



Università
Ca' Foscari
Venezia



UNIVERSITÀ
DEGLI STUDI
FIRENZE



Istituto di Linguistica
Computazionale
"Antonio Zampolli"

 Consiglio Nazionale delle Ricerche

ItAnt DSL

Luca Rigobianco – Federico Boschetti

Venezia, 20 settembre 2024 - PRIN 2017 Lingue e culture dell'Italia antica: linguistica storica e modelli digitali



Defining DSL

- Domain-Specific Languages (DSLs) are programming or markup languages created specifically for a certain field, providing more expressive power, simplicity, and efficiency.



Disadvantages of XML-TEI/EpiDoc encoding

- Encoding epigraphic contextual metadata and textual data in XML-TEI/EpiDoc is a complex, error-prone task
- XML-TEI is quite verbose and redundant

```
<tei:msIdentifier>
```

```
<tei:settlement ref="https://sws.geonames.org/3166930">Schiavi d'Abruzzo (Chieti)</tei:settlement>
```

```
<tei:institution ana="https://sabapchpe.beniculturali.it/soprintendenza/aree-tematiche/patrimonio-archeologico/schiavi-dabruzzo/">in situ (under the tutelage of the Soprintendenza Archeologia, Belle Arti e Paesaggio dell'Abruzzo)</tei:institution>
```

```
<tei:msName>Mosaic from the sanctuary of Schiavi d'Abruzzo</tei:msName>
```

```
<tei:altIdentifier type="trismegistos"> <tei:idno source="www.trismegistos.org/text/170843">TM 170843</tei:idno> </tei:altIdentifier>
```

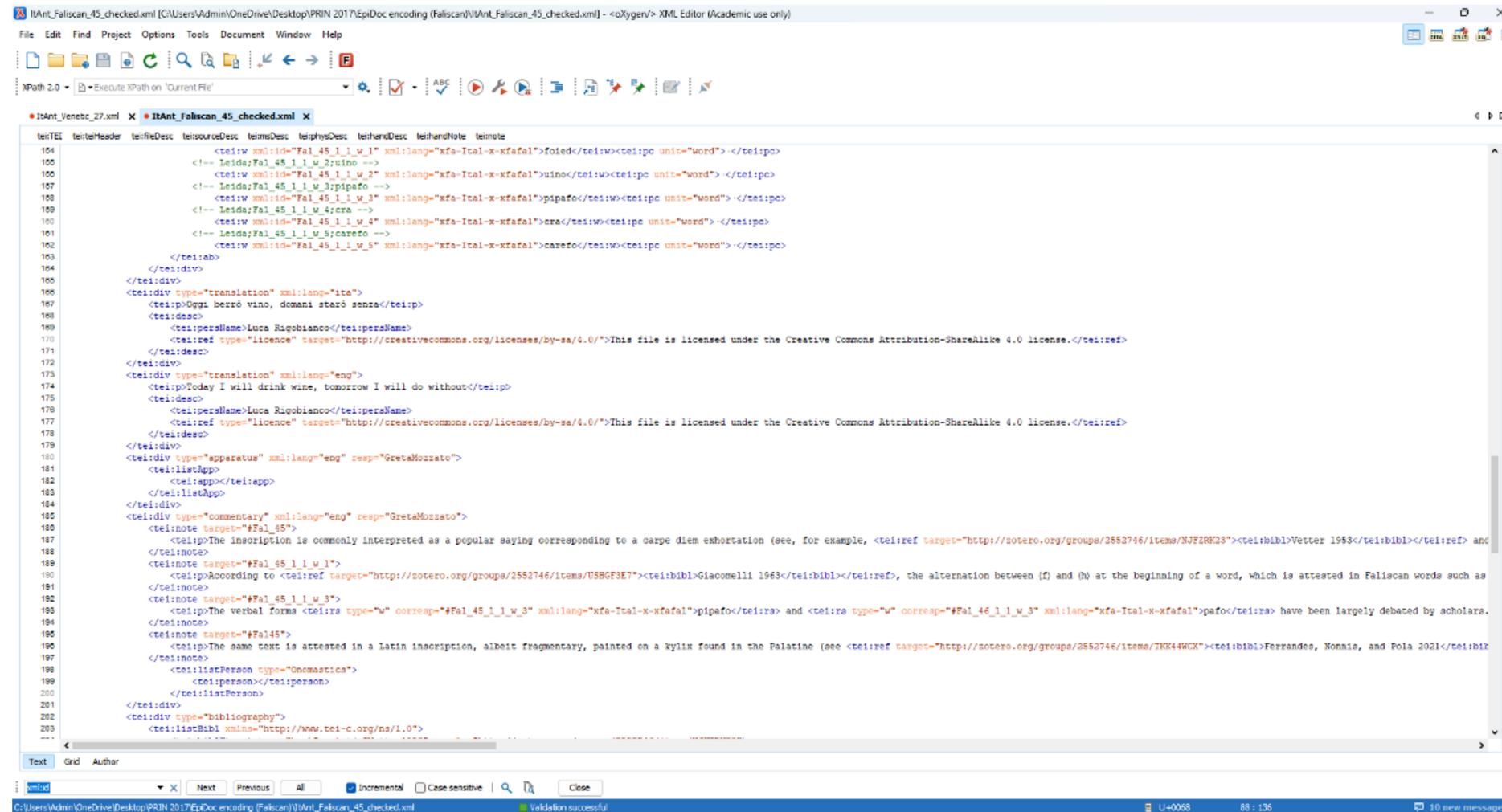
```
<tei:altIdentifier type="traditionalID"> <tei:idno source="http://zotero.org/groups/2552746/items/IMWMSAVE">ST Sa 2</tei:idno> </tei:altIdentifier>
```

```
<tei:altIdentifier type="traditionalID"> <tei:idno source="http://zotero.org/groups/2552746/items/IKUTPHDC">ImIt Teruentum 36</tei:idno>  
</tei:altIdentifier>
```

```
</tei:msIdentifier>
```



