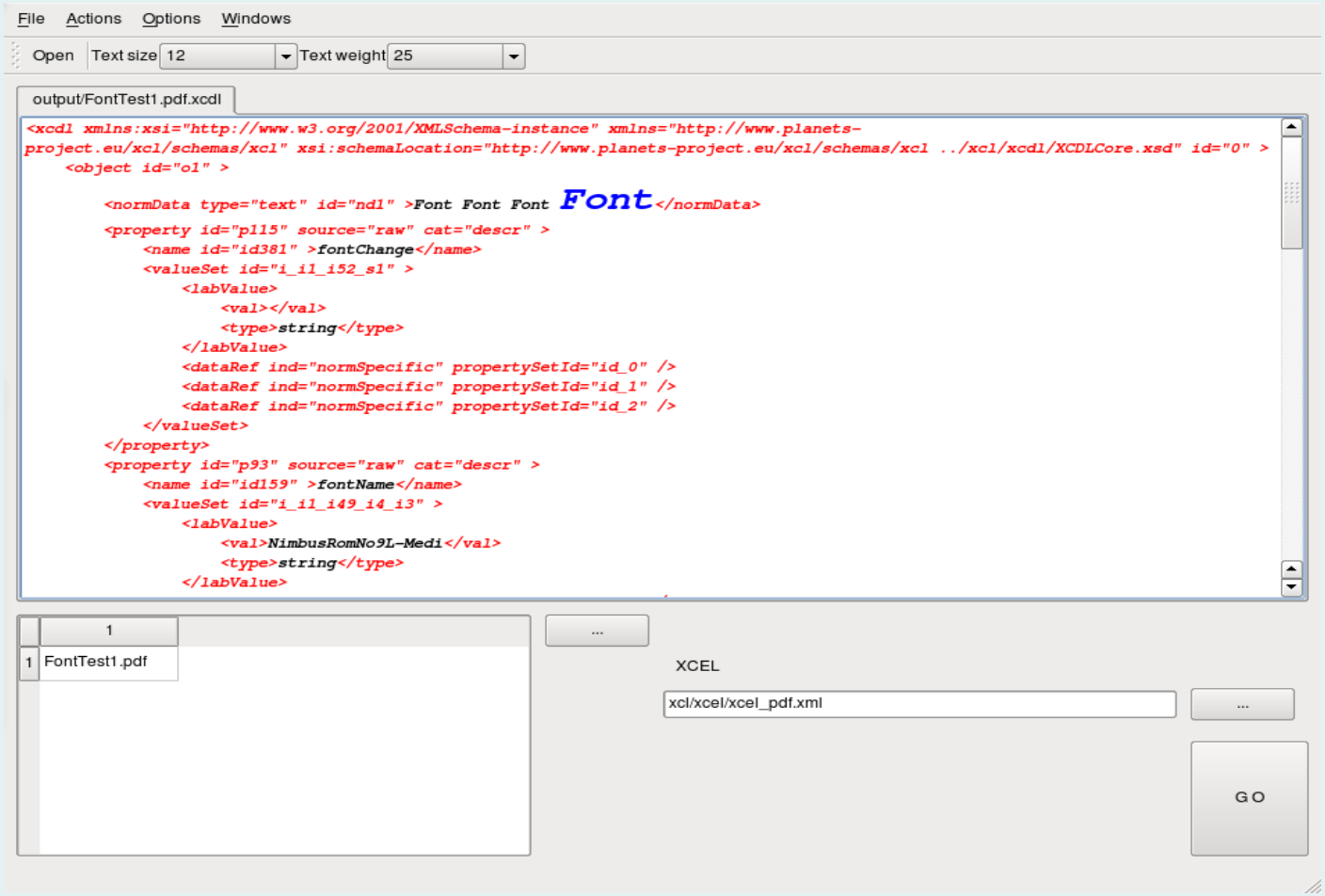
XCEL (Extensible Characterisation Extraction Language)

```
<symbol value="137 80 78 71 13 10 26 10"/>

<symbol interpretation="uint32" length="4"/>

<symbol value="IHDR" interpretation="ASCII">

 <symbol interpretation="uint32"
      name="imageWidth" length="4"/>
```

