

FRANCESCA GIOVANNETTI

Consortium  
**GARR**

THE ITALIAN  
EDUCATION  
& RESEARCH  
NETWORK



# DSEasLOD

## Digital Scholarly Editions as Linked Open Data

GIORNATA DI INCONTRO  
BORSE DI STUDIO GARR  
"ORIO CARLINI"  
6 DICEMBRE 2018 ROMA

Roma 06/12/2018

Giornata di incontro Borse di Studio "Orio Carlini"

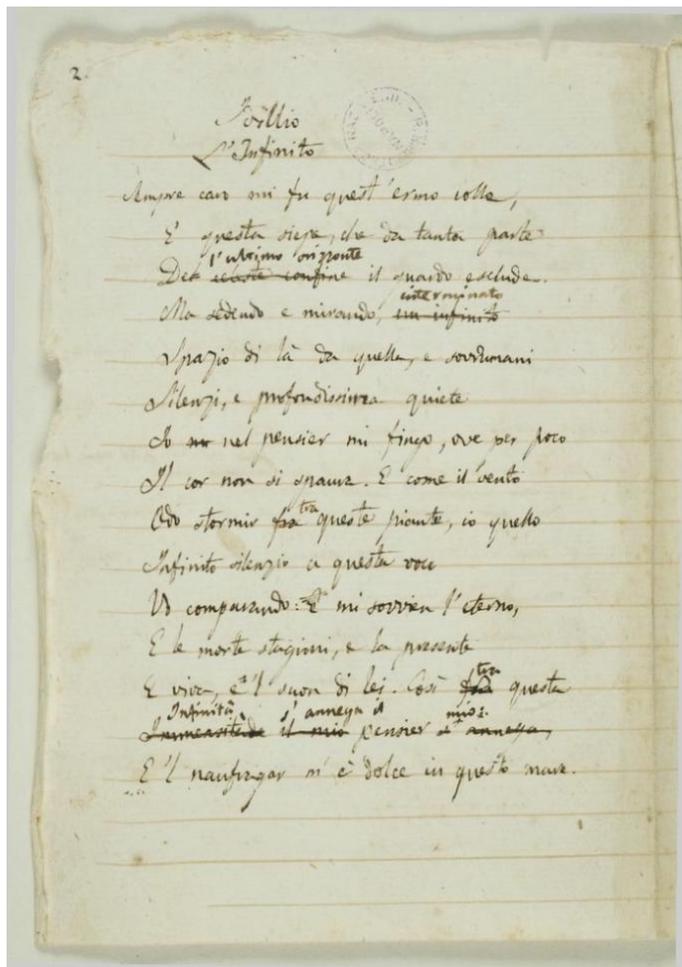


## Contenuto della presentazione

- I. Tema della borsa di studio
- II. Il paradigma Linked Open Data
- III. Una definizione di Digital Scholarly Edition
- IV. Il tool DSEasLOD
- V. Stato di avanzamento