Disadvantages of XML-TEI/EpiDoc encoding



```
ItAnt_Venetic_27.xml x ItAnt_Faliscan_45_checked.xml x
File Edit Find Project Options Tools Document Window Help
XPath 2.0 - Execute XPath on 'Current File'
ItAnt_Venetic_27.xml x ItAnt_Faliscan_45_checked.xml x
teiTEI teiHeader teiText
154 <teiw xml:id="Fal_45_1_1_w_1" xml:lang="xfa-Ital-x-xfafal">foied</teiw><teip unit="word"></teip>
155 <!-- Leida:Fal_45_1_1_w_2:uino -->
156 <teiw xml:id="Fal_45_1_1_w_2" xml:lang="xfa-Ital-x-xfafal">uino</teiw><teip unit="word"></teip>
157 <!-- Leida:Fal_45_1_1_w_3:pifafo -->
158 <teiw xml:id="Fal_45_1_1_w_3" xml:lang="xfa-Ital-x-xfafal">pifafo</teiw><teip unit="word"></teip>
159 <!-- Leida:Fal_45_1_1_w_4:cre -->
160 <teiw xml:id="Fal_45_1_1_w_4" xml:lang="xfa-Ital-x-xfafal">cre</teiw><teip unit="word"></teip>
161 <!-- Leida:Fal_45_1_1_w_5:carefo -->
162 <teiw xml:id="Fal_45_1_1_w_5" xml:lang="xfa-Ital-x-xfafal">carefo</teiw><teip unit="word"></teip>
163 </teitab>
164 </teidiv>
165 </teiText>
166 <teidiv type="translation" xml:lang="ita">
167 <teip>Oggi berrò vino, domani starò senza</teip>
168 <teidesc>
169 <tei:persName>Luca Rigobianco</tei:persName>
170 <tei:ref type="licence" target="http://creativecommons.org/licenses/by-sa/4.0/">This file is licensed under the Creative Commons Attribution-ShareAlike 4.0 license.</tei:ref>
171 </teidesc>
172 </teidiv>
173 <teidiv type="translation" xml:lang="eng">
174 <teip>Today I will drink wine, tomorrow I will do without</teip>
175 <teidesc>
176 <tei:persName>Luca Rigobianco</tei:persName>
177 <tei:ref type="licence" target="http://creativecommons.org/licenses/by-sa/4.0/">This file is licensed under the Creative Commons Attribution-ShareAlike 4.0 license.</tei:ref>
178 </teidesc>
179 </teidiv>
180 <teidiv type="apparatus" xml:lang="eng" resp="GretaMozzato">
181 <teitab>
182 <teitab></teitab>
183 </teitab>
184 </teidiv>
185 <teidiv type="commentary" xml:lang="eng" resp="GretaMozzato">
186 <teinote target="#Fal_45">
187 <teip>The inscription is commonly interpreted as a popular saying corresponding to a carpe diem exhortation (see, for example, <tei:ref target="http://zotero.org/groups/2552746/items/NJF2R23"><tei:bibl>Vetter 1953</tei:bibl></tei:ref> and
188 </teinote>
189 <teinote target="#Fal_45_1_1_w_1">
190 <teip>According to <tei:ref target="http://zotero.org/groups/2552746/items/US8GF3E7"><tei:bibl>Giacomelli 1963</tei:bibl></tei:ref>, the alternation between (f) and (h) at the beginning of a word, which is attested in Faliscan words such as
191 </teinote>
192 <teinote target="#Fal_45_1_1_w_3">
193 <teip>The verbal forms <teirs type="w" corresp="#Fal_45_1_1_w_3" xml:lang="xfa-Ital-x-xfafal">pifafo</teirs> and <teirs type="w" corresp="#Fal_46_1_1_w_3" xml:lang="xfa-Ital-x-xfafal">pifo</teirs> have been largely debated by scholars.
194 </teinote>
195 <teinote target="#Fal45">
196 <teip>The same text is attested in a Latin inscription, albeit fragmentary, painted on a kyllix found in the Palatine (see <tei:ref target="http://zotero.org/groups/2552746/items/TKK4WCK"><tei:bibl>Ferrandes, Nonnis, and Pola 2021</tei:bibl>
197 </teinote>
198 <teitab>
199 <teitab></teitab>
200 </teitab>
201 </teidiv>
202 <teidiv type="bibliography">
203 <teitab>
204 <teitab></teitab>
205 </teitab>
206 </teidiv>
207 </teiText>
208 </teiTEI>
Text Grid Author
xml:lang Next Previous All Incremental Case sensitive Close
C:\Users\Admin\OneDrive\Desktop\PRIN 2017\EpiDoc encoding (Faliscan)\V\Ant\Faliscan_45_checked.xml Validation successful U-0068 68 / 136 10 new message(s)
```



