

Università degli Studi di Modena e Reggio Emilia
Ingegneria 'Enzo Ferrari'

Corso di Laurea Magistrale (D. M. 270/04) in Ingegneria Informatica

**Linked Open Data: pubblicazione, arricchimento
semantico e linking di dataset pubblici attraverso
il sistema MOMIS**

Relatore:

Prof.ssa Sonia Bergamaschi

Correlatore:

Dott. Ing. Serena Sorrentino

Candidato:

Elisa Fusari

Open Data

Open Science

Open Government

www.linkedscience.org

www.data.gov

www.data.gov.uk

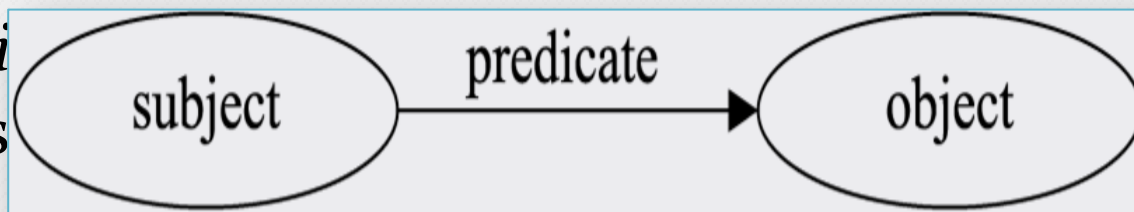
www.data.gov.it

Da Open Data a LINKED Open Data

Principi dei Linked Data :

- Usare *URI (Uniform Resource Identifier)*, per identificare le risorse
- Utilizzare il protocollo *HTTP (HyperText Transfer Protocol)*, per facilitare la ricerca e il recupero delle risorse
- Inserire risorse coperta di
- Utilizzare gli standard *W3C, **RDF** e SPARQL (Simple Protocol and RDF Query Language)*, per la pubblicazione e l'interrogazione delle risorse

Resource Description Framework



Da Open Data a LINKED Open Data

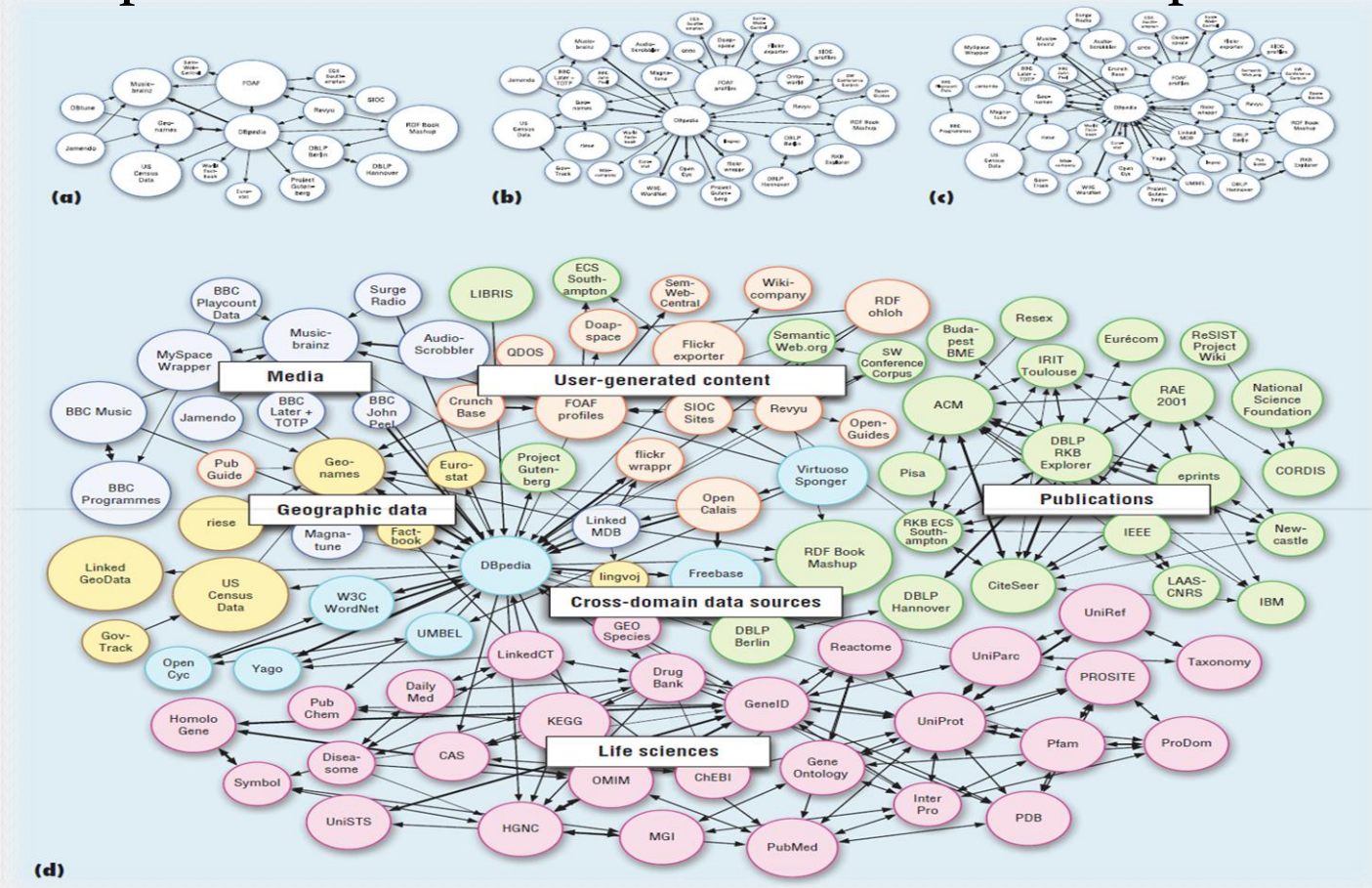
Classificazione “5 stelle”:

- 1 Stella = Dati nel Web e con Licenza Aperta
- 2 Stelle = Dati ‘machine-readable’
- 3 Stelle = Dati in formato NON-proprietario
- 4 Stelle = Dati in RDF
- 5 Stelle = Dati in RDF e con Link



LOD Project

Progetto Linking Open Data (LOD), attività del W3C (*World Wide Web Consortium*) iniziata nel 2007, che al momento comprende 295 dataset e 31.634.213.770 triple RDF.



