

Scuola estiva CADottorato

TEI (seconda parte): un'altro testo, metadati, ecc.

Simon Gabay

Verona, 17 luglio 2019

Metadati

Kézaco?

- Questi sono dati digitali usati per rappresentare o descrivere altri dati (digitali o no).
- Forniscono informazioni sulla fonte, natura, contenuto, storia, ecc. del documento che descrivono.
- Possono (devono?) essere standardizzati.

Utilità

- Forniscono un indice che rende più facile e veloce la ricerca.
- La standardizzazione semplifica lo scambio dei dati (parliamo di interoperabilità).

Metadati e TEI

- In un documento codificato in TEI, si trovano nel `<teiHeader>` (cf. TEI) i metadati del documento.
- Il `<teiHeader>` fornisce una descrizione strutturata dei dati contenuti nel documento XML.
- Alcuni elementi sono obbligatori, altri sono opzionali.
- La gerarchia dei dati è limitata dallo schema.

Un documento TEI minimale

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>Title</title>
      </titleStmt>
      <publicationStmt>
        <p>Publication Information</p>
      </publicationStmt>
      <sourceDesc>
        <p>Information about the source</p>
      </sourceDesc>
    </fileDesc>
  </teiHeader>
  <text>
    <body>
      <p>Some text here.</p>
    </body>
  </text>
</TEI>
```

La Santissima Trinità del <teiHeader>

1. <titleStmt> (cf. TEI) raggruppa le informazioni sul titolo di un'opera e sulle responsabilità del suo contenuto intellettuale.

```
<titleStmt>
  <title>Esercizio durante la scuola estiva</title>
</titleStmt>
```

2. <publicationStmt> (cf. TEI) raggruppa le informazioni riguardo la pubblicazione o la distribuzione di un documento elettronico.

```
<publicationStmt>
  <p>Simon Gabay, UniNE. CC-BY.</p>
</publicationStmt>
```

3. <sourceDesc> (cf. TEI) fornisce una descrizione relativa alla o alle fonti da cui è derivato o generato un documento elettronico.

```
<sourceDesc>
  <p>_La Commedia secondo l'antica vulgata_, a cura di Gio
    Petrocchi, 4 voll., Milano, A. Mondadori, 1966-67.</p>
</sourceDesc>
```

<titleStmt> vs <sourceDesc>

- **<titleStmt>** non è il nome dell'opera (letteraria) *codificata*, ma dell'edizione (critica) *prodotta*.
- Il nome dell'edizione di Petrocchi non è la *Commedia* ma *La Commedia secondo l'antica vulgata*. Il titolo dell'edizione contiene il titolo originale, ma non solo.
- **<titleStmt>** et **<sourceDesc>** sono simili, ma rimangono fondamentalmente diversi.
- Questa distinzione ha più senso nel caso di una monografia (sì, possiamo scrivere una tesi in TEI, forse dovremmo...), il cui titolo è necessariamente diverse delle fonti.

Perché essere semplici se possiamo complicarci?

Possiamo codificare cose molto diverse in TEI, il che spiega alcune cose strane. Ad esempio, perché codificare come segue:

```
<titleStmt>  
  <title>Esercizio durante la scuola estiva</title>  
</titleStmt>
```

E non:

```
<titleStmt>Esercizio durante la scuola estiva</titleStmt>
```

Perché? Perché si

Perché questa è la versione minima del `<titleStmt>`, in cui possiamo aggiungere altre informazioni oltre all'elemento `<title>` (cf. TEI).

```
<titleStmt>
  <title>Esercizio durante la scuola estiva</title>
  <author>Dante</author>
  <editor>
    <persName>
      <forename>Simon</forename>
      <surname>Gabay</surname>
    </persName>
  </editor>
</titleStmt>
```

`<author>` (cf. TEI) potrebbe essere codificato come `<editor>` (cf. TEI) con un `<persName>` (cf. TEI), un `<forename>` (cf. TEI) e un `<surname>` (cf. TEI).

L'encodage emmental

La codifica in XML-TEI è una codifica con dei buchi che possono essere riempiti in base alle nostre esigenze e in un modo semplice. Questo non è il caso per tutti linguaggi (<- critica infelice degli informatici che non capiscono che gli umanisti usano ancora XML).

Benefici della globalizzazione (in TEI)

Abbiamo appena visto apparire il tag `<persName>`. Questo tag non è unico per il `<teiHeader>`, e si trova ovunque in un documento TEI, come nel tag `<l>` (cf. TEI) del `<body>` (cf. TEI) che abbiamo visto in precedenza.

```
<l>quelli è <persName>0mero</persName> poeta sovrano;</l>  
<l>l'altro è <persName>0razio</persName> satiro che vene;
```

Idem per `<title>`, che anche si può trovare in diversi posti in un documento TEI, come ad esempio nel `<bibl>`:

```
<bibl>  
  <author>Dante</author>  
  <title>La Commedia</title>  
  <editor>Giorgio Petrocchi</editor>  
  <pubPlace>Milano</pubPlace>  
  <publisher>Mondadori</publisher>  
  <date>1966</date>  
</bibl>
```

Limiti della globalizzazione (sempre in TEI)

Tuttavia, non è possibile riciclare tutti i tag dal TEI (e.g. `<titleStmt>`)
e nel caso sia possibile non possiamo farlo in tutti i posti (e.g.
`<author>`):

```
<p><author>Dante</author> è l'autore della  
  <title>Commedia</title></p>
```

L'elemento `<title>` può essere messo in un `<p>` , ma non in un
`<author>` .

Limiti della globalizzazione (II)

A volte è permesso utilizzare un elemento, anche quando non è quello più adatto, come ad esempio `<persName>` in `<bibl>` (cf. TEI):

```
<bibl>
  <persName>Dante</persName>
  <title>La Commedia</title>
  <persName>Giorgio Petrocchi</persName>
  <pubPlace>Milano</pubPlace>
  <publisher>Mondadori</publisher>
  <date>1966</date>
</bibl>
```

Sebbene sia possibile specificare:

```
<bibl>
  <persName type="autore">Dante</persName>
  <title>La Commedia</title>
  <persName type="editore">Giorgio Petrocchi</persName>
  ...
</bibl>
```

TEI di base

- `<div>` per la divisione del testo in parti, capitoli, poemi, ecc.
- Utilizziamo un attributo `@type` per specificare il tipo (`<div type="chapter">`).
- `<pb/>` per l'inizio della pagina
- `<lb/>` per l'inizio della riga
- `<hi>` (*highlight*) per rendere quello che è in corsivo nel testo aggiungiamo un attributo (`<hi rend="it">`)

- `<lg>` (*line group*) per la stanza
- `<p>` (*paragraph*) per il paragrafo
- `<ab>` (*anonymous block*) per quello che non è né una stanza né un paragrafo (e.g. indirizzo postale)
- `<l>` (*line*) per il verso
- `<s>` (*sentence*) per la frase
- `<w>` (*word*) per la parola
- `<said>` per il discorso diretto

- `<name>` per i nomi
- `<persName>` per i nomi delle persone
- `<placeName>` per i nomi dei luoghi
- `<surname>` per il cognome
- `<forename>` per il nome
- `<country>` per il paese
- `<region>` per la regione
- `<settlement>` per la città
- `<note>` per le note