



Università
Ca' Foscari
Venezia



UNIVERSITÀ
DEGLI STUDI
FIRENZE



Istituto di Linguistica
Computazionale
"Antonio Zampolli"

 Consiglio Nazionale delle Ricerche

ItAnt e TEI/EpiDoc

Francesca Murano – Luca Rigobianco

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



Archivio digitale dei testi

- Il corpus dei testi selezionato per il progetto è stato codificato in TEI/EpiDoc
 - standard per la rappresentazione dei testi epigrafici in formato digitale
 - sottoinsieme del vocabolario standard della Text Encoding Initiative (TEI)
- Il formato XML favorisce compatibilità e interoperabilità



ItAnt e EpiDoc

- Lo schema di EpiDoc non è sembrato sufficientemente dettagliato per descrivere alcune questioni connesse con i testi di ItAnt
- Poiché ItAnt fa parte di uno studio linguistico storico di materiali dall'Italia, la descrizione di specifiche caratteristiche epigrafiche intrecciate con caratteristiche linguistiche è particolarmente rilevante per il progetto
 - natura frammentaria delle lingue dell'Italia antica
 - peculiarità della loro documentazione epigrafica
 - taglio linguistico del progetto ItAnt
 - formalizzazione delle informazioni epigrafiche interessanti per il recupero dei dati



Personalizzazione dello schema EpiDoc

- Aggiunte al set standard di tag TEI/EpiDoc
 - per descrivere in profondità le diverse condizioni e situazioni che si presentano nei nostri testi
- Differenziazione delle informazioni riguardanti le lingue e le scritture
 - lingua e scrittura appartengono a livelli concettuali distinti
 - in EpiDoc non sono distinte
- Le soluzioni che abbiamo adottato potrebbero essere utili per altri progetti



Integrazione dei dati

- ItAnt sfrutta concetti provenienti da vocabolari e gazetteer ampiamente riconosciuti
 - tipo di oggetto
 - materiali
 - tecnica
 - tipo di iscrizione
 - luoghi
 - identificativi



Integrazione dei dati

Tipologia dell'oggetto

- Per il supporto
 - The Art & Architecture Thesaurus (AAT), The Getty Research Institute
 - per i manufatti
 - iDAI.thesauri, Deutsches Archäologisches Institut
 - per i supporti naturali

<tei:support>

<tei:objectType ana="http://vocab.getty.edu/page/aat/300010699">tegula</tei:objectType>

</tei:support>



Integrazione dei dati

Materiale dell'oggetto

- Vocabolario: The Art & Architecture Thesaurus (AAT)

```
<tei:support>  
  <tei:objectType ana="http://vocab.getty.edu/page/aat/300002469">wall</tei:objectType>  
  <tei:material ana="http://vocab.getty.edu/page/aat/300011264">tuff</tei:material>  
</tei:support>
```



Integrazione dei dati

Tecnica di esecuzione

- Vocabolario: The Art & Architecture Thesaurus (AAT)

```
<tei:layoutDesc>  
  <tei:layout columns="1" writtenLines="2">  
    <tei:rs type="execution" ana="http://vocab.getty.edu/page/aat/300053829">engraved</tei:rs>  
  </tei:layout>  
</tei:layoutDesc>
```



Integrazione dei dati

Tipologia di iscrizione

- Vocabolario: EAGLE Vocabulary (AAT)

```
<tei:textClass>
```

```
  <tei:keywords scheme="https://www.eagle-network.eu/resources/vocabularies/typeins/">
```

```
    <tei:term ref="https://www.eagle-network.eu/voc/typeins/lod/86">building inscription</tei:term>
```

```
  </tei:keywords>
```

```
</tei:textClass>
```



Integrazione dei dati

Entità geografiche

- Pleiades
 - Entità geografiche antiche
- GeoNames
 - Entità geografiche moderne

```
<tei:origin>
```

```
<tei:origPlace type="composed">
```

```
<tei:placeName type="ancient" ref="https://pleiades.stoa.org/places/413126">Falerii Veteres</tei:placeName>
```

```
<tei:placeName type="modern" ref="https://www.geonames.org/3178595">Civita Castellana</tei:placeName>
```

```
</tei:origPlace>
```

```
</tei:origin>
```



Integrazione dei dati

Entità geografiche

- Non sono stati utilizzati vocabolari per date e periodi (es. PeriodO), poiché le iscrizioni relative al nostro dominio sono per lo più datate utilizzando intervalli di tempo estremamente ampi



Integrazione dei dati

Identificativi

- Riferimento a Trismegistos

```
<tei:altIdentifier type="trismegistos">
```

```
  <tei:idno source="www.trismegistos.org/text/218086">TM 218086</tei:idno>
```

```
</tei:altIdentifier>
```

```
<tei:altIdentifier type="traditionalID">
```

```
  <tei:idno source="http://zotero.org/groups/2552746/items/JQ7H6CNX">LV Es 31</tei:idno>
```

```
</tei:altIdentifier>
```