Natural Language

„The first eight bytes of a PNG datastream always contain the following (decimal) values: 137 80 78 71 13 10 26 10 [...] The four-byte chunk type field contains the decimal values 73 72 68 82[.] The IHDR chunk shall be the first chunk in the PNG datastream. It contains: Width 4 bytes [...]
Width and height give the image dimensions in pixels.
They are PNG four-byte unsigned integers. Zero is an invalid value.“ (http://www.w3.org/TR/PNG/)

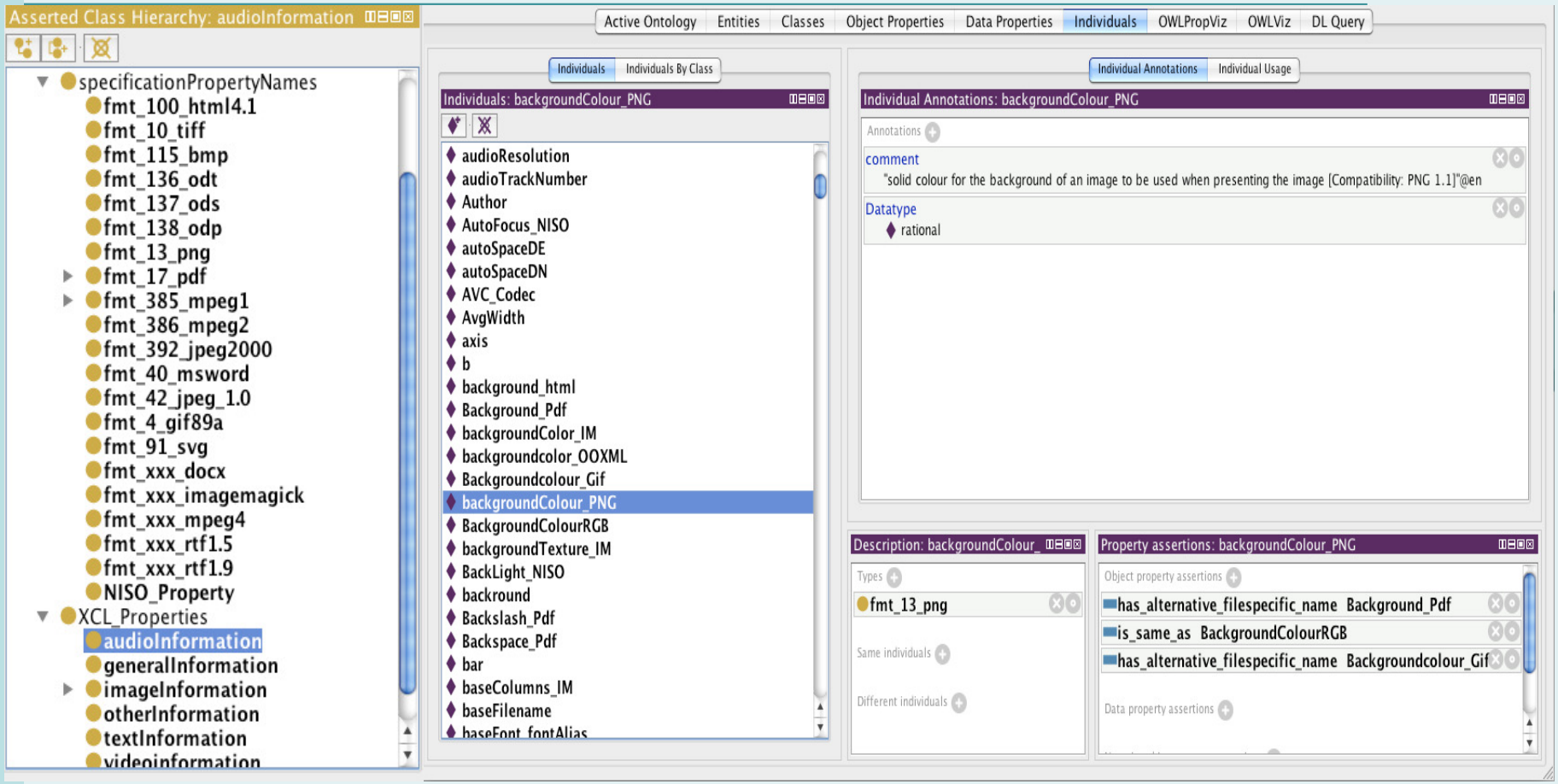
The Planets XCL Approach - Extractor



The Planets XCL Approach – General Data Representation with XCDL (Extensible Characterisation Definition Language)

```
    <object id="o1" >
        <normData type="image" id="nd1" >00 01 02 03 04 05 06 07 08 09
0a 0b 0c 0d 0e 0f 10 11 12 13 14 15 16 17 18 19 1a 1b 1c 1d ...
        </normData>
        <property id="p13" source="raw" cat="descr" >
            <name id="id2" >imageHeight</name>
            <valueSet id="i_i1_s10" >
                <labValue>
                    <val>32</val>
                    <type>int</type>
                </labValue>
            </valueSet>
        </property>
        <property id="p14" source="raw" cat="descr" >
            <name id="id30" >imageWidth</name>
            <valueSet id="i_i1_s8" >
                <labValue>
                    <val>32</val>
                    <type>int</type>
                </labValue>
            </valueSet>
        </property> ...
    </object>
```

