

**Historisch**  
**Kulturwissenschaftliche**  
**Informationsverarbeitung**

Tools: How to understand files!

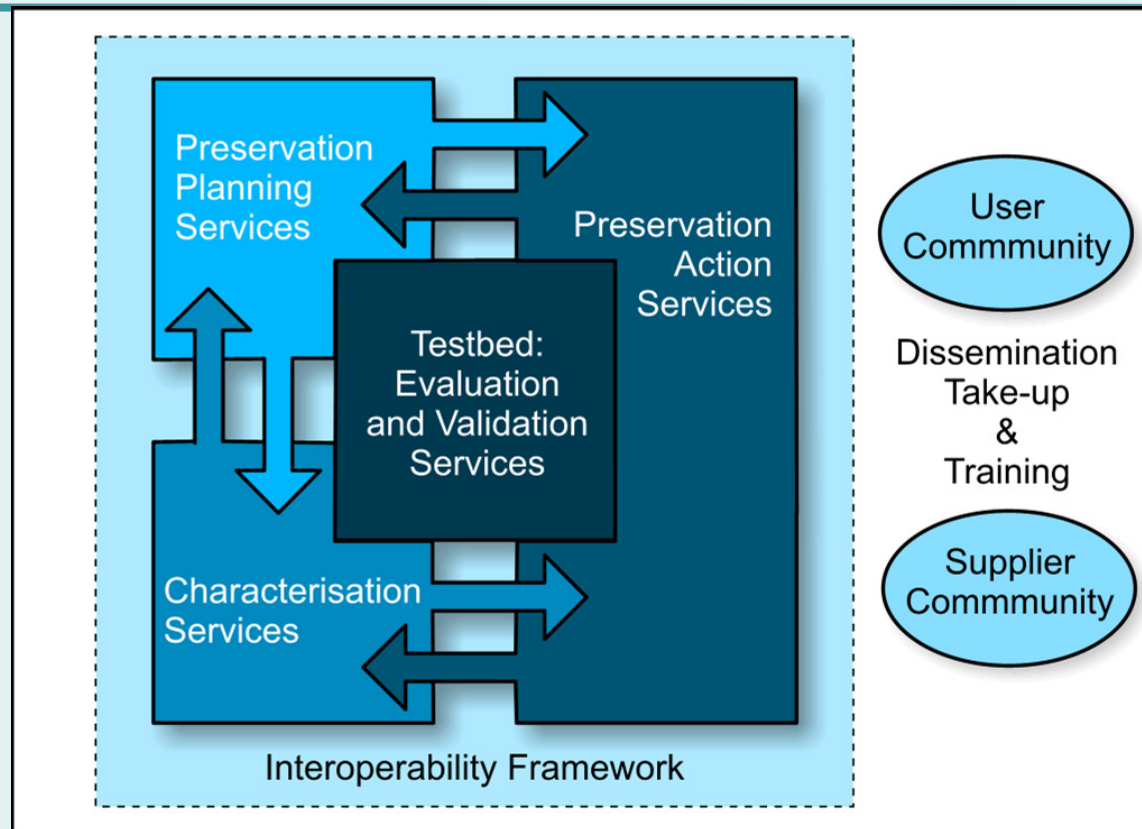
[planetarium.hki.uni-koeln.de](http://planetarium.hki.uni-koeln.de)