TEI/EpiDoc

- PRO
 - TEI/EpiDoc è uno strumento versatile e personalizzabile che può essere facilmente adattato alle varie esigenze
- CON
 - l'interoperabilità può essere compromessa dall'eccessiva personalizzazione dello schema



Personalizzazioni dello schema



Supporto

Riutilizzo del supporto scrittorio

- Uso di `<tei:rs type="reuse">` in `<tei:support>`
- Valore binario



`<tei:supportDesc>`

`<tei:support>`

`<tei:objectType ana="http://archwort.dainst.org/en/term/4572">stone slab</tei:objectType>`

`<tei:rs ana="objectShape">quadrangular</tei:rs>`

`<tei:material ana="http://vocab.getty.edu/aat/300011286">limestone</tei:material>`

`<tei:rs type="reuse" resp="#La_Regina2019">no</tei:rs>`

`</tei:support>`

Supporto

Condizioni del supporto scrittorio

- Uso di <tei:condition> in <tei:supportDesc>
- 4 valori: : ‘well-preserved’, ‘fragmentary’, ‘very fragmentary’, ‘lost inscription’

```
<tei:supportDesc>
  <tei:support>
    <tei:objectType ana="http://archwort.dainst.org/en/term/4572">stone slab</tei:objectType>
    <tei:rs ana="objectShape">quadrangular</tei:rs>
    <tei:material
ana="http://vocab.getty.edu/aat/300011286">limestone</tei:material>
    <tei:rs type="reuse" resp="#La_Regina2019">no</tei:rs>
  </tei:support>
  <tei:condition>very fragmentary</tei:condition>
</tei:supportDesc>
```

Layout

Tipo di esecuzione

- Uso di `<tei:rs type="execution">` in `<tei:layoutDesc>`
- Uso del vocabolario AAT per i valori

```
<tei:layoutDesc>  
  <tei:layout columns="1" writtenLines="48">  
    <tei:rs type="execution" ana="http://vocab.getty.edu/page/aat/300404794">engraved</tei:rs>  
    <tei:rs type="opistography">yes</tei:rs>  
  </tei:layout>  
</tei:layoutDesc>
```


Layout

Opistografia

- Uso di `<tei:rs type="opistography">` in `<tei:layoutDesc>`
- Valore binario



```
<tei:layoutDesc>  
  <tei:layout columns="1" writtenLines="48">  
    <tei:rs type="execution" ana="http://vocab.getty.edu/page/aat/300404794">engraved</tei:rs>  
    <tei:rs type="opistography">yes</tei:rs>  
  </tei:layout>  
</tei:layoutDesc>
```

Layout

Dimensione delle linee

- Uso di `<tei:dimensions type="lineDimension">` in `<tei:layoutDesc>`

```
<tei:layoutDesc>
```

```
  <tei:layout columns="1" writtenLines="6">
```

```
    <tei:rs type="execution" ana="http://vocab.getty.edu/page/aat/300404794">engraved</tei:rs>
```

```
    <tei:rs type="opistography">no</tei:rs>
```

```
    <tei:dimensions type="lineDimension" unit="cm" precision="medium" corresp="#Osc_4_l_1" resp="Francesca Murano">
```

```
      <tei:height quantity="3">3</tei:height>
```

```
      <tei:width quantity="23.8">23.8</tei:width>
```

```
    </tei:dimensions>
```

```
    <tei:dimensions type="lineDimension" unit="cm" precision="medium" corresp="#Osc_4_l_2" resp="Francesca Murano">
```

```
      <tei:height>3</tei:height>
```

```
      <tei:width>18.8</tei:width>
```

```
    </tei:dimensions>
```


Layout

Dimensione dei caratteri

- Uso di `<tei:dimensions type="characterDimension">` in `<tei:handNote>`

```
<tei:handDesc>
  <tei:handNote>
    <tei:dimensions type="characterDimension" unit="cm" precision="high" corresp="#Oscan_Text">
      <tei:height atLeast="4" atMost="5.5">4 to 5,5</tei:height>
    </tei:dimensions>
    <tei:dimensions type="characterDimension" unit="cm" precision="high" corresp="#Latin_Text">
      <tei:height atLeast="4" atMost="8">4 to 8</tei:height>
    </tei:dimensions>
  </tei:handNote>
</tei:handDesc>
```


Script

Separazione delle parole

- Uso di `<tei:rs type="wordDivision">` in `<tei:scriptNote>`
- 4 valori
 - ‘scriptio continua’, ‘punctuation’, ‘blank spaces’, ‘mixed’

```
<tei:scriptNote>
```

```
  <tei:rs type="wordDivision">punctuation</tei:rs>
```