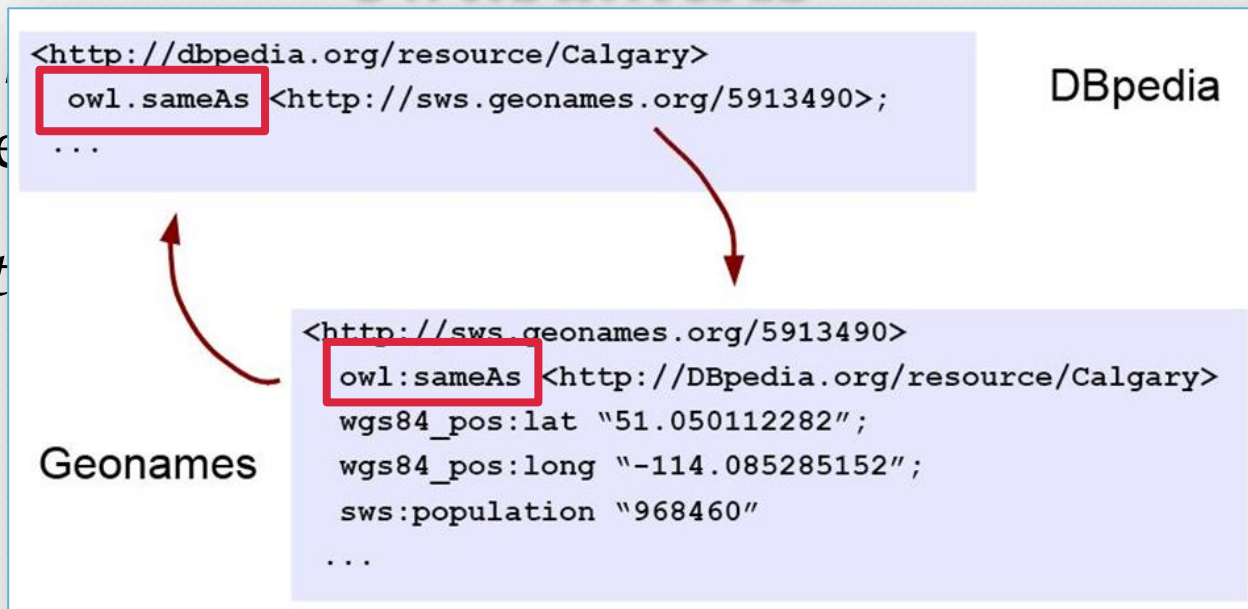
Semantica nei LOD

Tipi di RDF Link:

- *Relationship Link*: se legano fonti di dati diverse

owl:sameAs

- *Vocabolario*
dei termini
- *Identificazione*
fonti



oni
a loro
erse

Semantica nei LOD

Predicati mancanti:



Semantica nei LOD

Predicati mancanti:

owl:equivalentClass

owl:equivalentProperty

rdfs:subClassOf

rdfs:subPropertyOf

owl:AnnotationProperty

Semantica nei LOD

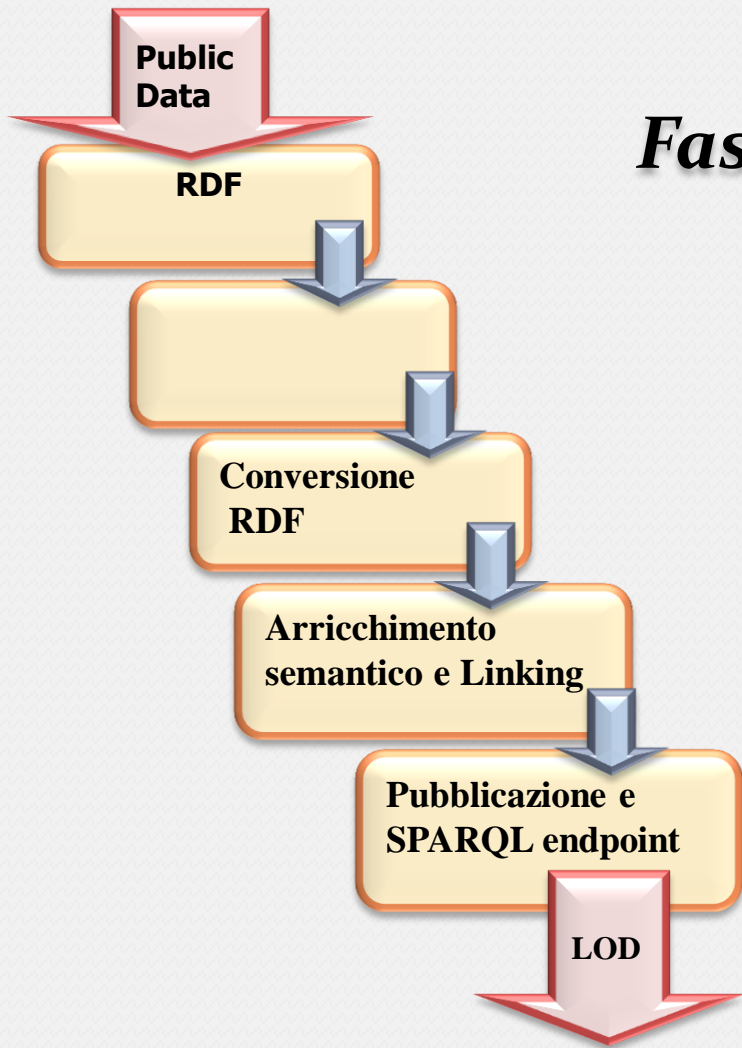
owl:AnnotationProperty



Annotazione Semantica:

Operazione che permette di esprimere in modo formale il significato delle risorse, rendendole contestualizzate e facilitandone l'integrazione.

Publicazione LOD



Fasi del processo:

- Trasformare i dati semplici in dati strutturati
- Convertire i dati in RDF
- Arricchire semanticamente i dati e creare link con altri dataset
- Rendere i dati pubblici ed interrogabili, attraverso punti d'accesso SPARQL

Progetto PRECARIETA' GIOVANILI

Raccolta dati utile allo studio della situazione di insicurezza e instabilità dei giovani, non solo in ambito lavorativo, ma anche formativo, familiare ed affettivo.