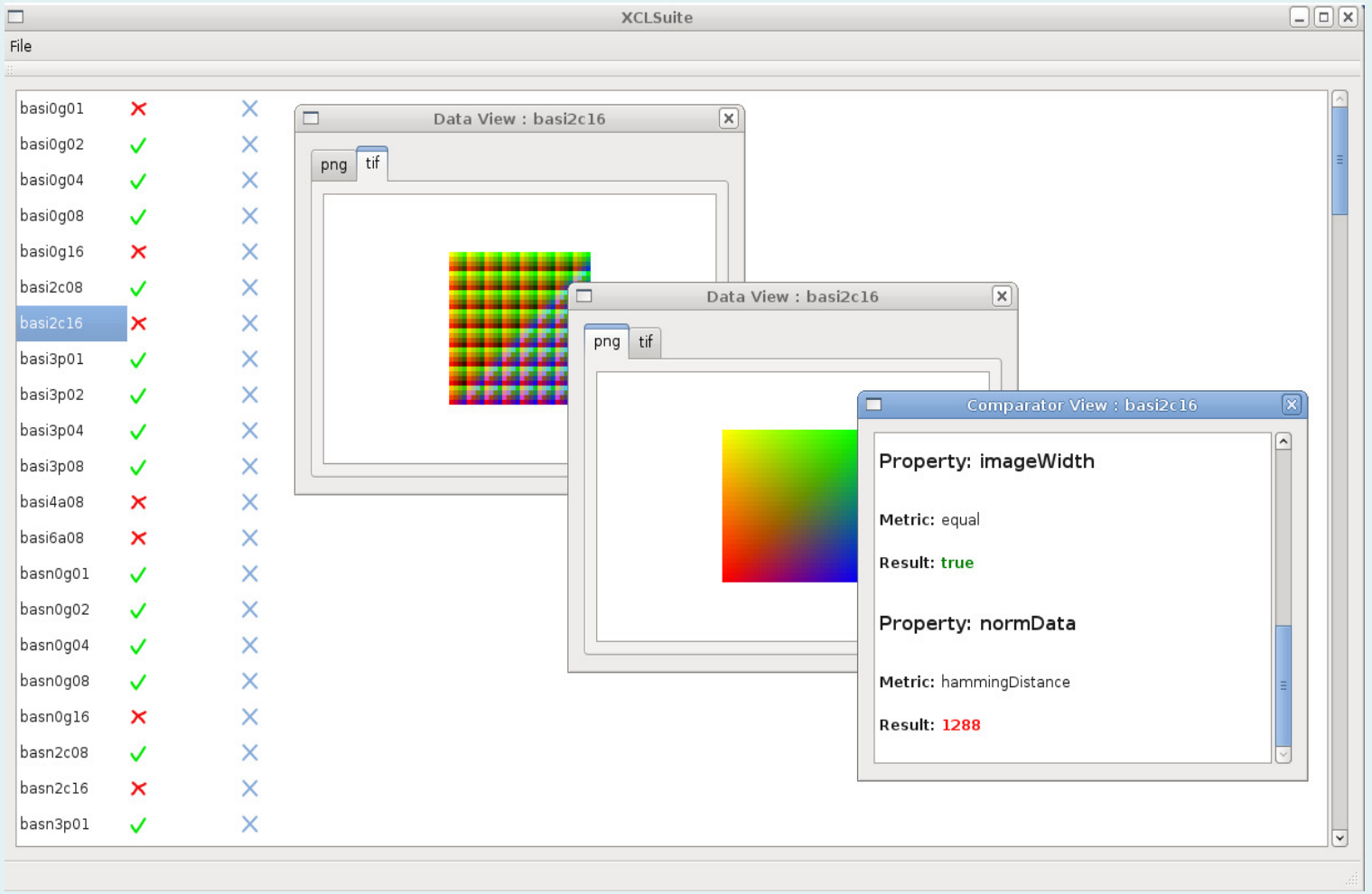
The Planets XCL Approach – The Ontology



CEI
CENTRAL EUROPEAN INITIATIVE