Jan Sch



November 17<sup>th</sup> Bern

## Zoom In



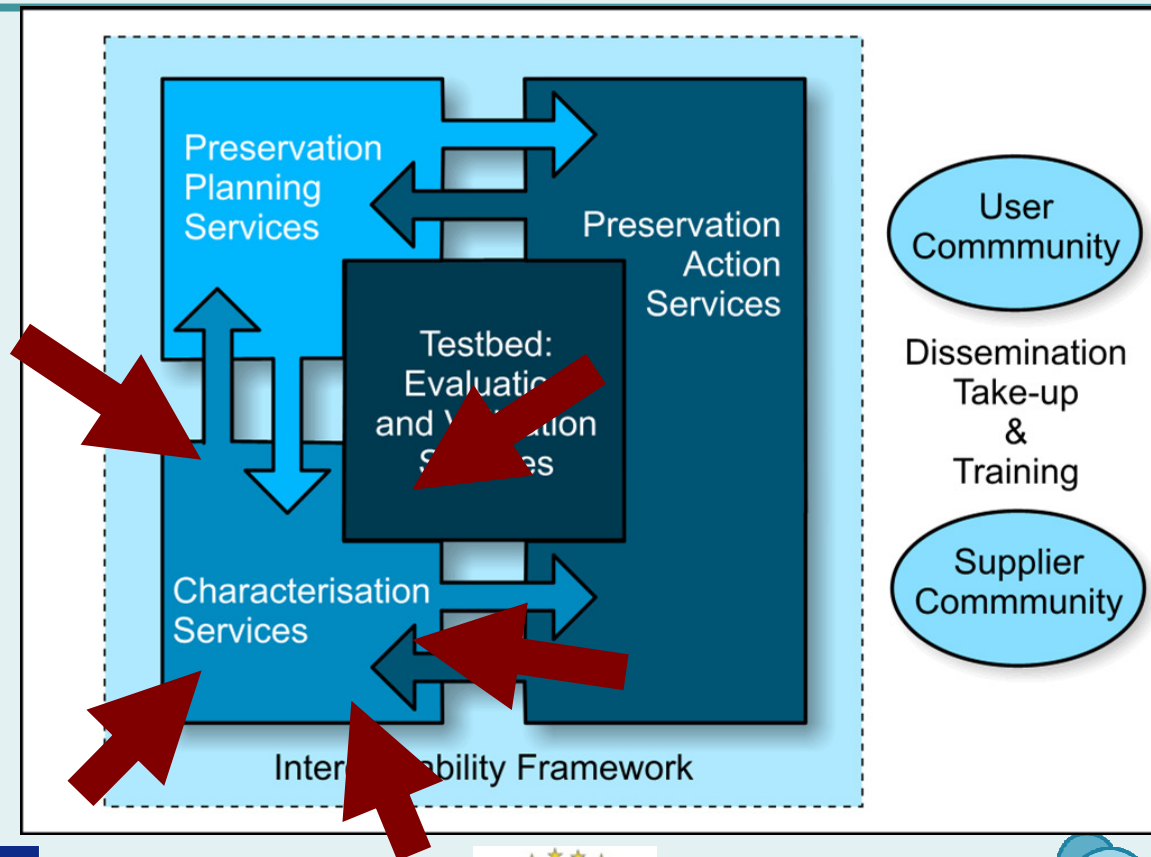
Jan Schnasse



November 17<sup>th</sup> Bern



## Zoom In

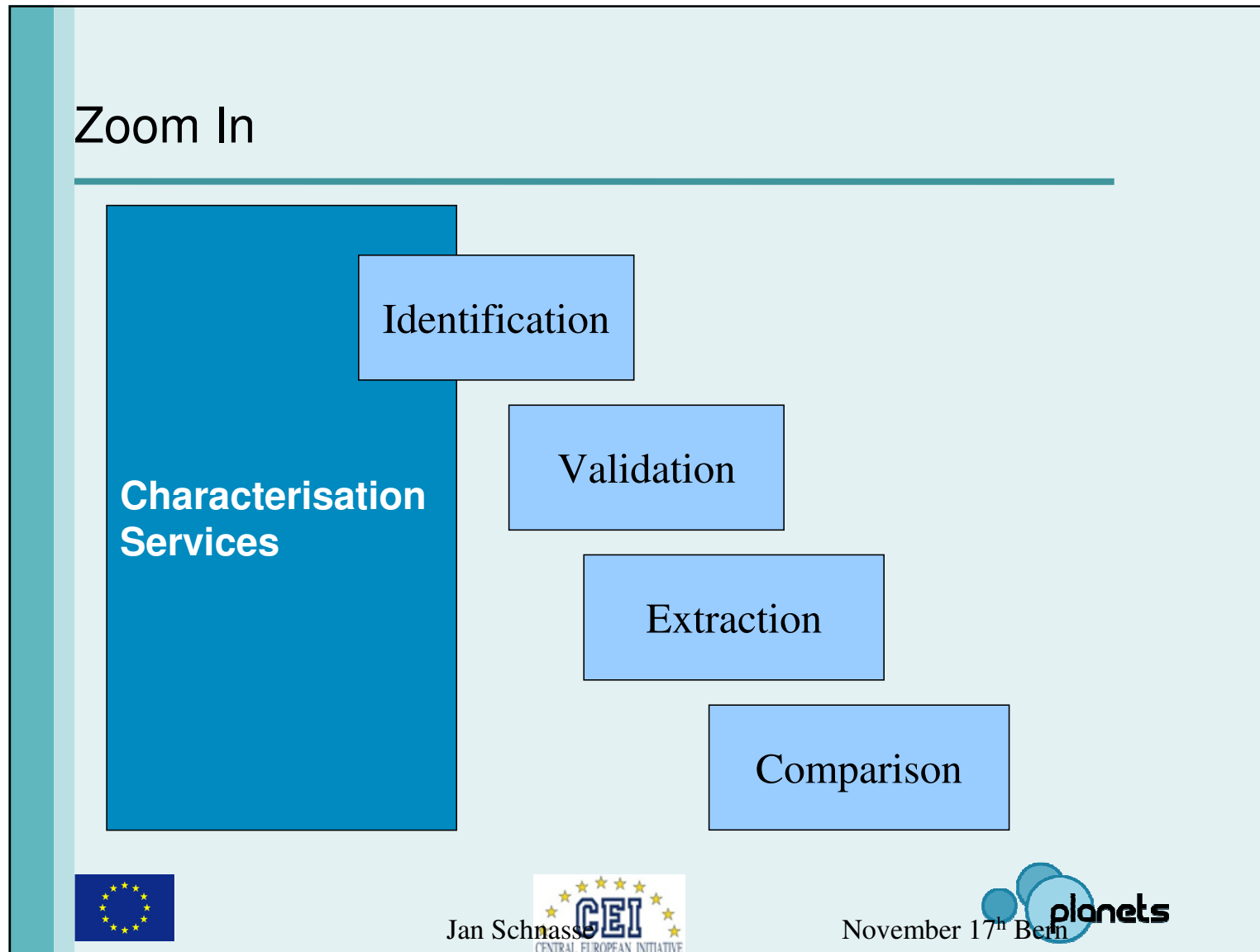


Jan Schnasse



November 17<sup>th</sup> Bern





## Focus

---

Extraction

Comparison



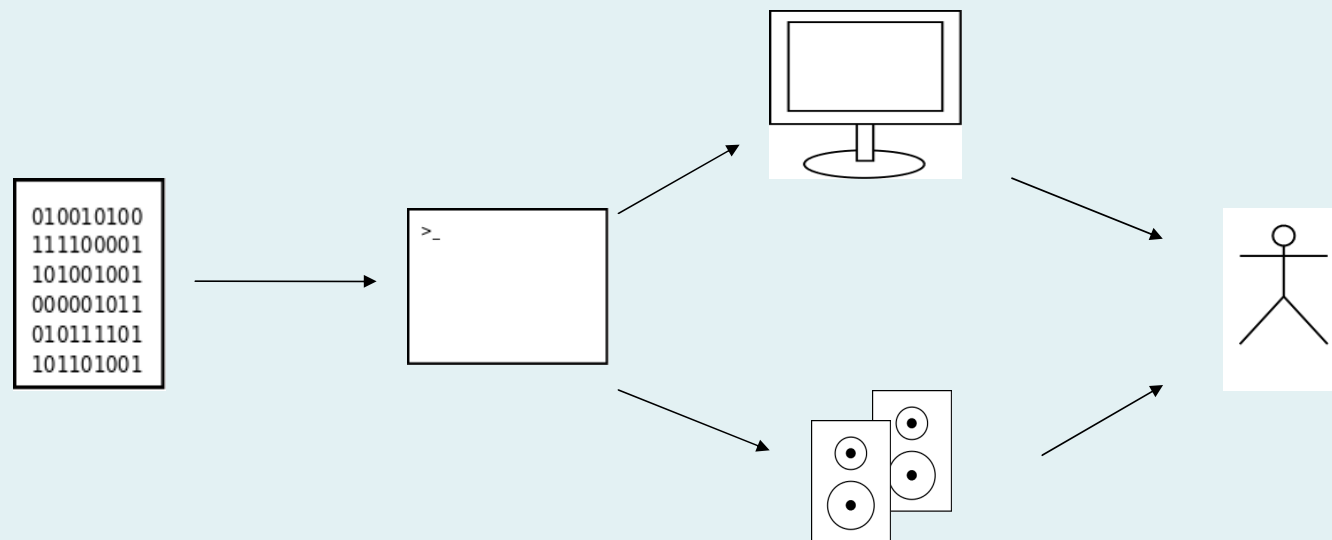
Jan Schnasse



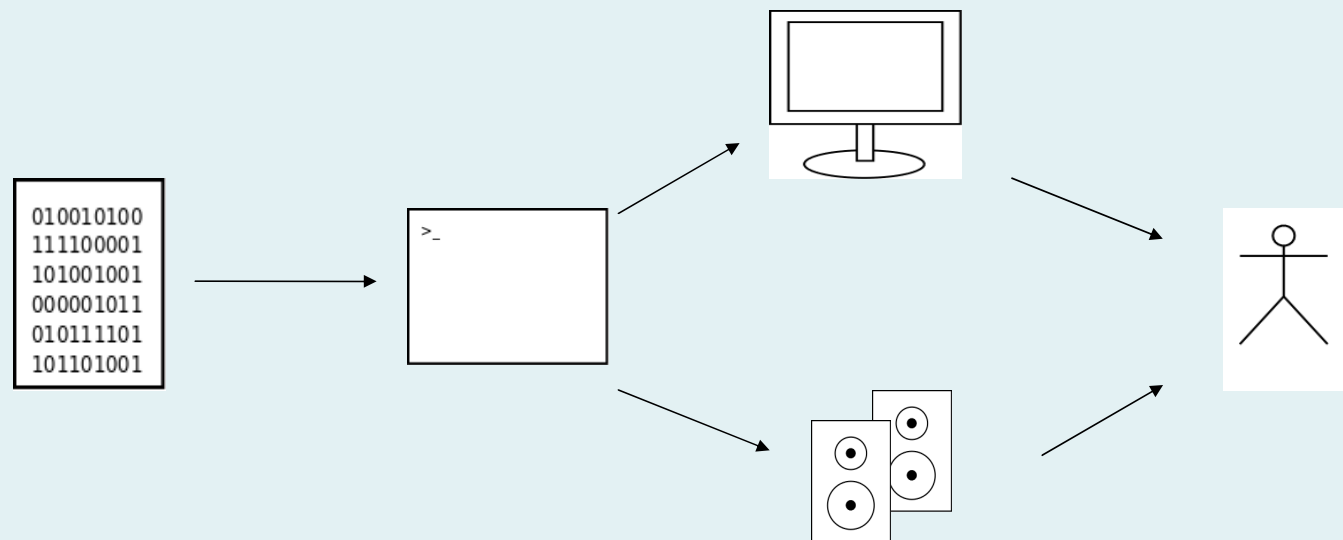
November 17<sup>th</sup> Bern



## Data, Perception, Information



## Data, Perception, Information



Data  
Representation

Processing

Presentation

Perception



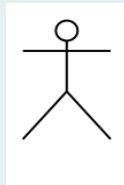
Jan Schnasse



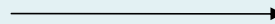
# What to preserve?

---

Perception



must be preserved



Jan Schnasse



November 17<sup>th</sup> Bern





## What to preserve?

---

What do you see?



Jan Schnasse

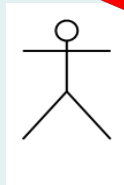


November 17<sup>th</sup> Bern

## What to preserve?

---