## Tema della borsa di studio GARR

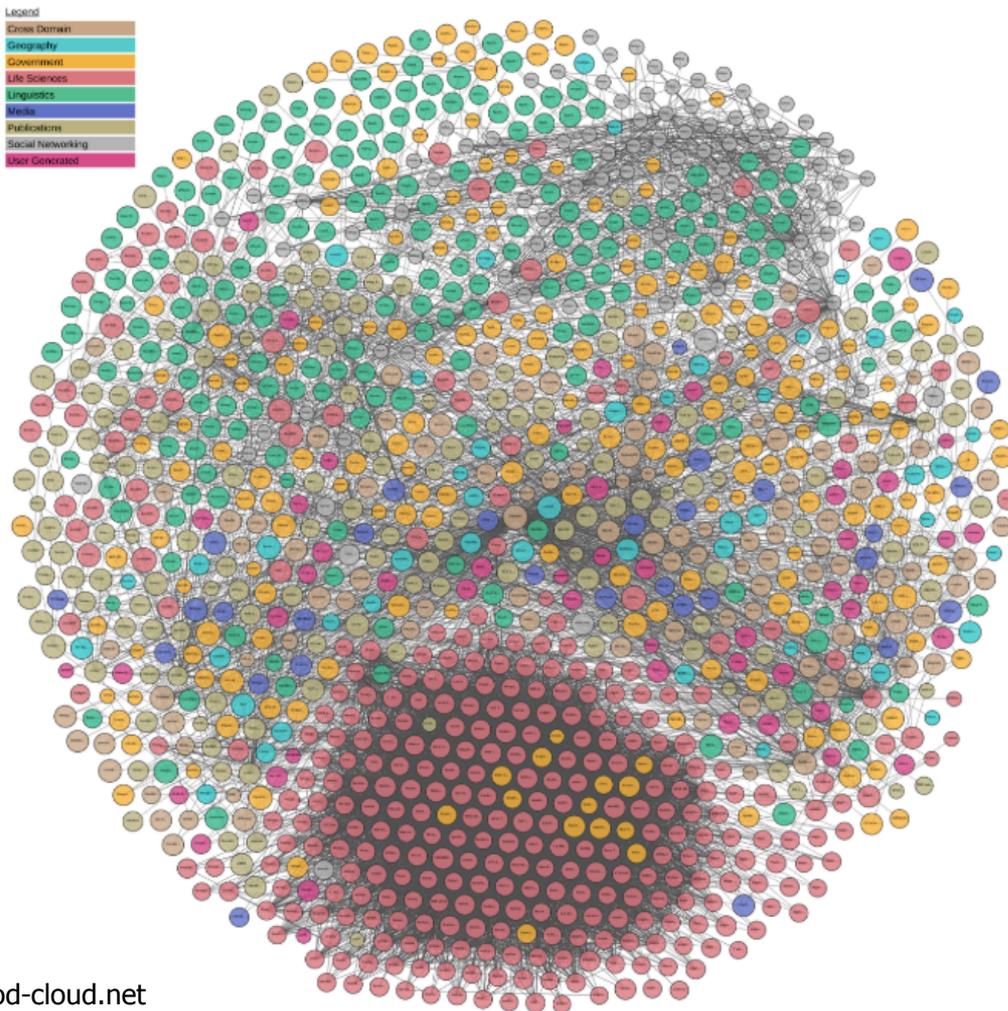


Sviluppo di un'applicazione web per  
l'arricchimento semantico di edizioni  
critiche digitali già codificate in XML/TEI,  
standard di riferimento per le edizioni  
critiche in ambito umanistico, attraverso  
l'estrazione user-friendly di grafi di  
conoscenza RDF.

"L'infinito" di Giacomo Leopardi (manoscritto  
autografo, 1819) Biblioteca Nazionale di Napoli