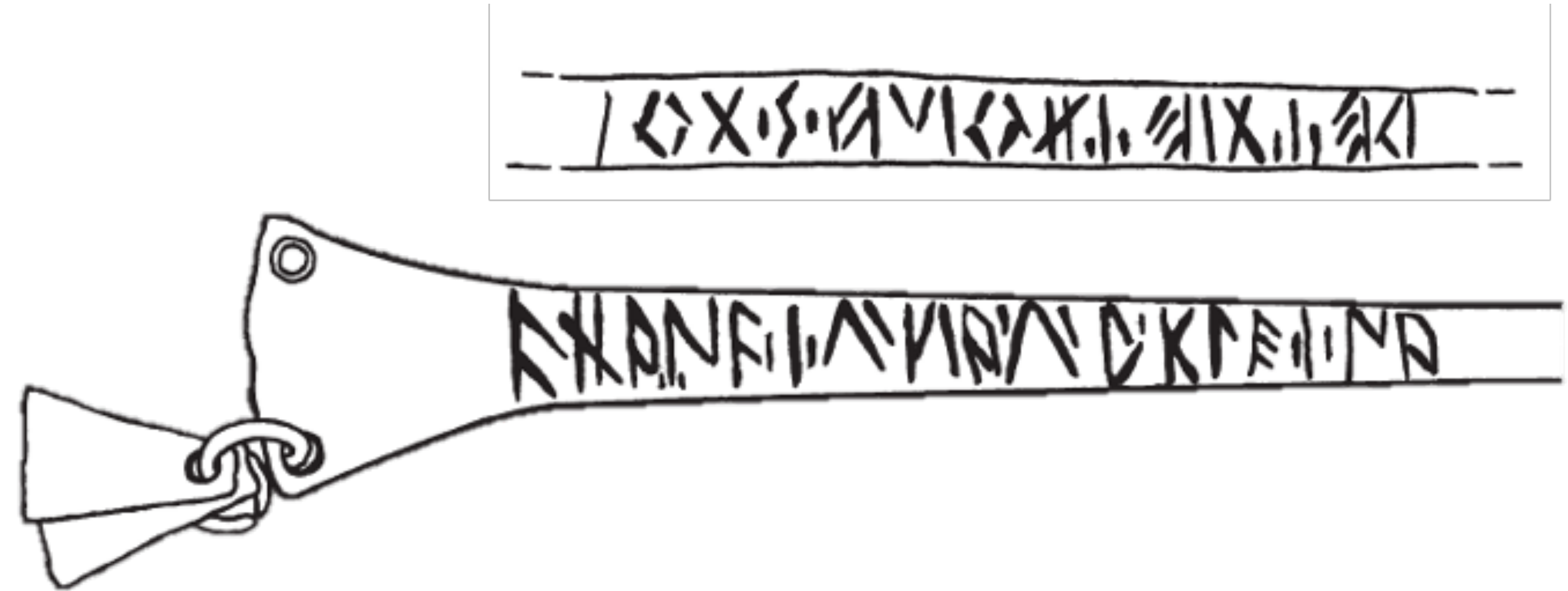
```
</tei:scriptNote>
```



Script

Separazione delle parole

- Puntuazione sillabica (per le iscrizioni venetiche)
- 4 valori
 - ‘simplified’, ‘unsimplified’;
 - ‘properly applied’, ‘unproperly applied’



<tei:scriptNote>

<tei:rs type="syllabicPunctuationApplication">yes</tei:rs>

<tei:rs type="syllabicPunctuationSimplification">yes</tei:rs>

</tei:scriptNote>

Lingue e scritture

- EpiDoc non ha specifiche istruzioni per codificare separatamente la lingua e la scrittura

Languages And Scripts



TEI/EpiDoc provides mechanisms for encoding languages and scripts (writing systems) as they relate to the contents of an EpiDoc file and to the text(s) described and transcribed therein. In doing so, we make use of Internet standards for the identification of these languages and scripts. This portion of the *Guidelines* addresses all relevant aspects.

Relevant element documentation (TEI):

- TEI definition: [langUsage](#); EpiDoc-specific customization: [langUsage](#)
- TEI definition: [language](#); EpiDoc-specific customization: [language](#)
- TEI definition: [textLang](#); EpiDoc-specific customization: [textLang](#)
- TEI definition: [foreign](#); EpiDoc-specific customization: [foreign](#)

Indicating languages and scripts used in an EpiDoc file

TEI and EpiDoc follow the best current practice outlined in the Network Working Group's [RFC 5646: Tags for Identifying Languages](#), which establishes the norms for same on an Internet-wide basis. The RFC and supporting documents define a syntax for creating short strings of characters ('language tags') that function as unique identifiers for any desired combination of language and script. These tags are composed of 'subtags' for language qua language, writing system (script), and regional and dialectal variation. The RFC also establishes a process for registration and maintenance of these subtags by the [Internet Assigned Numbers Authority](#).