Disadvantages of XML-TEI/EpiDoc encoding

EpiDoc Converter - PATRIMONIUM Project

EpiDoc Converter

This converter is based on the EpiDoc converter built for the [Patrimonium Editor](#). It allow users to import text written in Leiden Rules and several other standards: very useful for converting existing, not EpiDoc editions

You can paste below any text formatted according to main epigraphical standards. You can then directly edit the xml produced by the converter in the second editor, by direct input or through the toolbar. A preview pane at the bottom will give back the text in a standard format.

Warning: any action in the top editor will overwrite previous conversion. You may lock and unlock the editor with the lock button

Select source format

Starting line number (if not 1)

Enter Leiden-style text below 

```
1 [---]voltage tursapis patavnos do[---]
```

EpiDoc conversion 

Epigraphic annotations
(1) a á ã Ä å [a] [a?] [a] 'a' <a> [a] a(bc) a(b)c(d) a(bc?) a(- -) (abc) Lacuna Addition Illegible Vacat (- -) IX Symbols (!)

```
1 <lb n="1"/><gap reason="lost" extent="unknown" unit="character"/>voltage tursa<unclear>n</unclear>is patavnos do<gap reason="lost" extent="unknown" unit="character"/>
```



Advantages of ItAnt DSL encoding

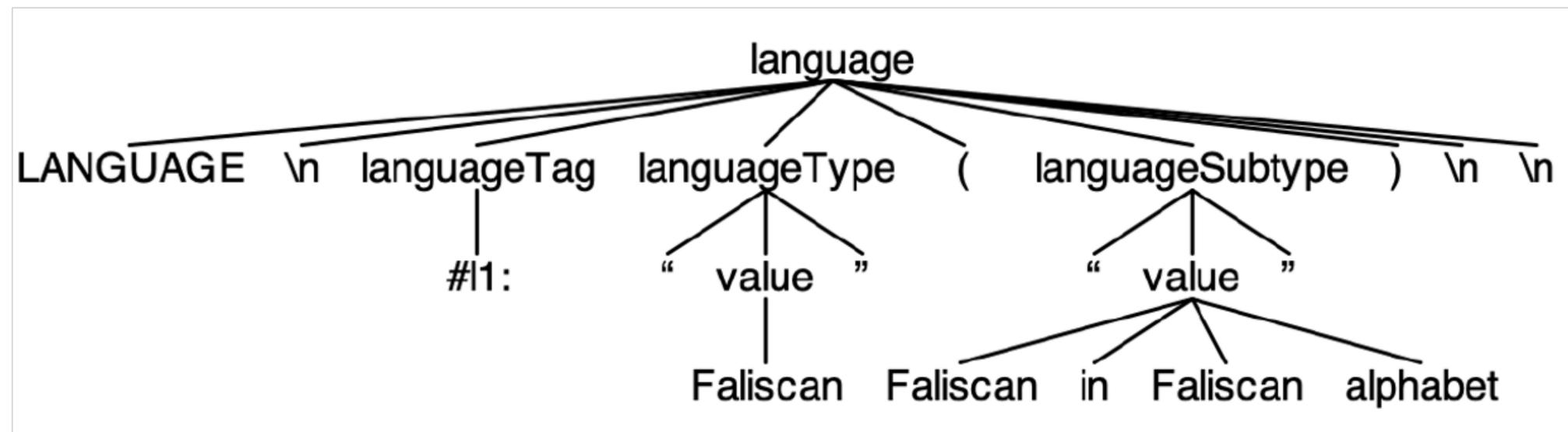
- Familiarity
- Transparency
- Compactness

```
4  
5 IDENTIFIERS  
6 #place: "Schiavi d'Abruzzo (Chieti)"  
7 #inst: "in situ (under the tutelage of the Soprintendenza Archeologia, Belle Arti e Paesaggio dell'Abruzzo)"  
8 #msName: mosaic from the sanctuary of Schiavi  
9 #tm: "TM_170843"  
10 #trad: "ST_Sa_2" "ImIt_Teruentum_36"
```