2. Sesso M F

3. Quali città di provenienza? Città: _____ Nazione: _____

4. Puoi dire di essere nato in un territorio? Sì No

5. Attualmente sei...:
 Studente (inclusi studenti lavoratori)
 Lavoratore assunto a tempo pieno
 Lavoratore a contratto
 Disoccupato
 In cerca di prima occupazione

6. Che cosa prevale (un'opzione per coppia)?

fiducia sfiducia

collaborazione competizione

presenza di regole libertà di agire liberamente

stabilità precarietà

difficoltà tranquillità

7. Che cosa prevale (un'opzione per coppia)?

fiducia sfiducia

collaborazione competizione

presenza di regole libertà di agire liberamente

stabilità precarietà

difficoltà tranquillità

8. Rispetto ad un anno fa, se pensi al tuo futuro che cosa vedi?

meno fiducia

9. Rispetto ad un anno fa, se pensi al tuo futuro che cosa vedi?

fiducia sfiducia

collaborazione competizione

presenza di regole libertà di agire liberamente

stabilità precarietà

difficoltà tranquillità

10. Riguardo a che cosa sei MENO fiducioso per il futuro?

scuola/lavoro

amore

amicizia

11. Riguardo a che cosa sei MENO fiducioso per il futuro?

scuola/lavoro

amore

amicizia



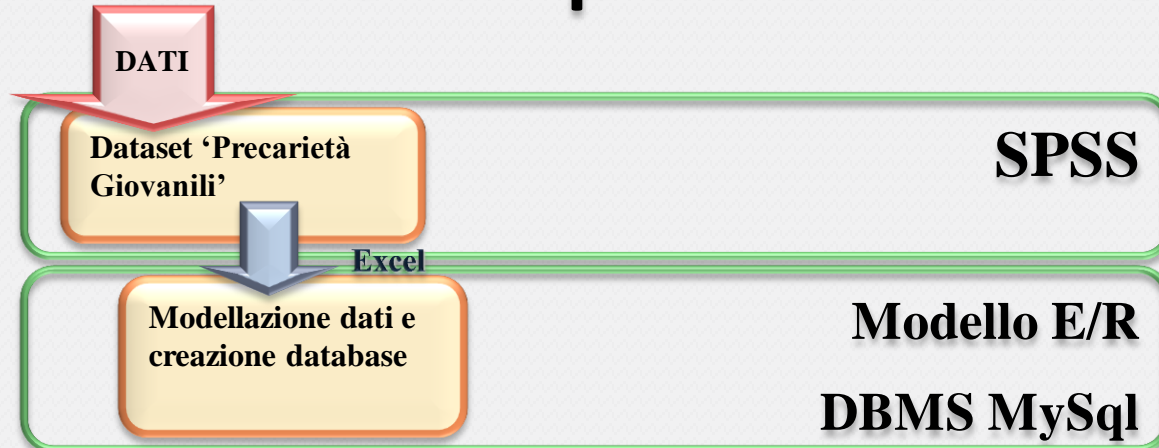
STRADANOVE



UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Dipartimento di Scienze del Linguaggio e della Cultura

Processo di pubblicazione



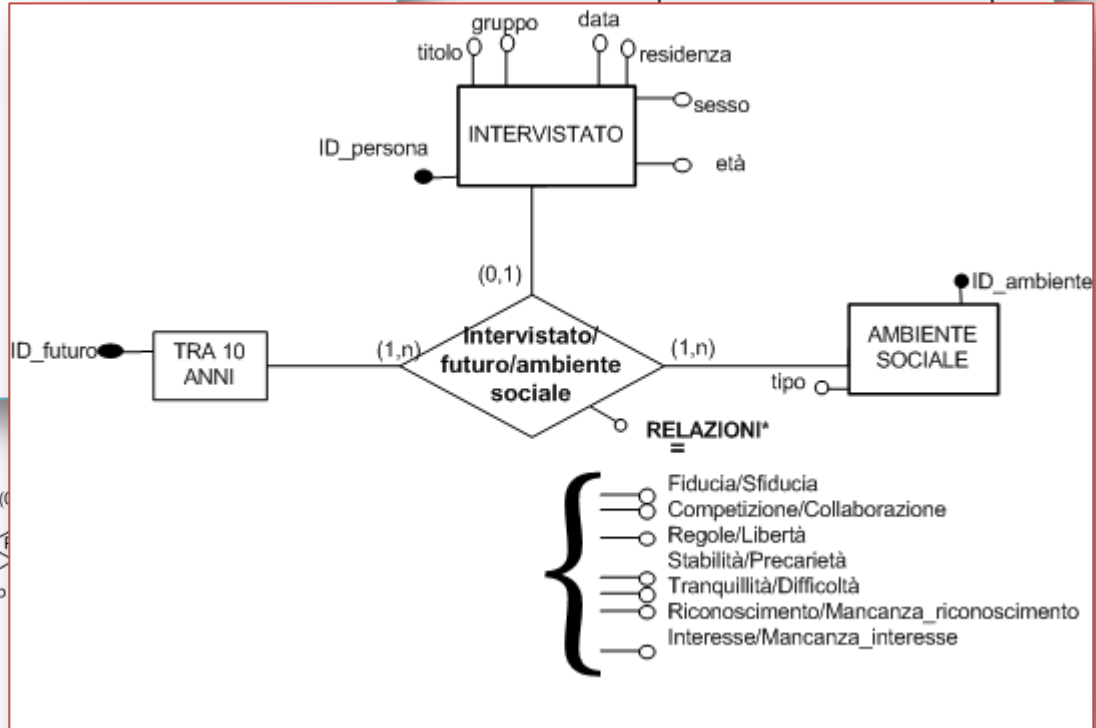
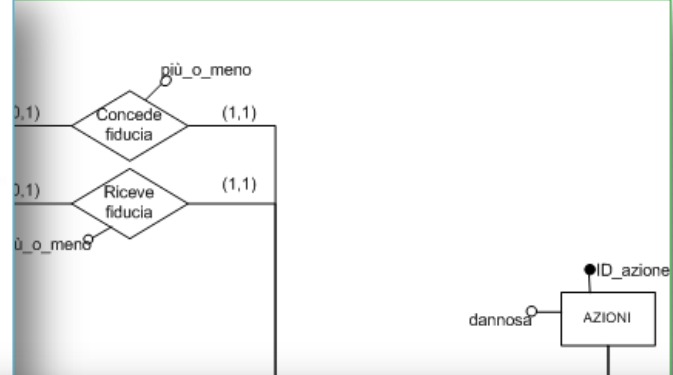
Statistical Package for the Social Sciences