Perception



must be preserved

**unpreservable**



Jan Schnasse



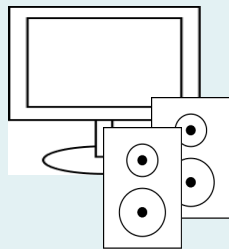
November 17<sup>th</sup> Bern



# What to preserve?

---

## Presentation



must be preserved  
→



Jan Schnasse



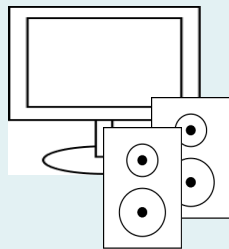
November 17<sup>th</sup> Bern



# What to preserve?

---

## Presentation



must be preserved →

adequate  
accurate  
authentic  
original  
significant



Jan Schnasse



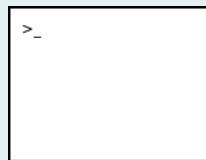
November 17<sup>th</sup> Bern



# What to preserve?

---

## Processing



must be preserved →



Jan Schnasse



November 17<sup>th</sup> Bern



# What to preserve?

---

## Processing



must be preserved  
→

sufficient



Jan Schnasse



November 17<sup>th</sup> Bern



# What to preserve?

---

## Data Representation

010010100
111100001
101001001
000001011
010111101
101101001

must be preserved →

arbitrary  
but  
complete!



