

FRANCESCA GIOVANNETTI

Consortium
GARR

THE ITALIAN
EDUCATION
& RESEARCH
NETWORK



Digital Scholarly Editions as Linked Open Data

GIORNATA DI INCONTRO
BORSE DI STUDIO GARR
"ORIO CARLINI"
ROMA

Roma 27/06/2019

Giornata di incontro Borse di Studio GARR "Orio Carlini"



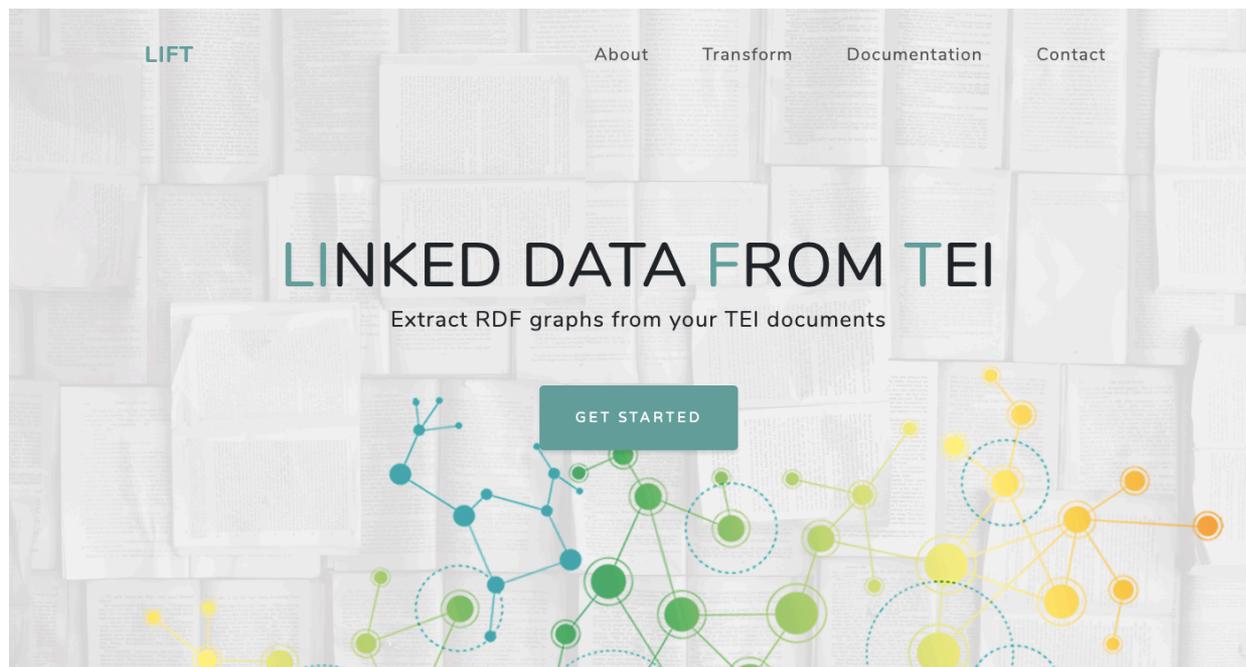
Contenuto della presentazione

- I. Tema della borsa di studio
- II. Scenario
- III. LIFT: descrizione
- IV. Critical Apparatus Ontology (CAO)
- V. LIFT: stato di avanzamento
- VI. Attività correlate



Tema della borsa di studio GARR

Sviluppo di **LIFT**,
un'applicazione web,
open source, per
l'estrazione di
knowledge graphs da
documenti **TEI/XML**