Modellazione dati

6. Che cosa prevale quando pensi al tuo futuro nel prossimo anno (un'opzione per coppia)?

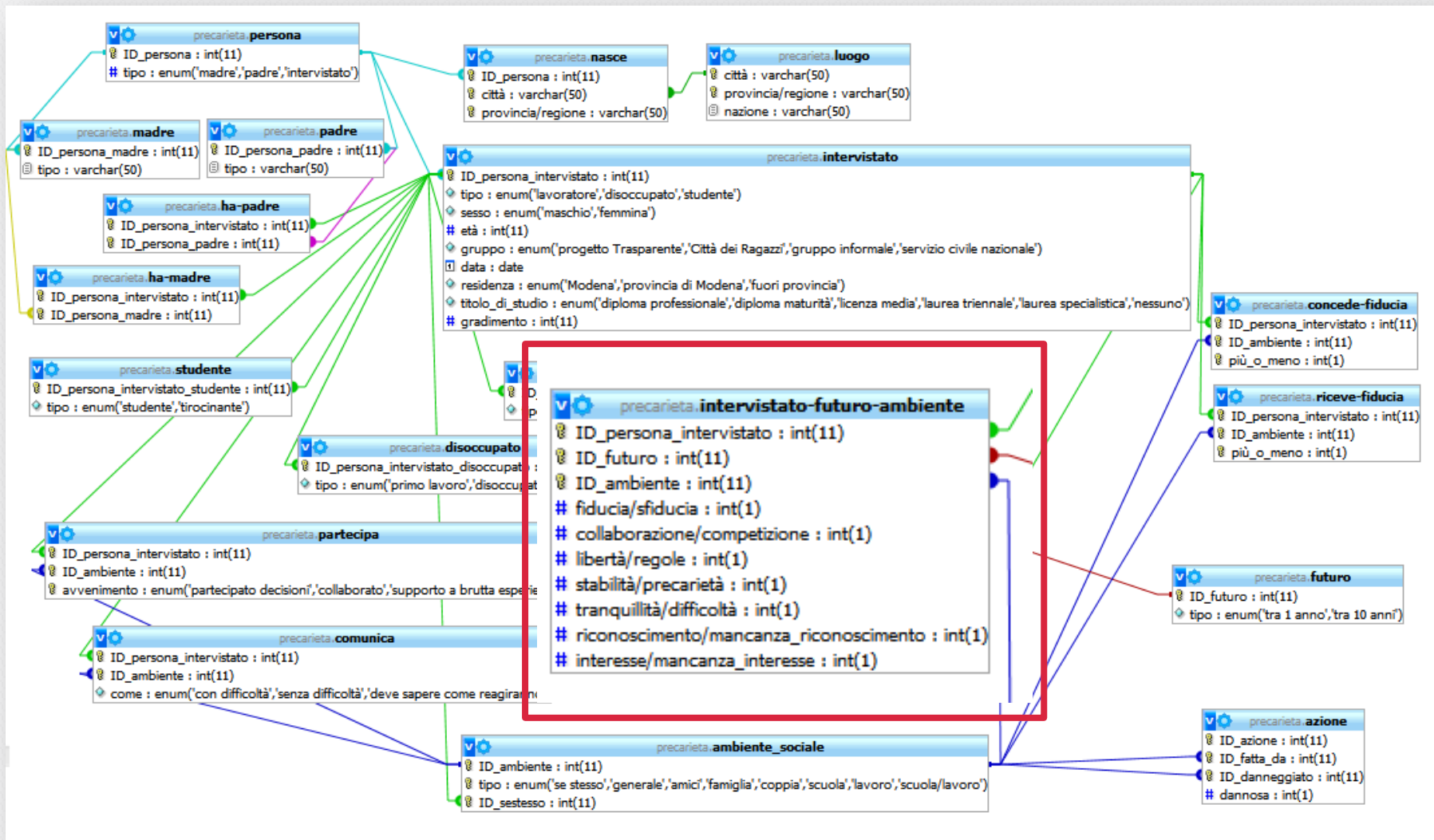
	In generale	Scuola/lavoro	Amici	Coppia	Famiglia
fiducia	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
sfiducia	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
collaborazione	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
competizione	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
presenza di regole	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
libertà di agire liberamente	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
stabilità	<input type="checkbox"/>				
precarietà	<input type="checkbox"/>				
difficoltà	<input type="checkbox"/>				
tranquillità	<input type="checkbox"/>				
riconoscimento delle mie capacità	<input type="checkbox"/>				
mancato riconoscimento delle mie capacità	<input type="checkbox"/>				
interesse	<input type="checkbox"/>				
manca di interesse	<input type="checkbox"/>				



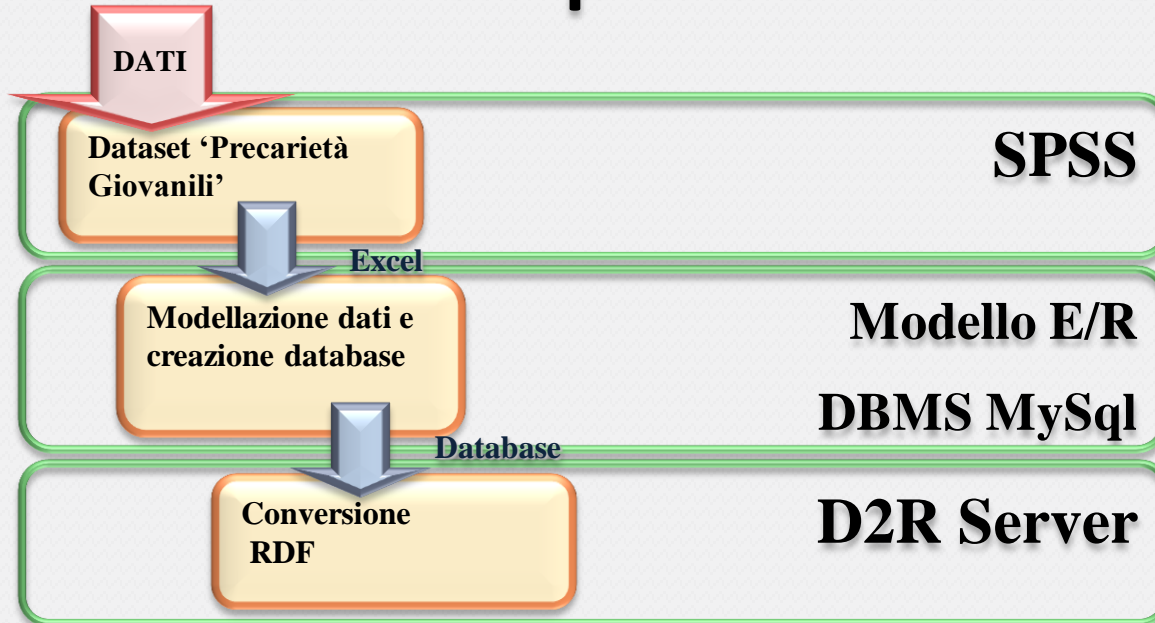
- RELAZIONI***
- Fiducia/Sfiducia
 - Competizione/Collaborazione
 - Regole/Libertà
 - Stabilità/Prearietà
 - Tranquillità/Difficoltà
 - Riconoscimento/Manca di riconoscimento
 - Interesse/Manca di interesse