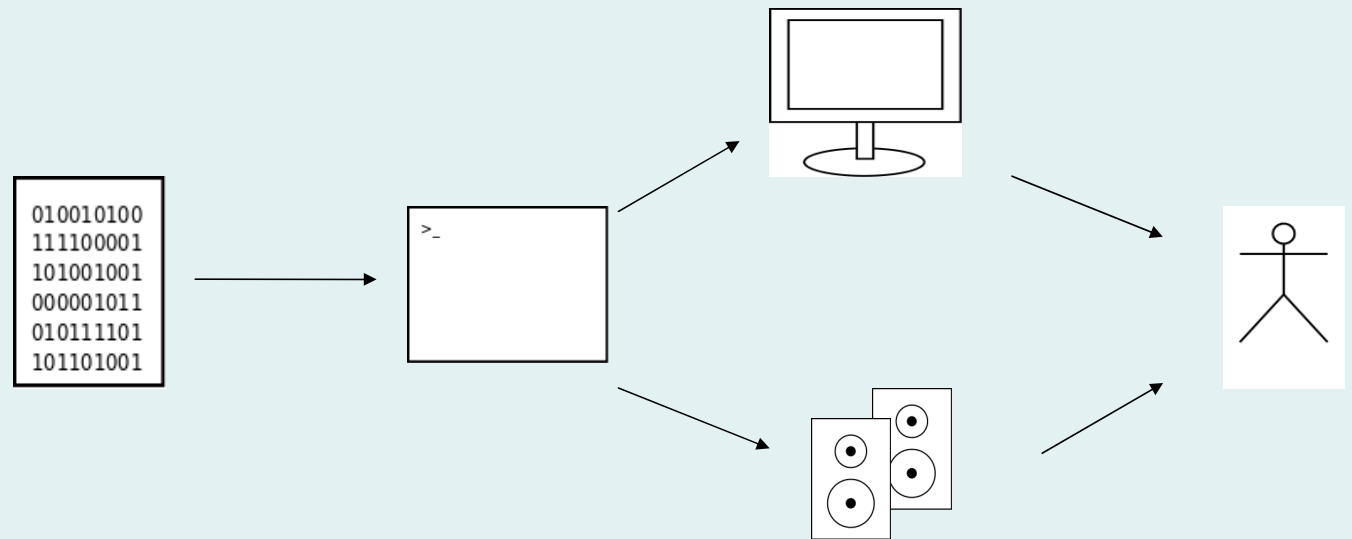
Jan Schnasse



November 17<sup>th</sup> Bern



## Data, Perception, Information



Data  
Representation

Processing

Presentation

Perception



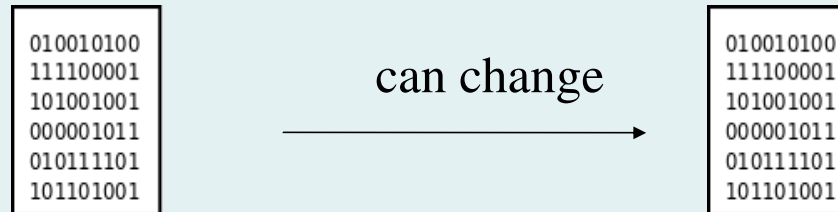
Jan Schnasse

November 17<sup>th</sup> Bern

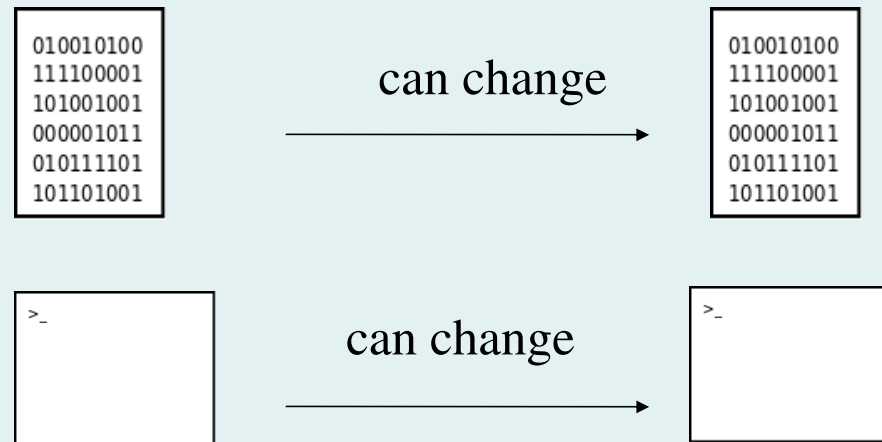




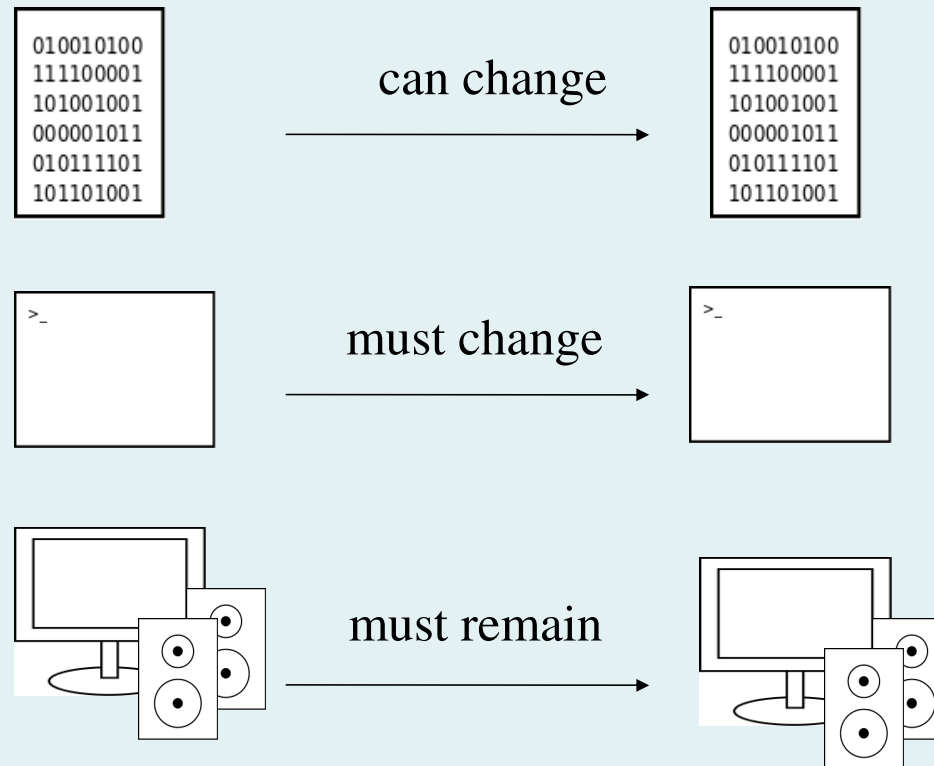
## Preservation Scenario: Format Conversion



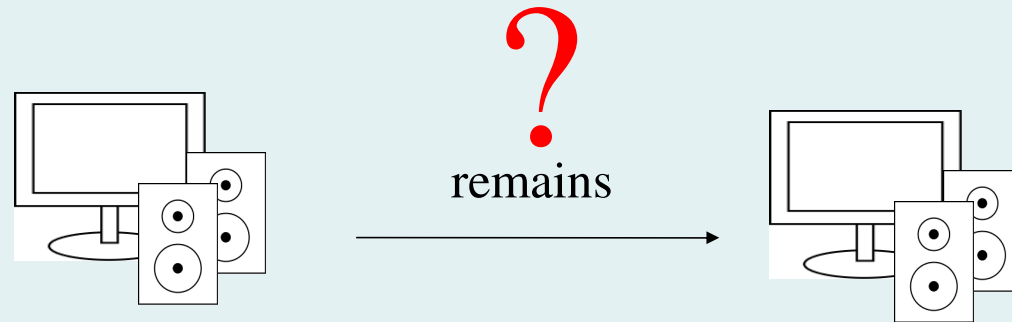
## Preservation Scenario: Format Conversion