<http://github.com/fgiovannetti/lift>



Tema della borsa di studio GARR

Knowledge graph =

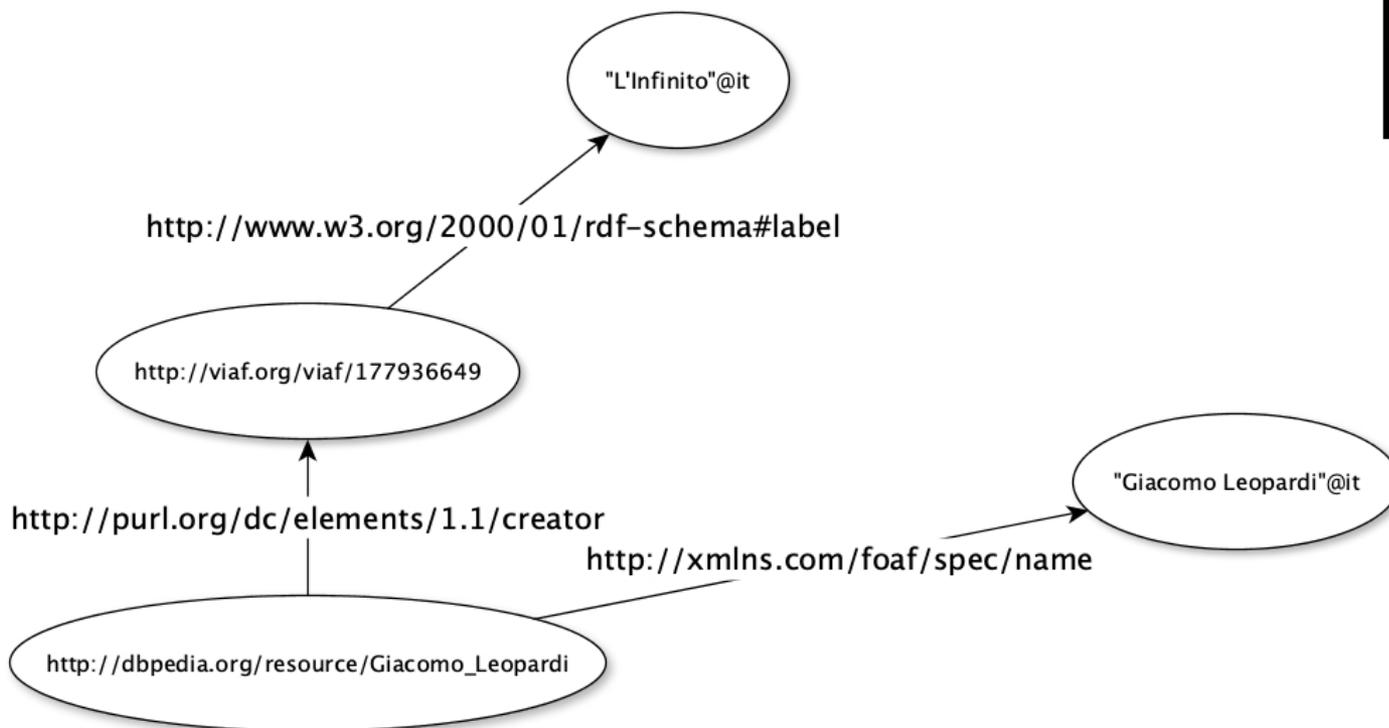
- grafo formato da entità, che rappresentano **concetti**, e link fra entità, che rappresentano **relazioni semantiche fra concetti**
- basato sul modello di dati **RDF** (Resource Description Framework)
- supportato da una o più **ontologie** (i.e. modelli di rappresentazione formale di un dominio di conoscenza) che ne determinano la semantica



Tema della borsa di studio GARR

Il data model RDF =

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





Tema della borsa di studio GARR

TEI (Text Encoding Initiative) =

- vocabolario XML per la codifica di testi umanistici

<https://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html>

```
<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
```



Tema della borsa di studio GARR

Scenario:

- **complessità intrinseca** legata alla produzione dell'edizione critica in TEI/XML
- 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 TEI/XML da entità semi-strutturata knowledge graph e dunque di ingresso nel **LOD cloud dei beni culturali**