- Fiducia/Sfiducia
- Competizione/Collaborazione
- Regole/Libertà
- Stabilità/Prearietà
- Tranquillità/Difficoltà
- Riconoscimento/Manca di riconoscimento
- Interesse/Manca di interesse

Creazione database

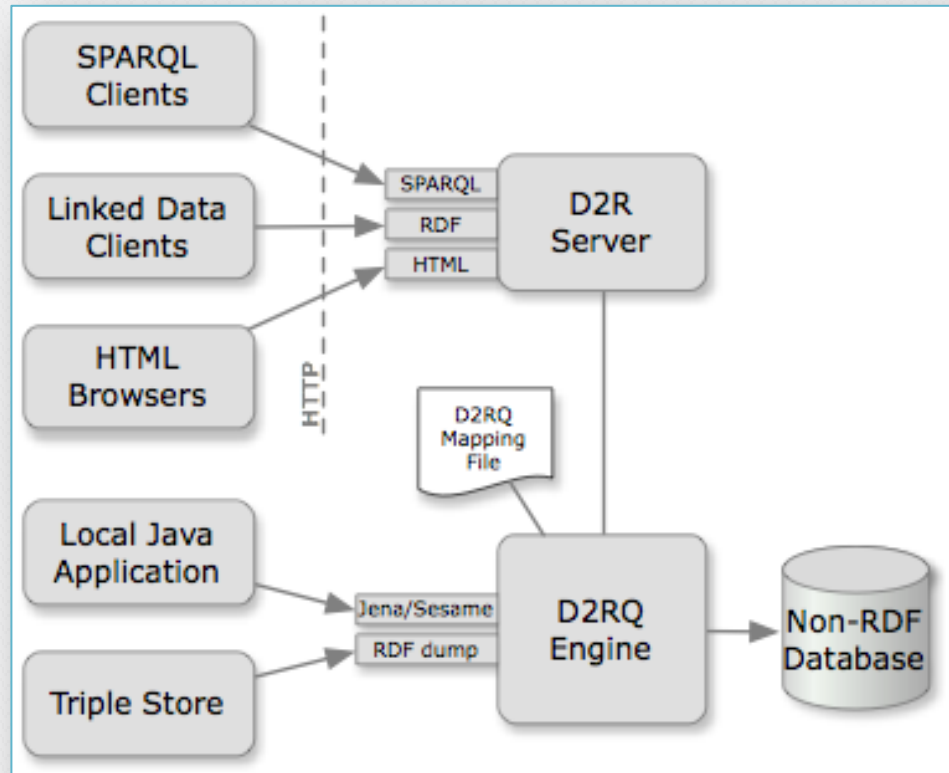


Processo di pubblicazione

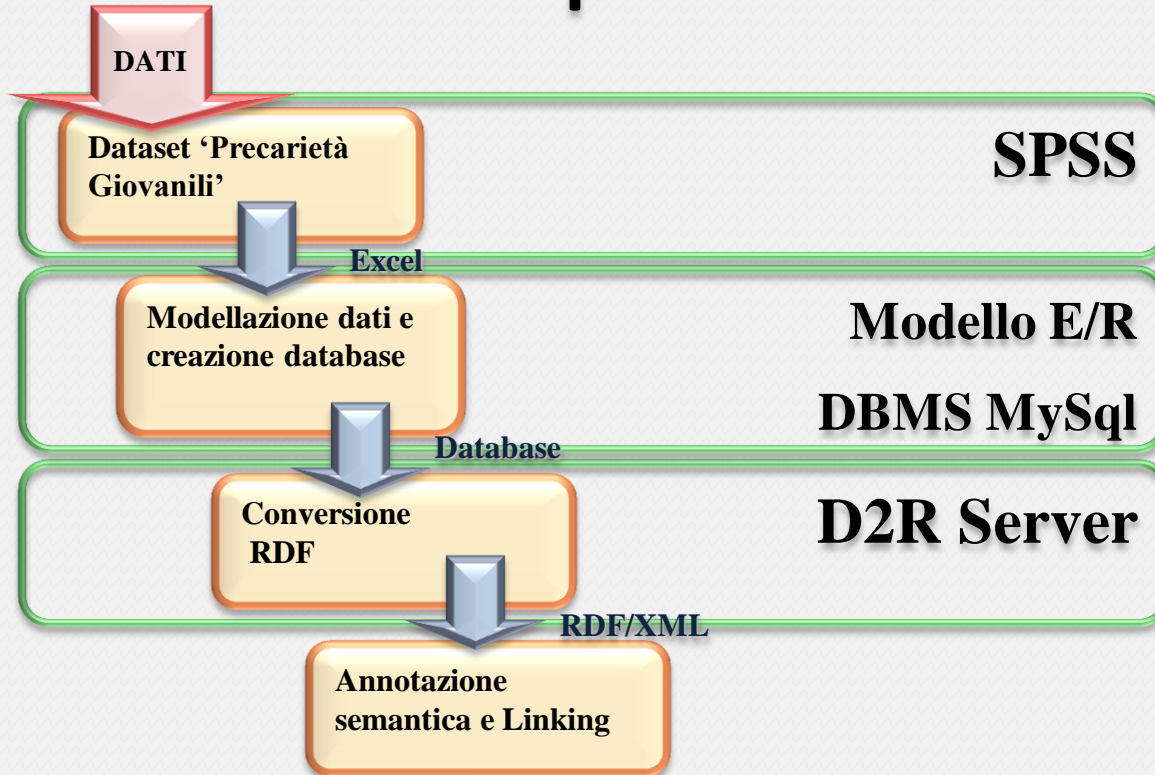


D2R Server

D2R (Database To RDF), è un server HTTP che permette la pubblicazione del contenuto dei database relazionali nel Semantic Web.