## Preservation Scenario: Format Conversion



## Preservation Scenario: Format Conversion



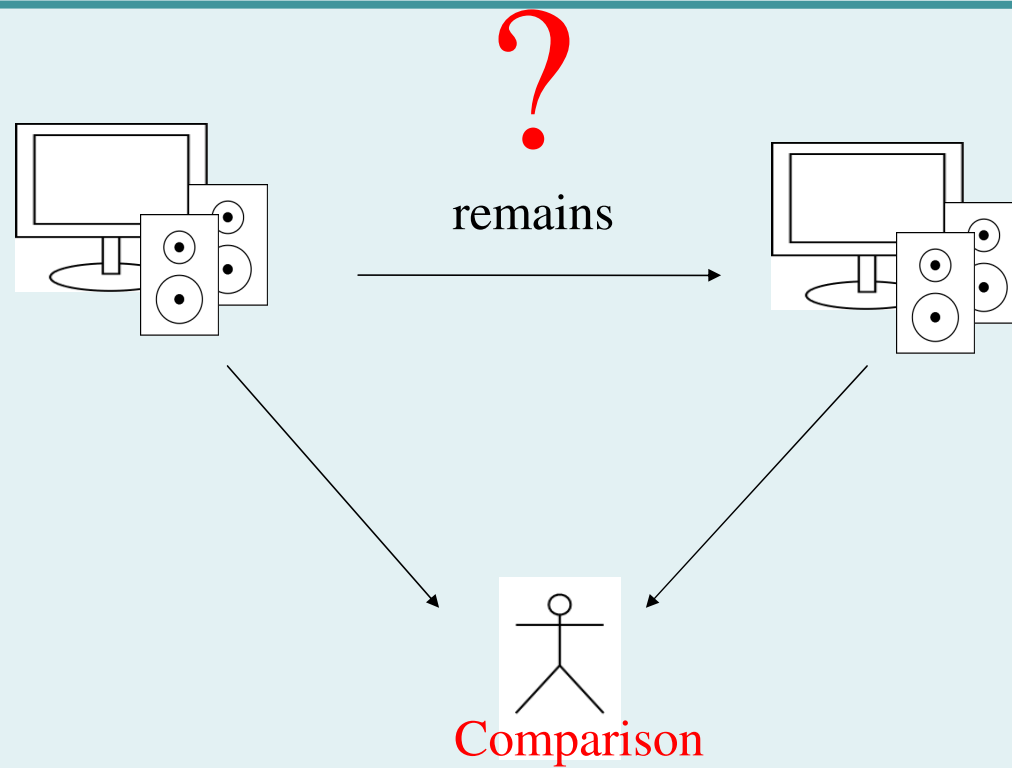
Jan Schnasse



November 17<sup>th</sup> Bern



## Evaluation of Format Conversion



Jan Schnasse



November 17<sup>th</sup> Bern



## 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

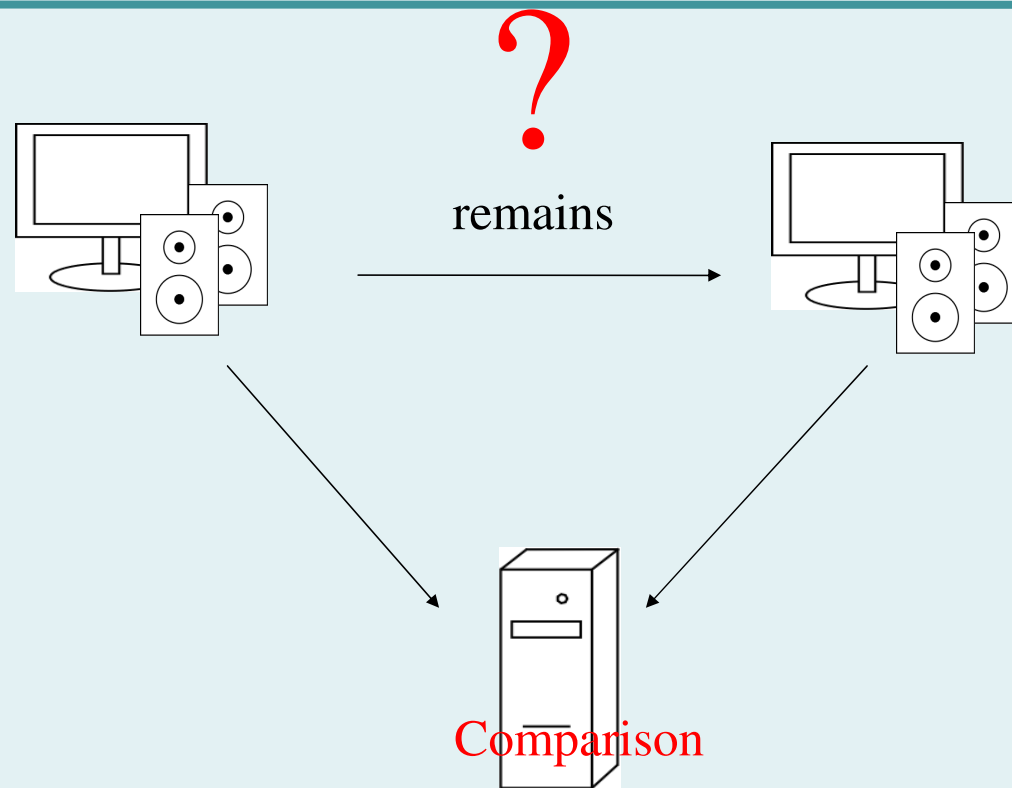
== way too much for anything



Jan Schnasse



## Evaluation of Migration



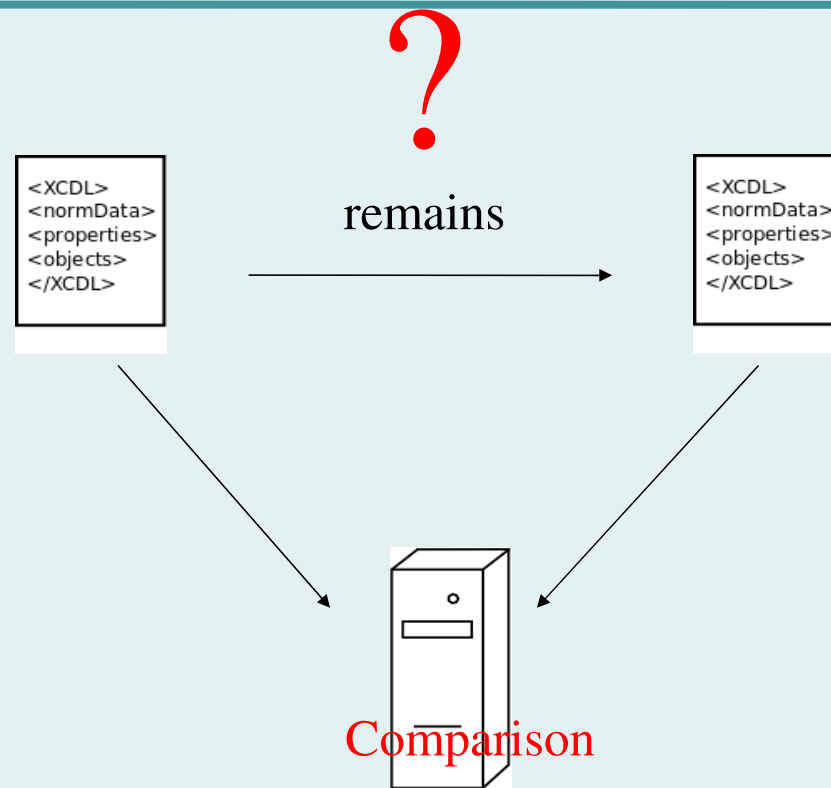
Jan Schnasse



November 17<sup>th</sup> Bern



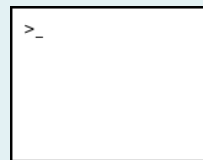
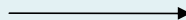
## Evaluation of Format Conversion