# The Linked Open Data cloud (giugno 2018)

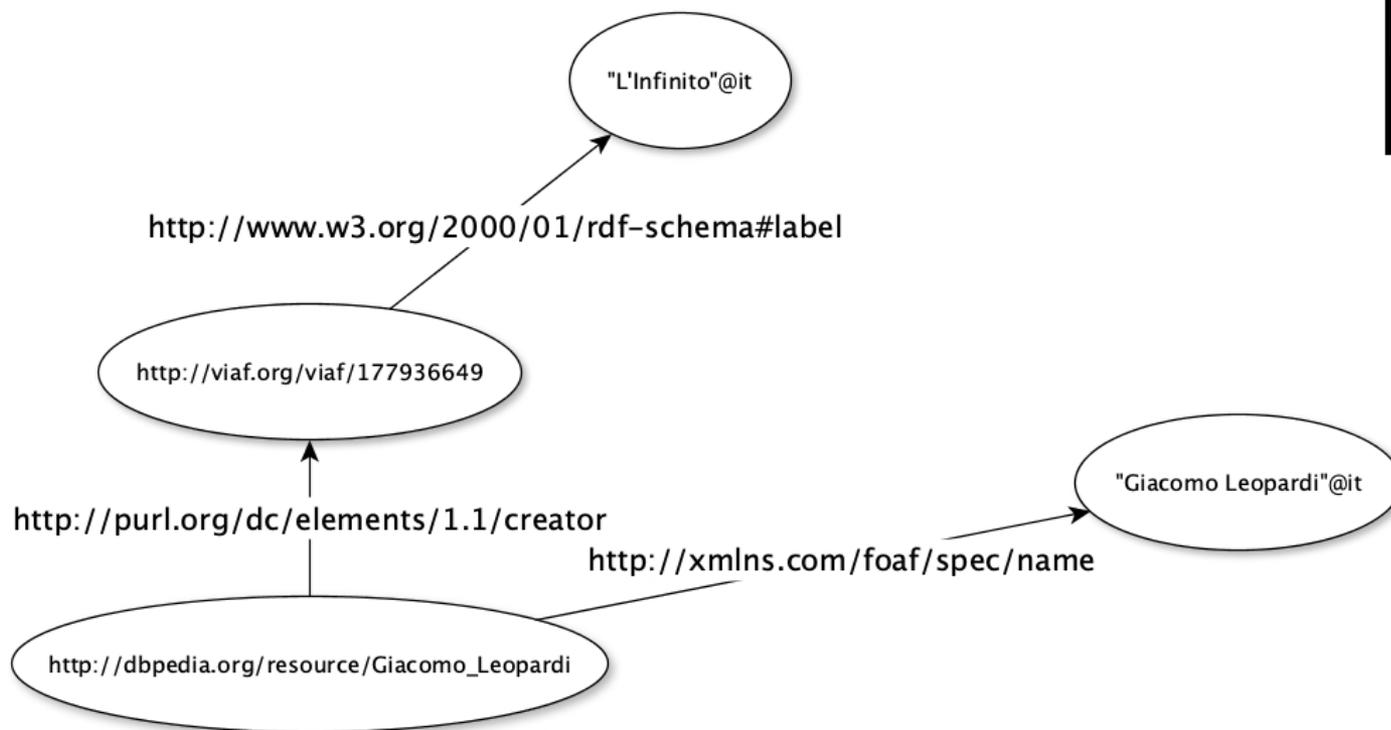


The Linked Open Data Cloud from lod-cloud.net



## Il LOD data model: RDF

Giacomo Leopardi è l'autore di *L'infinito*.





# Digital Scholarly Edition: definizione sintetica

(Sahle 2016)

---

"A scholarly edition is the critical representation of historic documents."

"A digital edition cannot be given in print without significant loss of content and functionality."



# Text Encoding Initiative (TEI)

## Standard per la codifica di DSE

```
<person xml:id="WM">
<!-- ... -->
<event type="marriage" when="1859-04-26">
  <label>Marriage</label>
  <desc>
    <name type="person" ref="#WM">William Morris</name> and <name type="person"
      ref="http://en.wikipedia.org/wiki/Jane_Burden">Jane Burden</name> were
      married at <name type="place">St Michael's Church, Ship Street, Oxford</name> on
    <date when="1859-04-26">26 April 1859</date>. The wedding was
      conducted by Morris's friend <name type="person" ref="#RWD">R. W.
        Dixon</name> with <name type="person" ref="#CBF">Charles
          Faulkner</name> as
        the best man. The bride was given away by her father,
    <name type="person" ref="#RB">Robert Burden</name>.
```

Estratto da <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/examples-person.html>



## Scenario

- **Complessità intrinseca** legata alla produzione dell'edizione critica in XML/TEI
- Adozione di **soluzioni di codifica ad hoc** che richiedono processi di trasformazione specifici
- **Carenza di competenze digitali** avanzate nell'ambito degli studi umanistici