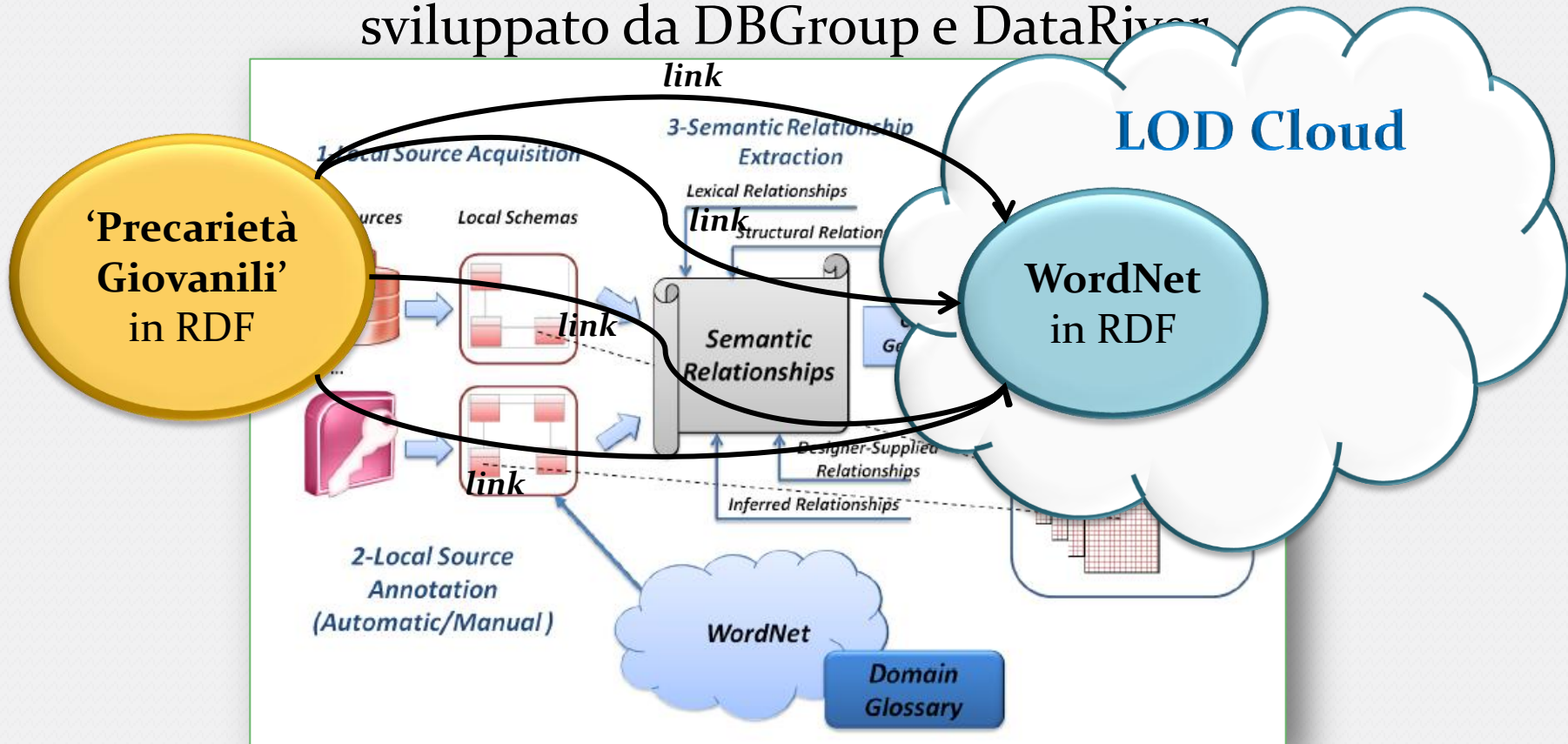


Processo di pubblicazione

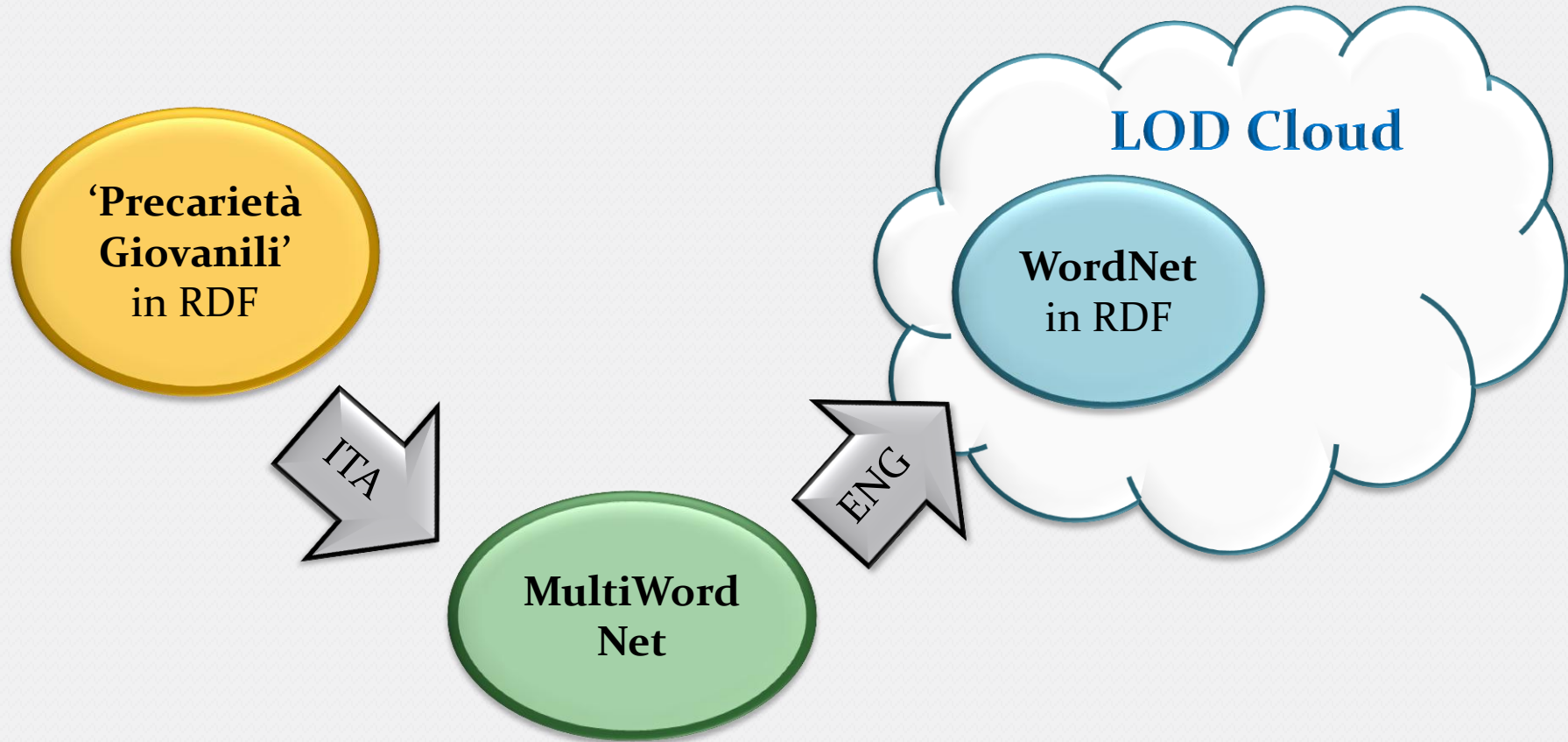


Annotazione semantica con WordNet

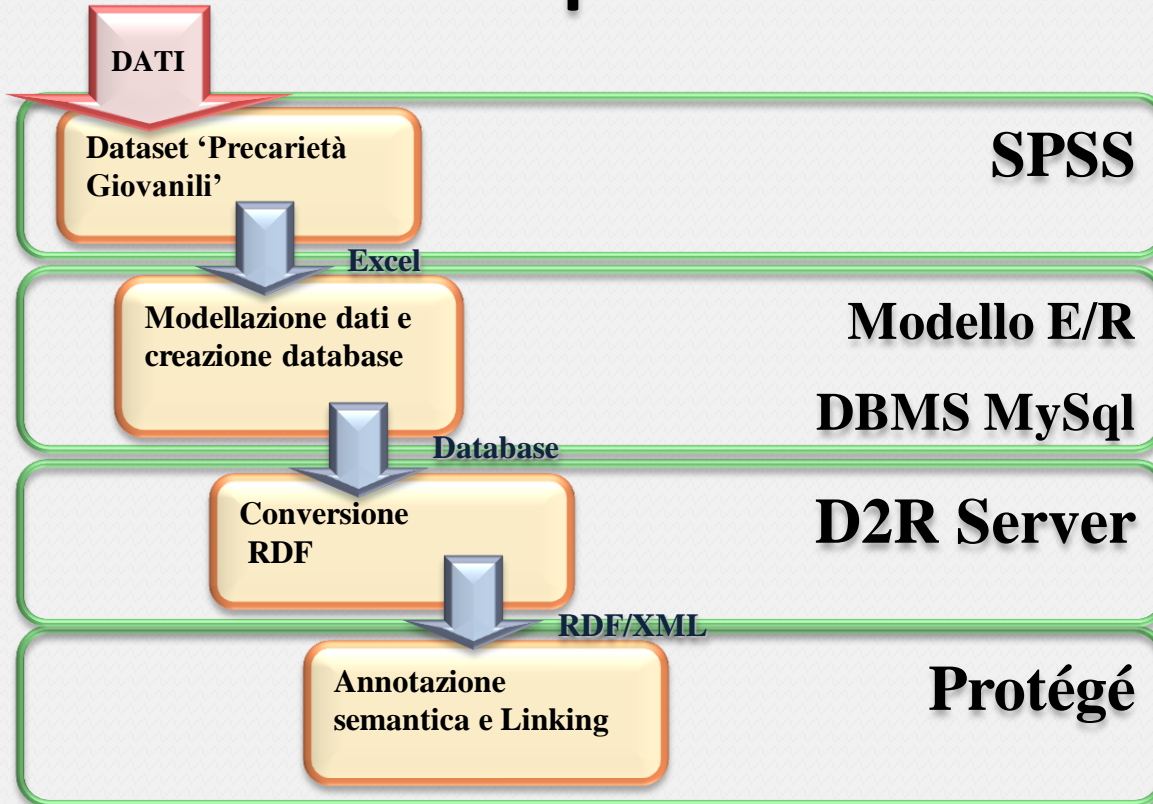
MOMIS (Mediator enviroNment for Multiple Information Sources), sistema per l'integrazioni di sorgenti dati eterogenee sviluppato da DBGroup e DataRiver