A valid EpiDoc file must make use of subtags recorded in the [IANA Language Subtag Registry](#). Many EpiDoc creators will already be familiar with some of these codes from other digital projects, for example:

- Grek = Greek script
- Latn = Latin script
- en = English language (assumed to be in its standard script: Latn)
- fr = French language (assumed to be in its standard script: Latn)
- el = Modern Greek language (1453-; assumed to be in its standard script: Grek)
- grc = Ancient Greek language (to 1453; assumed to be in its standard script: Grek)
- grc-Latn = Ancient Greek Language (to 1453), rendered in Latin script

When the IANA registry does not provide appropriate codes, then an EpiDoc project may devise "private use subtags", so long as they are internally defined in the EpiDoc file as outlined in the following paragraph and so long as they conform syntactically to the specifications laid out in RFC 5646, sections [2.1: Syntax](#) and [4.6: Considerations for Private Use Subtags](#). For example, the Campà Inscriptions team determined that the two Cham language subtags (cja = Western Cham and cjm = Eastern Cham) and the associated script subtag (Cham) were substantively different from the ancient Cham language and script represented in the inscriptions. Therefore the private use subtag "x-oldcam-latn-ci" was invented and given the project-specific meaning "Old Cam language in Old Cam script transliterated in Latin characters." Whenever possible, EpiDoc projects and practitioners should undertake to register new subtags with the IANA for the benefit of others. A procedure for same is set out in RFC 5646 [Section 3.5](#).

```
<langUsage>
<language id="ar">Arabic</language>
<language id="cop">Coptic</language>
<language id="egy-Egyd">Egyptian in Demotic script</language>
<language id="egy-Egyh">Egyptian in Hieratic script</language>
<language id="egy-Egyp">Egyptian Hieroglyphic</language>
<language id="etr">Etruscan</language>
<language id="el">Modern Greek</language>
<language id="grc">Ancient Greek</language>
<language id="grc-Latn">Ancient Greek written in Latin script</language>
<language id="he">Hebrew</language>
<language id="la">Latin</language>
<language id="la-Grek">Latin written in Greek script</language>
</langUsage>
```

Lingue e scritture

La soluzione proposta da ItAnt

<tei:langUsage>

<tei:language ident="osc" source="https://iso639-3.sil.org/code/osc">Oscan</tei:language>

<tei:language ident="osc-Ital-x-oscetr" source="https://www.prin-italia-antica.unifi.it/p166.html" ana="https://unicode.org/iso15924/iso15924-codes.html">Oscan in Oscan national alphabet</tei:language>

</tei:langUsage>

Oscan [osc]						
Identifier	Language Name(s)	Status	Code Sets	Scope	Language Type	Denotations
osc	Oscan	Active	639-3	Individual	Historical	Ethnologue, Glottolog, Wikipedia

Ital	210	Old Italic (Etruscan, Oscan, etc.)	ancien italique (étrusque, osque, etc.)	Old_Italic	3.1	2004-05-29
------	-----	------------------------------------	---	------------	-----	------------

Lingue e scritture

La soluzione proposta da ItAnt

<tei:scriptDesc>

<tei:scriptNote>

<tei:rs type="writingSystem" subtype="alphabet" ref="https://www.prin-italia-antica.unifi.it/p166.html">Oscan
national alphabet</tei:rs>

</tei:scriptNote>

</tei:scriptDesc>



Onomastica

La soluzione proposta da ItAnt

Cfr. Bl 1 ***Enoni*** ***Ontei*** ***Appioi*** ***sselboisselboi*** ***Andeticobos***
ecupetaris



Onomastica

La soluzione proposta da ItAnt

```
<tei:listPerson type="onomastics">
  <tei:person xml:id="p1" corresp="#Osc_2_L_1">
    <tei:persName>
      <tei:name type="praenomen">pakis</tei:name>
      <tei:name type="gentilicium">heleviis</tei:name>
      <tei:name type="patronymic">tre(bieís)</tei:name>
    </tei:persName>
  </tei:person>
```



Bibliografia

- Elliott, Tom, Gabriel Bodard, Elli Mylonas, Simona Stoyanova, Charlotte Tupman, e Scott Vanderbilt. 2024. «EpiDoc Guidelines: Ancient documents in TEI XML (Version 9)». 2024.
- Murano, Francesca, Valeria Quochi, Angelo Mario Del Grosso, Luca Rigobianco, e Mariarosaria Zinzi. 2023. «Describing Inscriptions of Ancient Italy. The ItAnt Project and Its Information Encoding Process». Journal on Computing and Cultural Heritage, 1–14.