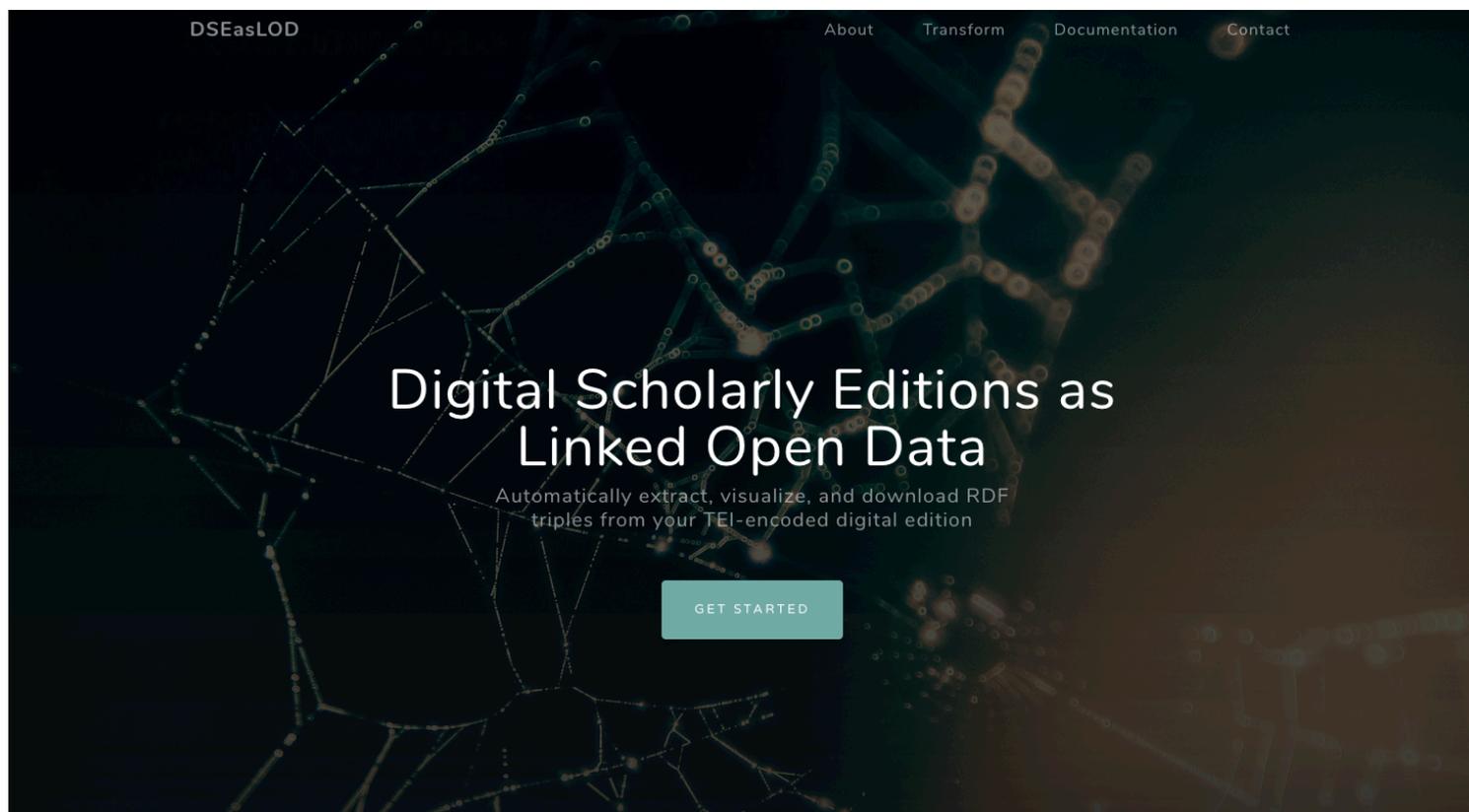


Difficoltà di trasformazione dell'edizione critica XML/TEI da entità semi-strutturata a grafo RDF e dunque di ingresso nel **LOD cloud dei beni culturali**



## DSEasLOD: che cos'è?

Applicazione web per l'estrazione, la visualizzazione e il download di triple RDF a partire da edizioni critiche XML/TEI





