XML support in the libraries
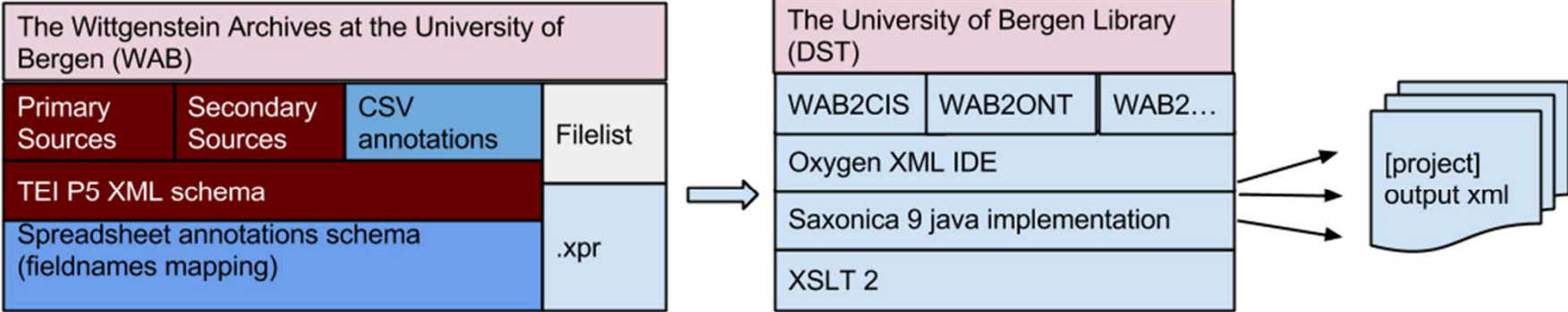
Øyvind Liland Gjesdal

Digital Systems and Services

University of Bergen Library

Digital Systems and Services

- Support development tasks for our University partners.
  - Text partners (XML coding)
    - **Wittgenstein Archives**
    - Ludvig Holberg
    - (The Library special collections)
  - … Institutional repositories (BORA)
  - … Journals (BOAP)

Different versions and output from WAB

Source files

- TEI P5 primary and secondary sources.
- Transcriptions with sections marked by identifiers
- Excel spreadsheets with annotations
  - hasVersion
  - Discusses
  - Points to the identifiers in the TEI P5 documents

Different versions and output from WAB

- **WAB2CIS**
- WAB2Ontology (agora)
- WAB2DM2E
- WAB2overview

XSLT

- eXtensible Stylesheet Language Transformation.
- «XSLT is a language for transforming the structure and content of an XML document» (Kay, 2008)
- Input file (optional) -> xslt -> output file(s)
- XSLT is both xslt and xpath.
- oXygen XML IDE

WAB2CIS

- Cooperative project with CIS-München.
  - We deliver four versions of WAB to München
    - CIS –normalized xml
    - CIS –diplomatic xml
    - CIS – text only (with formatting/spacing)
    - CIS – text-siglum
    - Specifications by Max Hadersbeck, Daniel Bruder and Alois Pichler.

Example: rules in CISNORM and DIPL

```xsl
 <xsl:template match="*:corr[starts-with(@type,'npc')]"
priority="0.8">
    <xsl:choose>
        <xsl:when test="$show='norm'"></xsl:when>
        <xsl:when test="$show='dipl'">
            <xsl:apply-templates/>
        </xsl:when>
    </xsl:choose>
 </xsl:template>
```

- Content in blue is an xpath expression to match.

- Content in red is a standard «switch» construct in XSLT.

Example input string

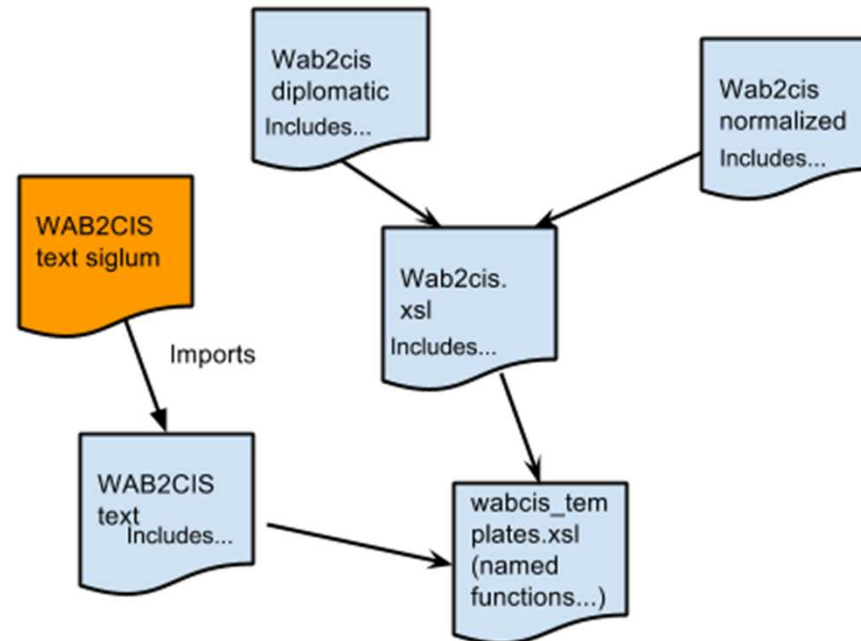die Definition immer auch el<corr type="npcn">l</corr>iminieren

Output normalized: die Definition immer auch eliminieren

Output diplomatic: die Definition immer auch elliminieren

(From: Ms-114,1v[1]et2r[1]et2v[1]et3r[1]et3v[1]et4r[1] sentence 35)

