



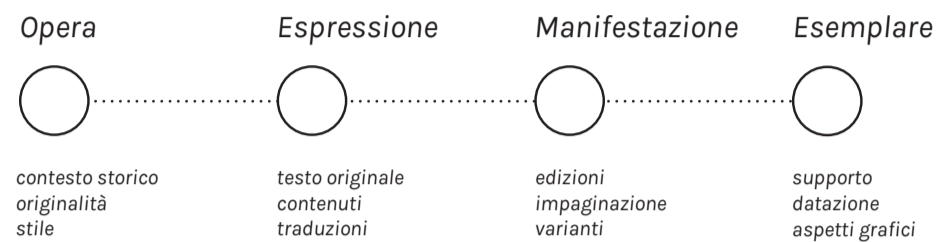
Manoscritti e testi digitali: trascrizione, marcatura e trasformazione

Le problematiche delle edizioni scientifiche digitali

La manipolazione del testo da parte dell'editore, nonché le scelte effettuate per la rappresentazione digitale influiscono sul risultato finale.



Le problematiche del testo



Digitalizzazione e Trascrizione

Il manoscritto viene digitalizzato/ scannerizzato ad alta risoluzione ed è disponibile per il confronto durante la trascrizione. In questa fase l'editore analizza i contenuti e le scelte dell'autore (e.g. cancellature, aggiunte, grafemi, varianti, etc.).

rich text editor

image markup



Codifica del testo

Il testo viene **codificato** per essere riprodotto su supporto digitale.

In questa fase l'editore digitale interpreta le scelte dell'autore, degli (eventuali) editori precedenti e vi aggiunge le proprie interpretazioni (e.g. riconoscimento di nomi, luoghi, date, citazioni, note, apparato, etc.).

```
<bibl resp="#PB">
  <persName key="PVM" role="author">
    <choice>
      <sic>Verg.</sic>
      <expan>Vergilius</expan>
    </choice>
  </persName>
  <title>
    <choice>
      <sic>Aen.</sic>
      <expan>Aeneis</expan>
    </choice>
  </title>
  <biblScope>II</biblScope>
</bibl>
```

Trasformazione

Il documento viene **trasformato** per poter essere **visualizzabile** su una macchina. In questa fase, lo sviluppatore interpreta la volontà degli editori per la visualizzazione dei fenomeni testuali (e.g. impaginazione, paratesto, note, etc.).

XML parser

TEI validator

browser

Visualizzazione

Il documento viene **visualizzato online**: il testo e la digitalizzazione del manoscritto sono esplorabili mediante un sito web. In questa fase il web designer interpreta le richieste dell'editore per la visualizzazione dei fenomeni grafici (e.g. colori, font, paginazione, etc.).

image viewer

transcription viewer

Arricchimento semantico

Il documento viene **arricchito**: l'espressività del testo viene valorizzata. In questa fase esperti di dominio arricchiscono la conoscenza estratta durante il processo con conoscenza disponibile in altre fonti o fornita dagli utenti (e.g. content analysis, annotazione semantica, relazioni tra fonti).

```
<person xml:id="PVM">
  <persName sameAs="
    http://viaf.org/viaf/339144647673697600501
    http://it.dbpedia.org/resource/Publio_Virgilio_Marone">
    Virgilio Marone, Publio (70-19 a.c.)
  </persName>
</person>
```

semantic annotation

text analysis and mining

XML (eXtensible Markup Language)
il linguaggio standard per la codifica (o marcatura) dei testi

TEI (Text Encoding Initiative)
è lo schema XML per descrivere i fenomeni del testo umanistico

XSLT (eXtensible Stylesheet Language - Transformations)
linguaggio di trasformazione di documenti XML

HTML (HyperText Markup Language)
linguaggio di markup per ipertesti

CSS (Cascading Style Sheets)
linguaggio per la grafica di ipertesti

JS (JavaScript)
linguaggio di scripting per la dinamicità dei contenuti ipertestuali

RDF (Resource Description Framework)
modello per la rappresentazione dei dati

OWL (Web Ontology Language)
linguaggio per la formalizzazione di ontologie