## Triple RDF: risultato dell'estrazione

```
<https://w3id.org/janeburdenbiography/person/WM> a foaf:Person ;  
  owl:sameAs <http://viaf.org/viaf/22146194> ;  
  rdfs:label "William Morris"@en ;  
  pro:holdsRoleInTime <https://w3id.org/janeburdenbiography/role-in-time/WM-at-event001> .
```

```
<https://w3id.org/janeburdenbiography/role-in-time/WM-at-event001> a pro:RoleInTime ;  
  pro:withRole <http://purl.org/vocab/bio/0.1/partner> ;  
  <tv:atTime rdf:resource="https://w3id.org/janeburdenbiography/tvc/event001-time"/>  
  <pro:relatesToEntity rdf:resource="https://w3id.org/janeburdenbiography/event/event001"/>  
  <proles:relatesToPlace rdf:resource="https://www.geonames.org/2640729"/> .
```

```
<https://w3id.org/janeburdenbiography/event/event001> a bio:Marriage ;  
  rdfs:label "Marriage"@en .
```



## DSEasLOD: motivazioni

- aggirare problematiche quali la compresenza di **molteplici gerarchie** all'interno del testo
- associare agli asserti la **provenance** rendendo possibile la presenza simultanea di più interpretazioni relative allo stesso oggetto
- facilitare l'**interoperabilità** fra edizioni
- incentivare l'ingresso delle edizioni nella **LOD cloud** dei beni culturali



## Da TEI a RDF: stato dell'arte (1)

### TEI Stylesheets

- diversi formati di output, fra cui RDF
- utilizzabili online o sul proprio computer

TUTTAVIA

- focus sui formati per la visualizzazione (es. HTML) e sulla generazione di schemas, RDF è marginale
- trasformazione RDF solo per un numero ristretto di elementi TEI
- CIDOC-CRM come modello concettuale



## Da TEI a RDF: stato dell'arte (2)

**Progetti editoriali specifici** che hanno generato e messo a disposizione le proprie stylesheets (es. *Paolo Bufalini: Appunti 1981-1991; Sharing Ancient Wisdoms*)

- costruiti per rispondere ad esigenze ed **obiettivi particolari**
- difficilmente generalizzabili



## DSEasLOD: componenti

- **Applicazione web** sviluppata interamente in Python
- **Repository GitHub** contenente:
  - codice sorgente dell'intera applicazione
  - documentazione estesa
  - fogli di stile XSLT personalizzabili (per gli utenti più esperti)



## DSEasLOD: caratteristiche

- **SPAR (Semantic Publishing and Referencing) Ontologies** come modello per la descrizione RDF
- **Modulare**
  - possibilità di scegliere quali triple estrarre dall'edizione
  - diversi livelli di controllo dell'estrazione disponibili
  - possibilità di customizzazione delle trasformazioni sia Python che XSLT
- **Valore didattico** (estrinsecato attraverso la documentazione)



## DSEasLOD: stato di avanzamento (1)

### DONE

- Stato dell'arte delle DSE e dei tool di conversione da TEI a RDF
- Scelta dei modelli ontologici da riutilizzare, ossia quali relazioni instaurare fra le entità dell'edizione
- Sviluppo dei fogli di stile XSLT

### IN PROGRESS

- Sviluppo dell'applicazione web (framework web.py)
- Implementazione dell'estrazione in Python



## DSEasLOD: stato di avanzamento (2)

### TO DO

- Perfezionamento della trasformazione da TEI a RDF con l’inserimento di meccanismi di rappresentazione della **provenance**
- Implementazione di metodi automatici di disambiguazione delle **Named Entities** all’interno dell’applicazione web
- Ampliamento della **documentazione**
- **Pubblicazione** del progetto su GitHub



## Riferimenti

Berners-Lee, T. (2006) *Linked Data*. <https://www.w3.org/DesignIssues/LinkedData.html>

Immagini tratte da Berners-Lee, T. (2007) *Levels of Abstraction: Net, Web, Graph*.  
<https://www.w3.org/DesignIssues/Abstractions.html>

Sahle, P. (2016) "What Is a Scholarly Digital Edition?" In *Digital Scholarly Editing: Theories and Practices*, edited by Matthew James Driscoll and Elena Pierazzo. Cambridge, UK: Open Book Publishers.