WAB2CIS/XSLT

- Main stylesheets to transform
- Include, import…

Original TEI file from WAB

```xml
<reloc type="relocate-nec" n="Ts-310,3a_Ts-310,3" corresp="Ts-310#1" status="unremarkable"> <emph rend="indl_5"/>
<emph rend="slilm_h"><seg type="edinst" part="N"><emph rend="cap">Insert</emph>:</seg><emph rend="blankspace_4"/>
<c type="c" part="N">T</c>he connection between these two ideas is that the<lb/> mental
experiences which accompany the use of a sign undoubt<lb rend="shyphen"/>edly are caused
by our usage of the sign in a particular<lb/> system of
language.</emph> </reloc>


 <pb facs="Ts-310_4" n="pagename_Ts-310,4 pageref_Ts-310,11"/>
<fw type="pagen" place="top right"><del type="dnpc_h" status="unremarkable">3</del><add rend="el_h" status="unrem

words by comparing the
sensations which they produce.</s></rs>
<s type="es" part="N">Ask<lb/> yourself what means we have of finding out the feelings which<lb/>

they produce in different people and on different occasions.</s> <lb/>
<s type="es" part="N">Ask yourself: "<c type="c" part="N">W</c>hen I said, '<c type="c" part="N">G</c>ive me
```

WAB2CIS Normalized output

```
    <space, quantity="4" unit="chars"/>, The connection between these two ideas is that the<lb/> mental experiences which acc
<s n="Ts-310,2[2]et3[1]et3a[1]et4[1]_24"
    ana="facs:Ts-310,2 abnr:3 satznr:45">Ask<lb/> yourself what means we have of finding out the feelings which<lb/> they p
<lb/>
<s n="Ts-310,2[2]et3[1]et3a[1]et4[1]_25"
    ana="facs:Ts-310,2 abnr:3 satznr:46">Ask yourself: "When I said, 'Give me an apple <emph rend="us1">&amp;</emph> a pear,
    <lb/> leave the room', had I the same feeling when I pronounced the<lb/> two words '&amp;'?"</s>
<s n="Ts-310,2[2]et3[1]et3a[1]et4[1]_26"
    ana="facs:Ts-310,2 abnr:3 satznr:47">But we do not deny that the people who use the<lb/> word "but" as "not" is used in H
<s n="Ts-310,2[2]et3[1]et3a[1]et4[1]_27"
    ana="facs:Ts-310,2 abnr:3 satznr:48">And the word "but" in the two<lb/> languages will on the whole be accompanied by di
<lb rend="hl"/>
<s n="Ts-310,2[2]et3[1]et3a[1]et4[1]_28"
    ana="facs:Ts-310,2 abnr:3 satznr:49">
    <seg type="series-number" part="N">2).<space quantity="2" unit="chars"/>
    </seg> Let us now look at an extension of language 1).</s>
```

WAB2CIS text output

```
Ts-310_3a



                        The connection between these two ideas is that the
 mental experiences which accompany the use of a sign undoubt-
edly are caused by our usage of the sign in a particular
 system of language.
Ts-310_4

   words by comparing the sensations which they produce. Ask
 yourself what means we have of finding out the feelings which
 they produce in different people and on different occasions.
Ask yourself: "When I said, 'Give me an apple & a pear, &
 leave the room', had I the same feeling when I pronounced the
 two words '&'?" But we do not deny that the people who use the
 word "but" as "not" is used in English will broadly speaking
 have similar sensations accompanying the word "but" as the Eng-
lish have when they use "not". And the word "but" in the two
 languages will on the whole be accompanied by different sets of
 experiences.)

                      2).
                        Let us now look at an extension of language 1). The
```

WAB2CIS text-siglum output

```
Ts-310,1[2]et2[1];Suppose a man described a game of chess, without mentioning the existence and operations of the
Ts-310,2[2]et3[1]et3a[1]et4[1];(Note: Objection: The word "brick" in language 1) has not the meaning which it has
Ts-310,4[2]et5[1];(Remark: We stressed the importance of learning the series of numerals by heart because there wa
Ts-310,5[2];What have the demonstrative explanations of the numerals in common with those of the words "slab", "co
Ts-310,5[3]et6[1];(Remark: This difference does not lie, however, in the act of pointing and pronouncing the word
Ts-310,6[2]et7[1];(Remark: It has been suggested that such words as "there", "here", "now", "this" are the "real 1
Ts-310,7[2];— "What does the word 'exactness' mean? Is it real exactness if you are supposed to come to tea at 4.
Ts-310,7[3]et8[1];Systems of communication as for instance 1), 2), 3), 4), 5) we shall call "language-games". They
Ts-310,8[2];When the boy or grown-up learns what one might call special technical languages, e.g., the use of char
Ts-310,8[3]et9[1];(Remark: Our use of expressions like "names of numbers", "names of colours", "names of materials
Ts-310,9[2]et10[1];Let us now look at the different kinds of signs which we have introduced. First let us distingu
Ts-310,10[2]et11[1];Reflections such as the preceding will show us the infinite variety of the functions of words
Ts-310,11[2];(Remark: Think of the reasons for and against classifying "O" with the other cardinals. "Are black a1
Ts-310,11[3]et12[1]et13[1]et14[1];It is natural for us to call gestures, as those employed in 4), or pictures as :
Ts-310,14[2]et15[1]et16[1];When we call these three preceding cases cases of comparing from memory, we feel that 1
Ts-310,16[2]et17[1];"But surely in case 14c) B acted entirely automatically. If all that happened was really what
Ts-310,17[2]et18[1];If our case 14c) troubles you, you may be inclined to say: "But why did he bring just this bo:
Ts-310,18[2]et19[1];"But then B didn't really recognize the material as the right one". — You needn't reckon 14c)
```

Fragen? ☺