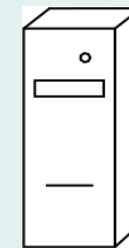
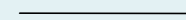


## Evaluation of Format Conversion

```
010010100
111100001
101001001
000001011
010111101
101101001
```



```
<XCDL>
<normData>
<properties>
<objects>
</XCDL>
```



Data  
Representation

Processing

Machine-readable  
Presentation

Calculation



Jan Schnasse

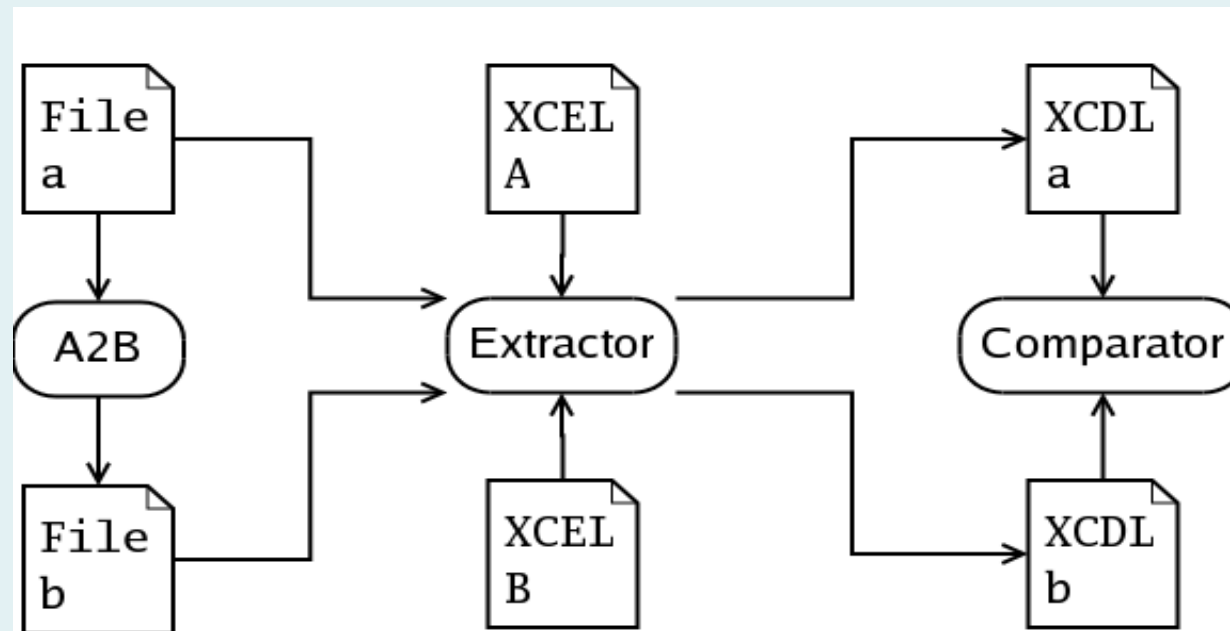


November 17<sup>th</sup> Bern



## 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/>)



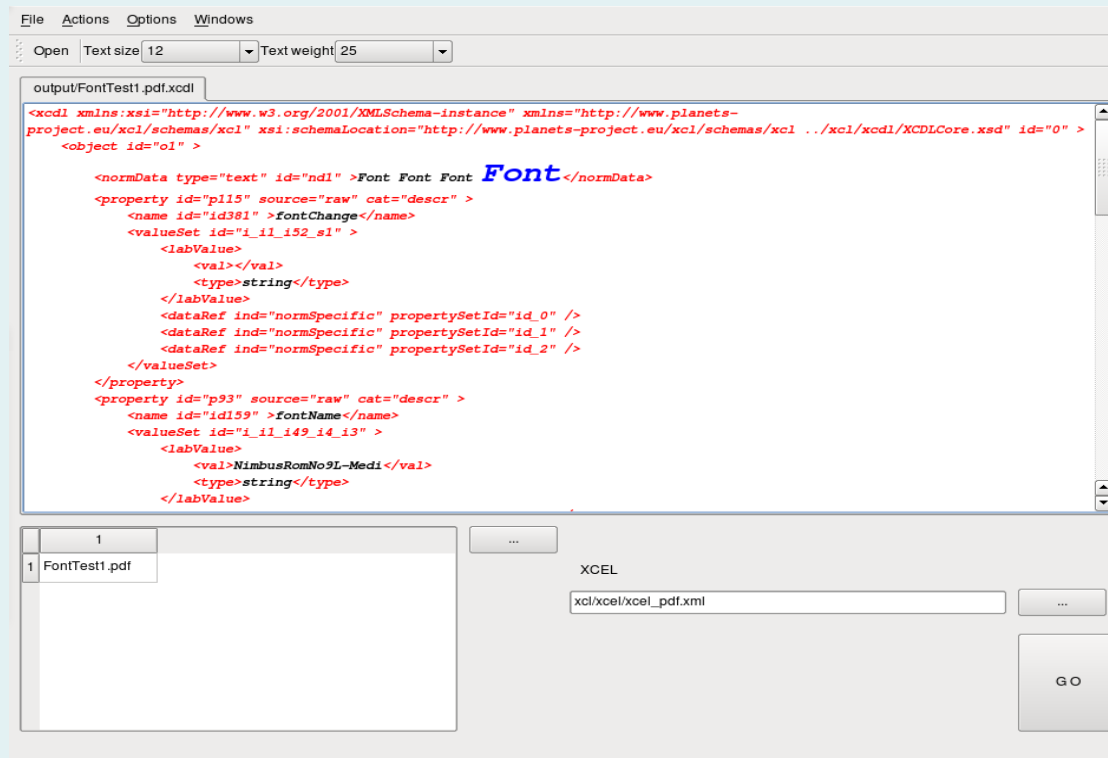
Jan Schnasse



November 17<sup>th</sup> Bern



## 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_il_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_il_s8" >
      <labValue>
        <val>32</val>
        <type>int</type>
      </labValue>
    </valueSet>
  </property> ...