TEI Consortium, eds. (2018) *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. TEI Consortium.  
<http://www.tei-c.org/Guidelines/P5/>

TEI Consortium, eds. (2018) TEI: XSL stylesheets for TEI XML. In *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. TEI Consortium. <http://www.tei-c.org/tools/stylesheets/>

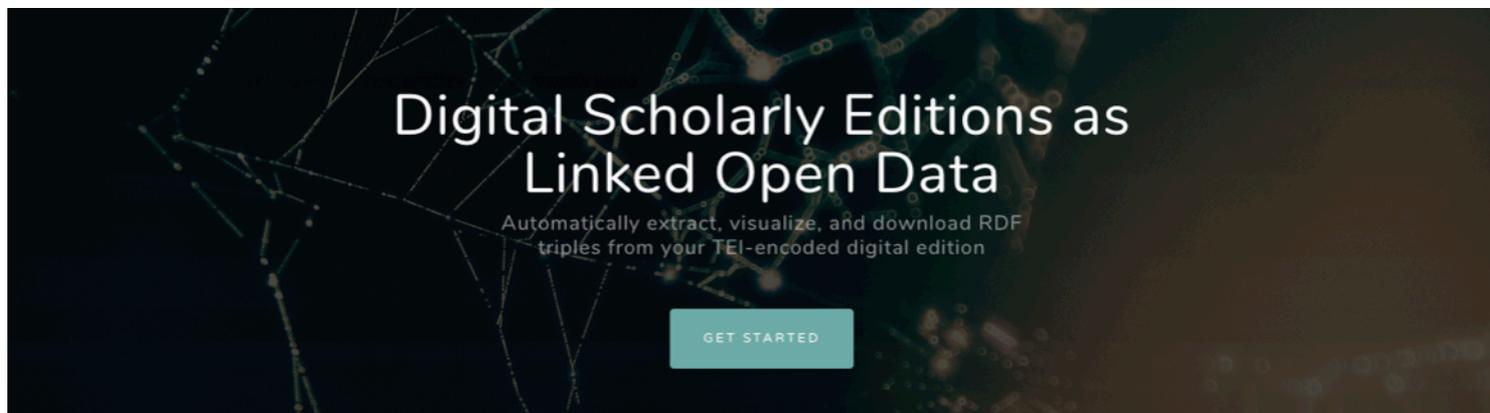
Daquino, M., Giovannetti, F., Tomasi, F. (2018) *Paolo Bufalini: Appunti 1981-1991*,  
<https://github.com/marilenadaquino/bufalinis-notebook>

Jordanous, A., Roueché, C., Tupman, C., Lawrence, K. F., Hedges, M., Wakelnig, E., & Searby, D. (2013). XSLTs created by SAWS. In *Sharing Ancient Wisdoms (SAWS)*. <http://www.ancientwisdoms.ac.uk/method/xslt/>

Giovannetti, F. (2018) DSEasLOD GitHub Repository, <https://github.com/fgiovannetti/dseaslod>

Peroni, S., Shotton, D. (2018). The SPAR Ontologies. In Proceedings of the 17th International Semantic Web Conference (ISWC 2018): 119-136. DOI: [https://doi.org/10.1007/978-3-030-00668-6\\_8](https://doi.org/10.1007/978-3-030-00668-6_8), <http://www.sparontologies.net/>

Web framework web.py. <http://webpy.org/>



Grazie per l'attenzione

Francesca Giovannetti

[francesc.giovannett5@studio.unibo.it](mailto:francesc.giovannett5@studio.unibo.it)

<https://www.linkedin.com/in/francesca-giovannetti-03489489/>

Quest'opera è stata rilasciata con licenza Creative Commons Attribuzione 4.0 Internazionale.  
Per leggere una copia della licenza visita il sito web <http://creativecommons.org/licenses/by/4.0>.