planets

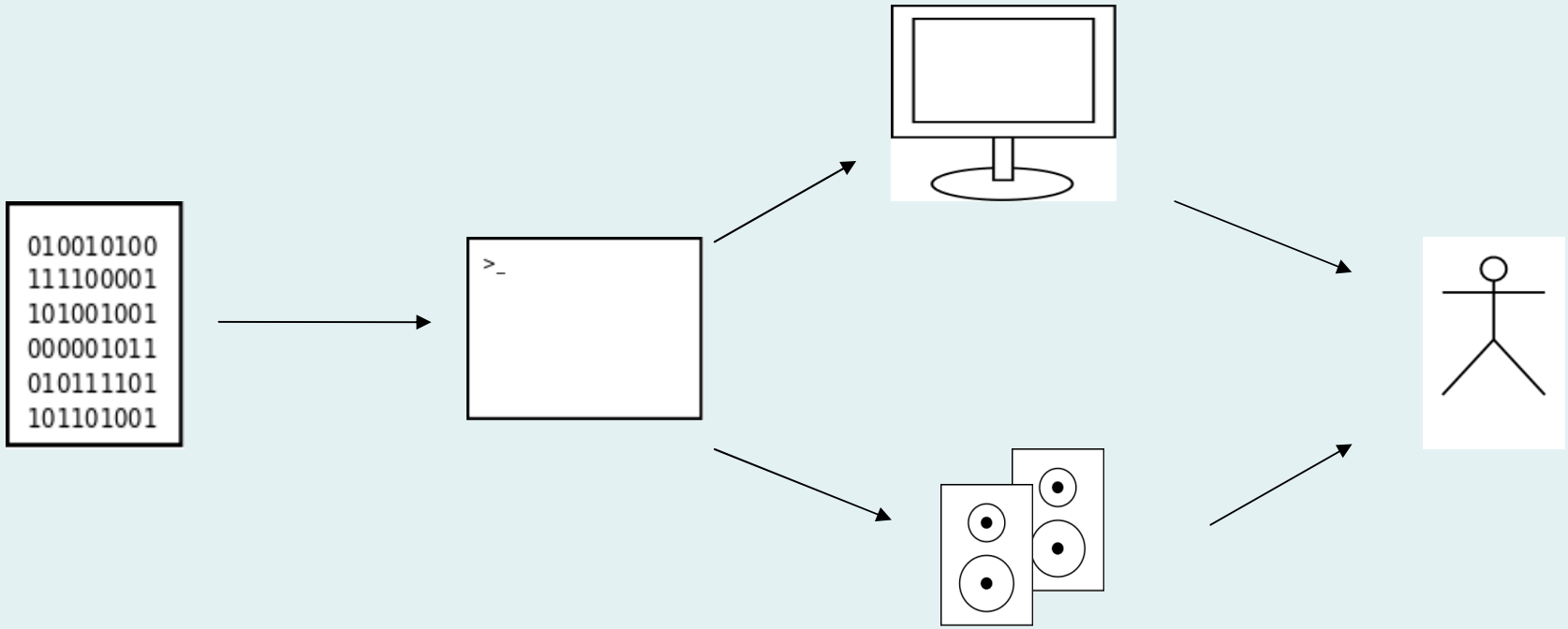
The Planets XCL Approach - Comparator



How to understand files!