Tema della borsa di studio GARR

Alcune triple RDF estratte dall'esempio TEI precedente tramite LIFT:

```
<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 .
```



LIFT (Linked data from TEI): descrizione

- Applicazione web sviluppata interamente con Python
- Modulare
 - possibilità di scegliere quali triple estrarre dall'edizione
 - **Persone**
 - **Luoghi**
 - **Eventi**
 - **Relazioni**
 - **Apparato critico**
 - **Riferimenti bibliografici**
 - possibilità di customizzare le trasformazioni
 - diverse serializzazioni (RDF/XML, N-Triples, N3...)
- **Valore didattico** (veicolato dalla documentazione, visualizzazione, Jupyter notebook)



LIFT (Linked data from TEI): descrizione

Principali ontologie riusate:

- Object-oriented Functional Requirements for Bibliographic Records (FRBRoo) <http://iflastandards.info/ns/fr/frbr/frbroo/>
- Publishing Roles Ontology (PRO) <http://purl.org/spar/pro/>
- Provenance Ontology (PROV) <http://www.w3.org/ns/prov#>
- Schema.org (SCHEMA) <https://schema.org/>
- Critical Apparatus Ontology (CAO) <http://purl.org/cao/>



Critical Apparatus Ontology (CAO)

- Permalink: <http://purl.org/cao/>
- Nuova ontologia per la rappresentazione delle **varianti** e delle **note filologiche** che formano l'apparato critico di un'edizione:
 - Utilizza il pattern n-ary relation per rappresentare clusters di varianti testuali
 - Consente di rappresentare una variante testuale sia come interpretazione che come entità testuale
 - Riutilizza il Web Annotation Data Model (OA) per l'introduzione di nuove voci di apparato
- Presentata presso:
 - *GraphSDE2019: Workshop on Scholarly Digital Editions, Graph Data-Models and Semantic Web Technologies*, Université de Lausanne, Switzerland, 3–4 Giugno 2019.
 - *ISWS International Semantic Web Research Summer School*, Bertinoro International Center for Informatics, 1-5 Luglio 2019.



LIFT (Linked data from TEI): avanzamento

21 Giugno – 15 Luglio:

1. Perfezionamento e pubblicazione documentazione
2. Implementazione dell'estrazione di apparato critico e riferimenti bibliografici nell'applicazione web
3. Aggiunta della serializzazione JSON
4. Aggiornamento Jupyter notebook a RDFLib

15 Luglio – :

Stesura di un articolo scientifico su LIFT



Attività correlate

- **Conferenza e pubblicazione**

Giovannetti, F. “The Critical Apparatus Ontology (CAO): Modelling the TEI's Critical Apparatus as a Knowledge Graph”. *GraphSDE2019: Workshop on Scholarly Digital Editions, Graph Data-Models and Semantic Web Technologies*, Université de Lausanne, Switzerland, 3–4 Giugno 2019, <https://wp.unil.ch/graphsde>.

Atti in pubblicazione in SIDE (Schriftenreihe des Instituts für Dokumentologie und Editorik).

- **Borsa di studio**

ISWS International Semantic Web Research Summer School, Bertinoro International Center for Informatics, 1-5 Luglio 2019.

- **Idoneità conseguita**

Dottorato in Computer Science and Engeneering, Ciclo XXXV, Università di Bologna. Borsa CNR a tema knowledge organization e beni culturali.



LINKED DATA FROM TEI
Extract RDF graphs from your TEI documents

GET STARTED

The banner features a background of overlapping document pages. In the foreground, there are several network graphs with nodes and connecting lines in shades of blue, green, and yellow. A prominent teal button with the text 'GET STARTED' is centered in the lower half of the banner.

Grazie per l'attenzione!

Francesca Giovannetti

francesc.giovannett5@studio.unibo.it

www.linkedin.com/in/f-giovannetti

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>