</object>
```



Jan Schnasse



November 17<sup>th</sup> Bern

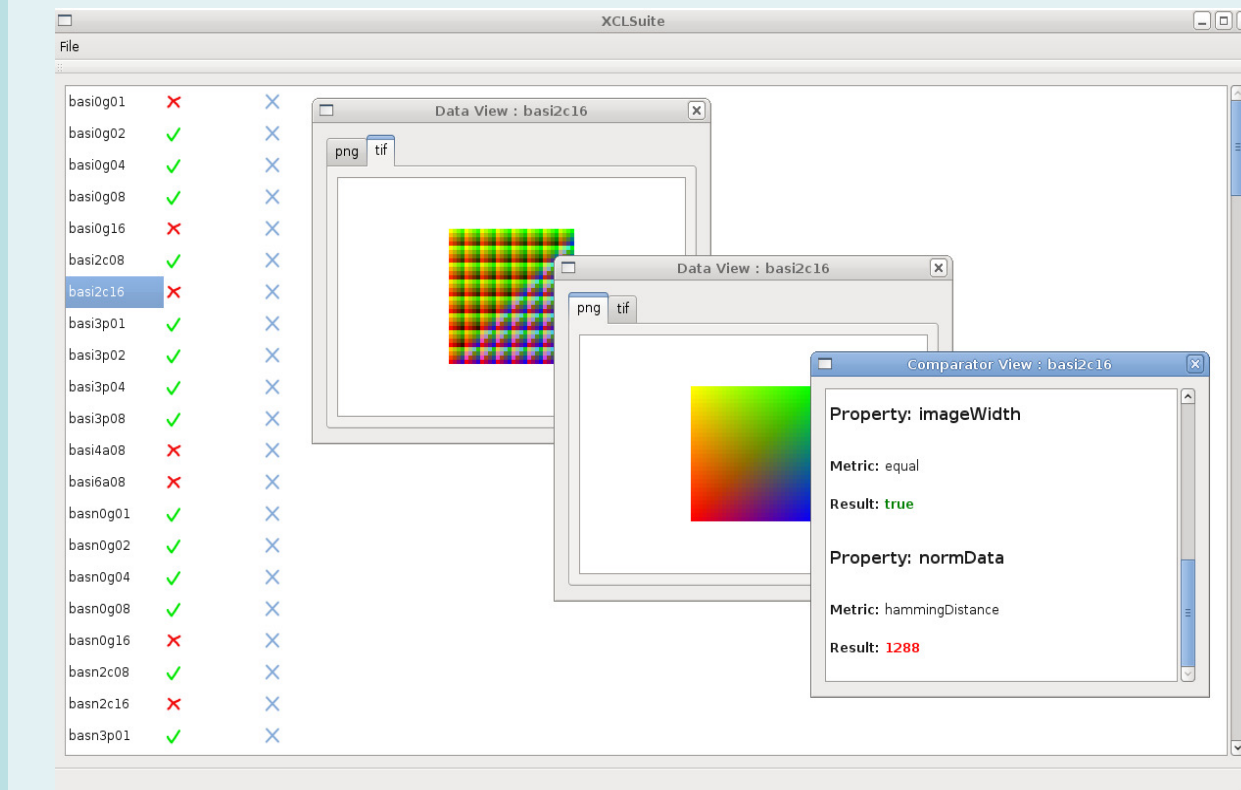


## The Planets XCL Approach – The Ontology

The screenshot displays the Planets XCL ontology editor interface. The left pane shows the 'Asserted Class Hierarchy: audiolInformation' with a tree structure including 'specificationPropertyNames' and 'XCL\_Properties'. The 'audiolInformation' class is highlighted. The middle pane shows the 'Individuals: backgroundColour\_PNG' list, with 'backgroundColour\_PNG' selected. The right pane shows the 'Individual Annotations: backgroundColour\_PNG' for the selected individual, including a comment and a datatype of 'rational'. Below the annotations, there are sections for 'Description: backgroundColour\_PNG' and 'Property assertions: backgroundColour\_PNG'.