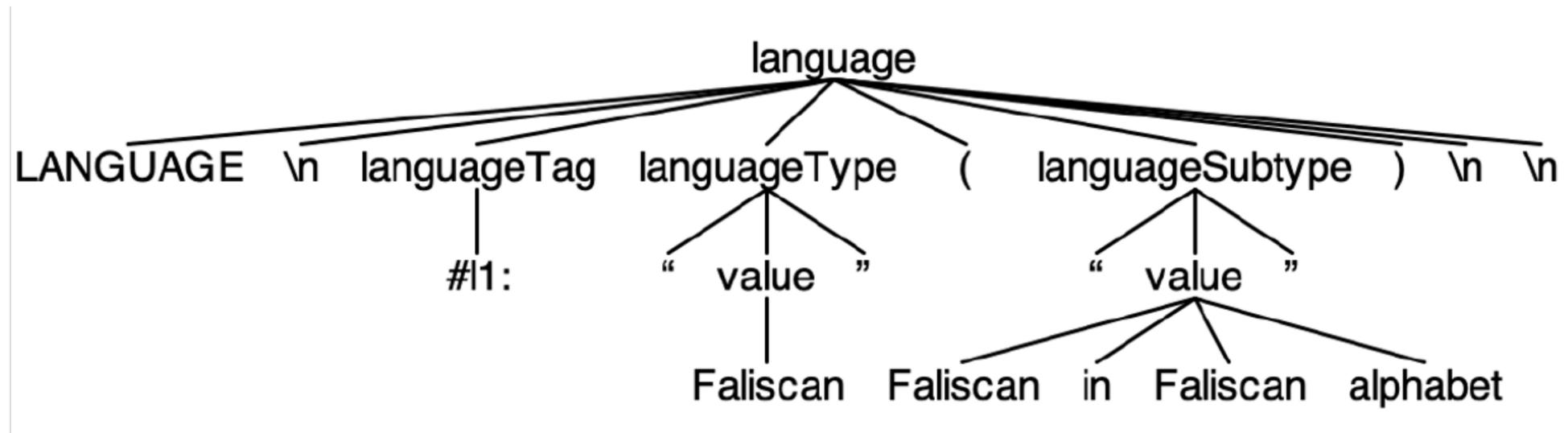
ItAnt DSL parsing

LANGUAGE

#1: "Faliscan" ("Faliscan in Faliscan alphabet")



ItAnt DSL production of XML files



```
<language>
LANGUAGE
<languageTag>#l1:</languageTag>
<languageType>"<value>Faliscan</value>"</languageType>
(<languageSubtype> "<value>Faliscan in Faliscan alphabet</value>"</languageSubtype>)
</language>
```

ItAnt DSL storage of shared information

```
list:
  type: "language"
  items:
    - key: "Faliscan"
      values:
        - type: "ident"
          content: "xfa"
        - type: "source"
          content: "https://iso639-3.sil.org/code/xfa"
    - key: "Faliscan in Faliscan alphabet"
      values:
        - type: "ident"
          content: "xfa-Ital-x-xfafal"
        - type: "source"
          content: "https://www.prin-italia-antica.unifi.it/"
        - type: "ana"
          content: "https://unicode.org/iso15924/iso15924-codes.html"
```



```
<list xmlns="http://itant.eu" type="language">
  <item key="Faliscan">
    <value type="ident">xfa</value>
    <value type="source">https://iso639-3.sil.org/code/xfa</value>
  </item>
  <item key="Faliscan in Faliscan alphabet">
    <value type="ident">xfa-Ital-x-xfafal</value>
    <value type="source">https://www.prin-italia-antica.unifi.it/</value>
    <value type="ana">https://unicode.org/iso15924/iso15924-codes.html</value>
  </item>
</list>
```



ItAnt DSL production of TEI/EpiDoc XML files

XML (proprietary schema)

YAML to XML (proprietary schema)

XSLT

```
<xsl:template match="/dsl:start/dsl:language/dsl:languageType/dsl:value|/dsl:start/dsl:language/dsl:languageSubtype/dsl:value">
  <tei:language>
    <xsl:variable name="key" select="."/>
    <xsl:variable name="languageIdent" select="document('database.xml')/dsl:data/dsl:list/dsl:item[@key=$key]/dsl:value[@type='ident']"/>
    <xsl:variable name="languageSource" select="document('database.xml')/dsl:data/dsl:list/dsl:item[@key=$key]/dsl:value[@type='source']"/>
    <xsl:variable name="languageAna" select="document('database.xml')/dsl:data/dsl:list/dsl:item[@key=$key]/dsl:value[@type='ana']"/>
    <xsl:attribute name="ident">
      <xsl:value-of select="$languageIdent"/>
    </xsl:attribute>
    <xsl:if test="$languageSource">
      <xsl:attribute name="source">
        <xsl:value-of select="$languageSource"/>
      </xsl:attribute>
    </xsl:if>
    <xsl:if test="$languageAna">
      <xsl:attribute name="ana">
        <xsl:value-of select="$languageAna"/>
      </xsl:attribute>
    </xsl:if>
    <xsl:value-of select="$key"/>
  </tei:language>
</xsl:template>
```

```
<tei:langUsage>
  <tei:language ident="xfa" source="https://iso639-3.sil.org/code/xfa">Faliscan</tei:language>
  <tei:language ident="xfa-Ital-x-xfafal"
    source="https://www.prin-italia-antica.unifi.it/"
    ana="https://unicode.org/iso15924/iso15924-codes.html">Faliscan in Faliscan alphabet</tei:language>
</tei:langUsage>
```



Bibliography

- Boschetti, Federico, Luca Rigobianco, e Valeria Quochi. 2024. «Domain-Specific Languages for Epigraphy: The Case of ItAnt». In Selected papers from the CLARIN Annual Conference, 191–202. Linköping Electronic Conference Proceeding.