What is not in a file?!

Data, Perception, Information



Data Representation

Processing

Presentation

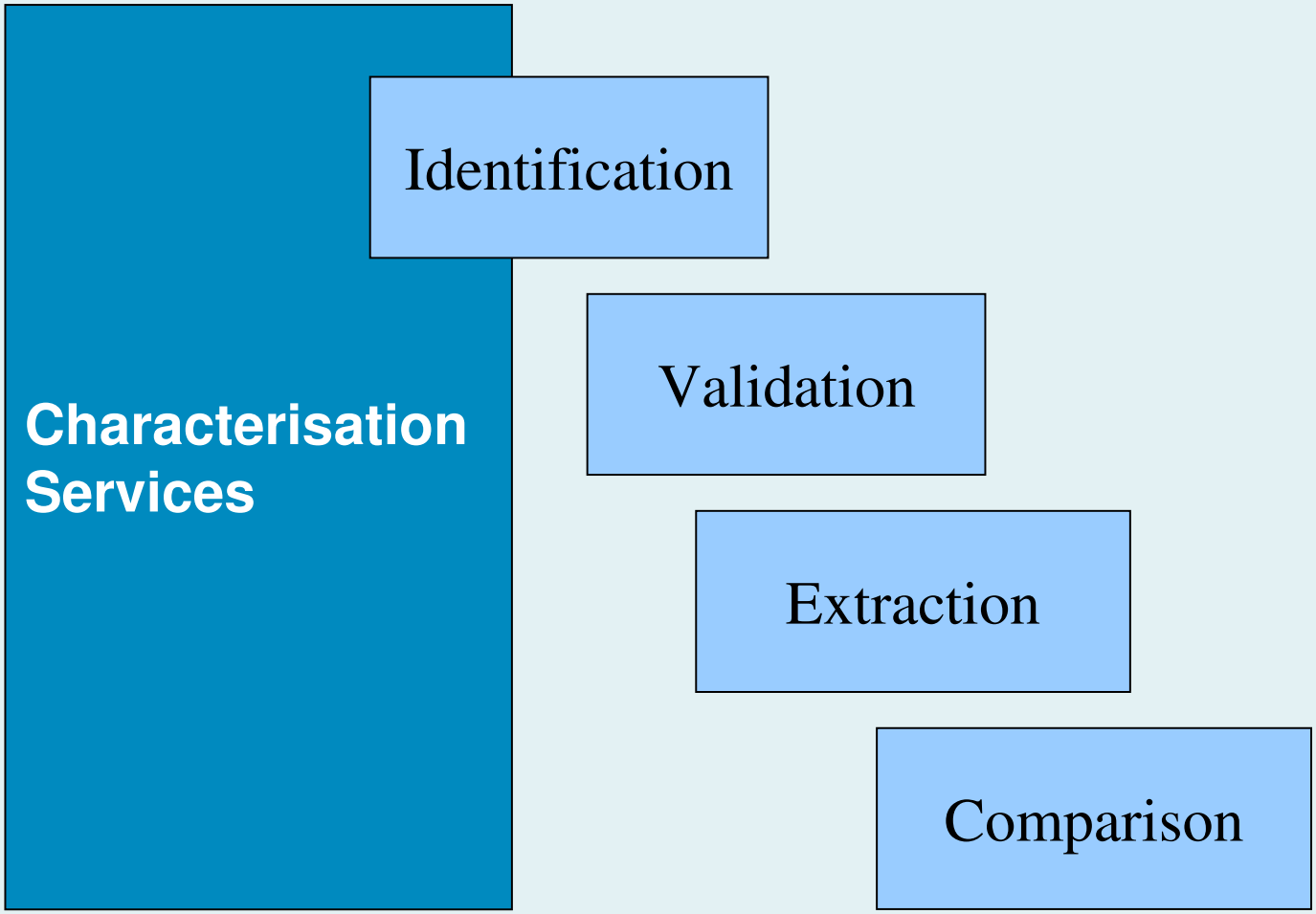
Perception

Focus

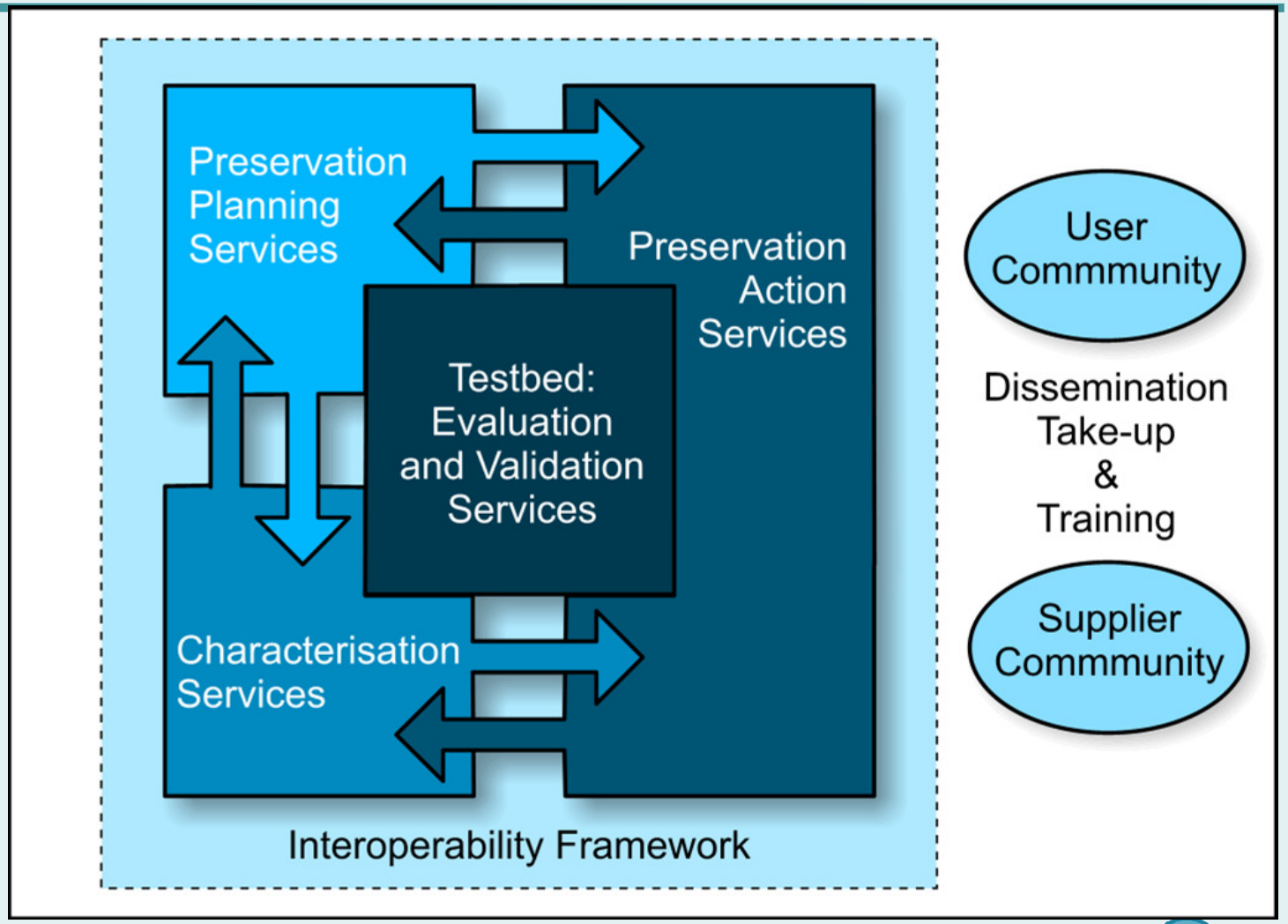
Extraction

Comparison

Zoom Out


Characterisation Services
Identification
Validation
Extraction
Comparison

Zoom Out


Preservation Planning Services
Preservation Action Services
Testbed: Evaluation and Validation Services
Characterisation Services
Interoperability Framework
User Commmunity
Dissemination
Take-up
&
Training
Supplier Commmunity

How to understand files!

The End