## The Planets XCL Approach - Comparator



# How to understand files!

---

## What is not in a file?!



Jan Schnasse

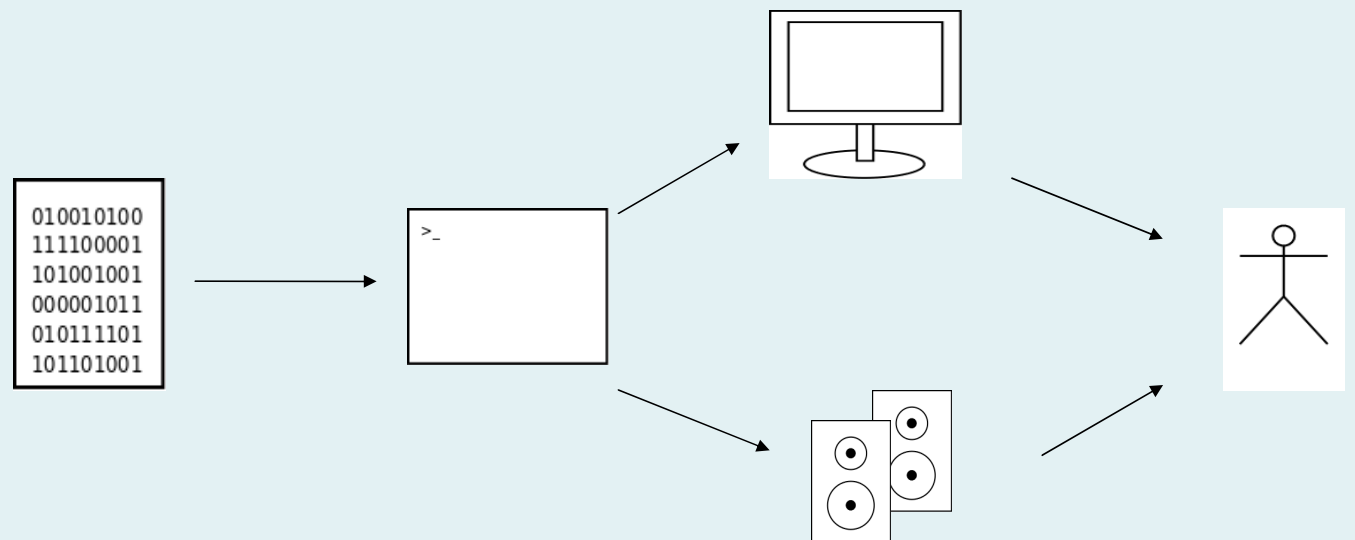


November 17<sup>th</sup> Bern





## Data, Perception, Information



Data  
Representation

Processing

Presentation

Perception



Jan Schnasse



## Focus

---

Extraction

Comparison

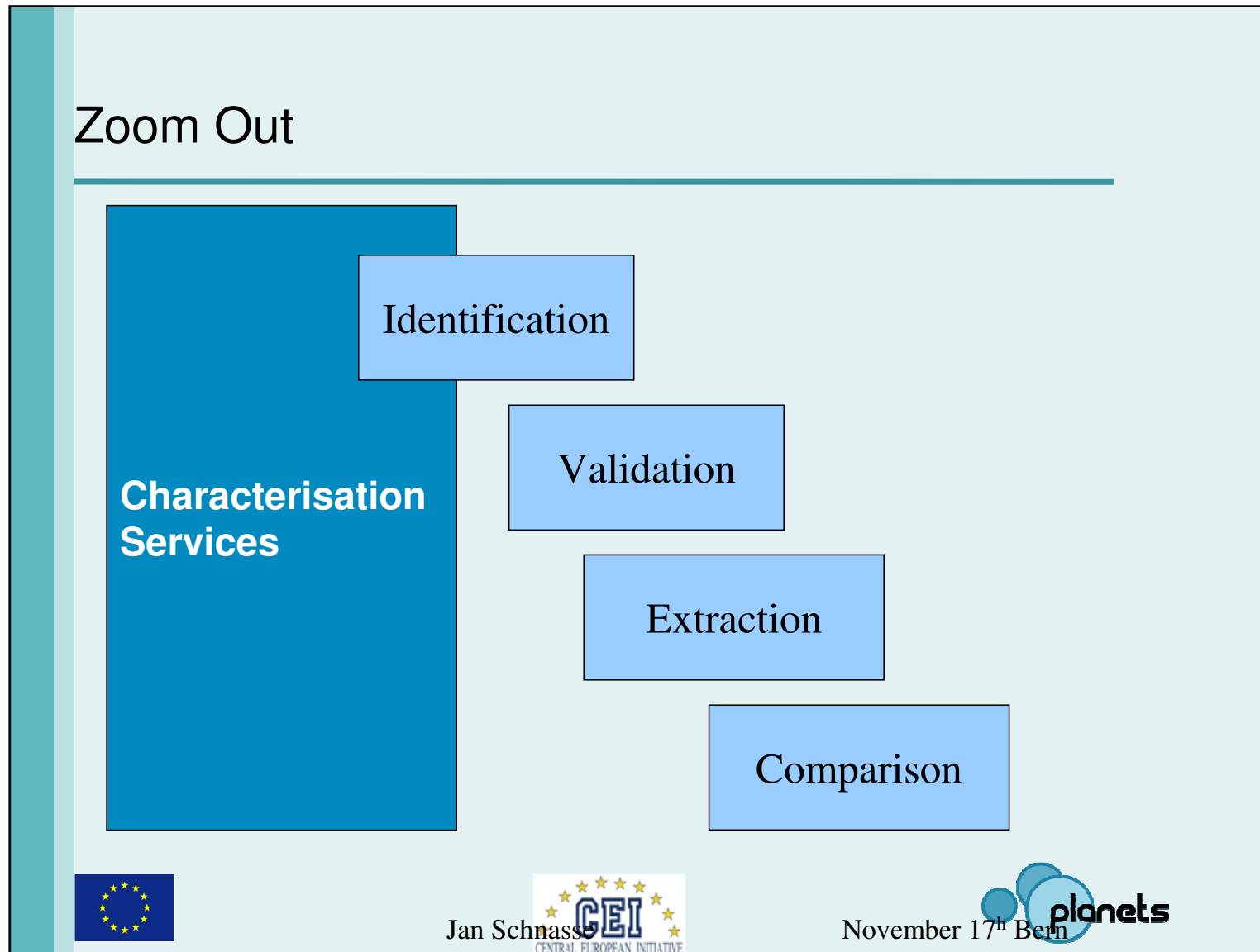


Jan Schnasse

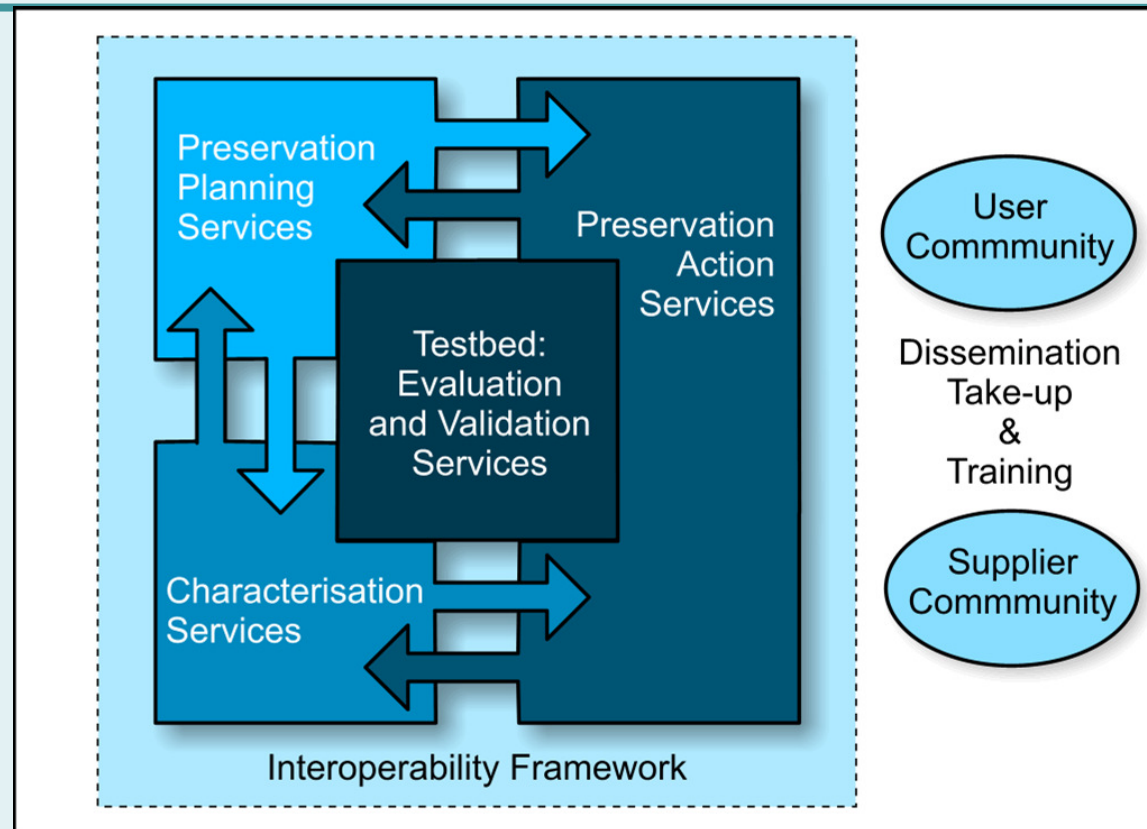


November 17<sup>th</sup> Bern





## Zoom Out



Jan Schnasse



November 17<sup>th</sup> Bern



## Further Information

---

[planetarium.hki.uni-koeln.de](http://planetarium.hki.uni-koeln.de)



Jan Schnasse



November 17<sup>th</sup> Bern



# How to understand files!

---

# The End



Jan Schnasse



November 17<sup>th</sup> Bern