Annotazione semantica con WordNet

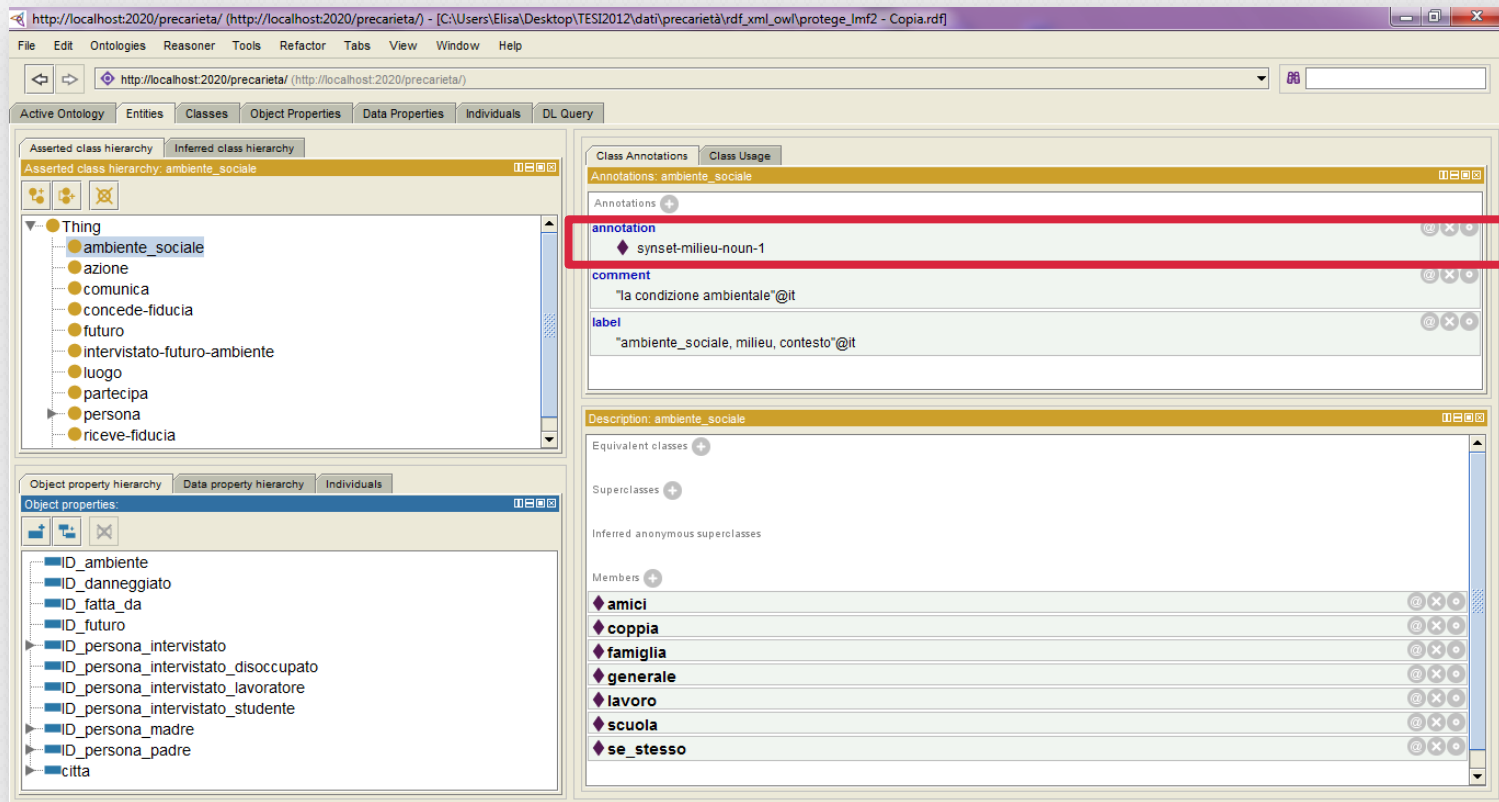


Processo di pubblicazione

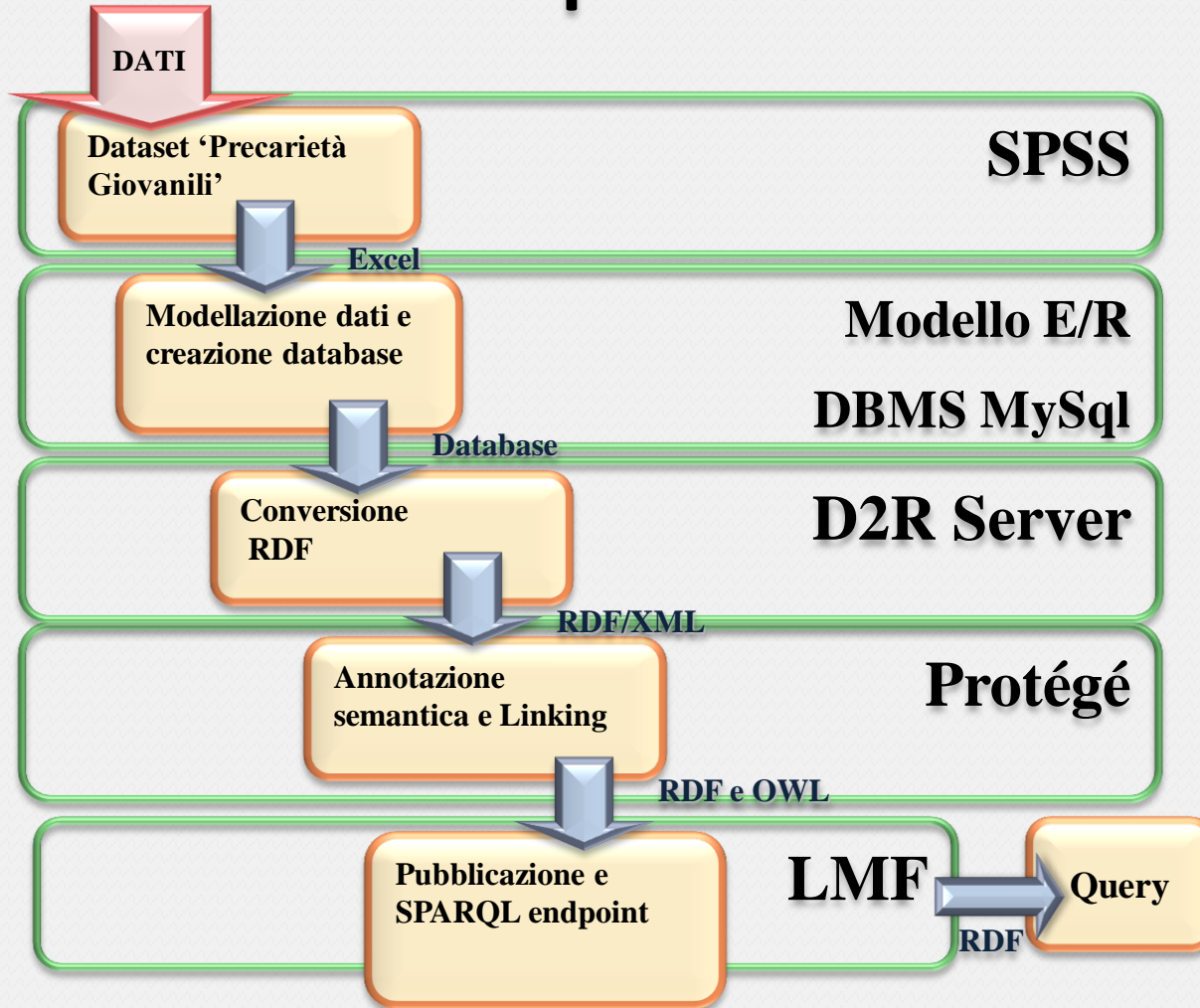


Protégé

Protégé, editor open-source di ontologie OWL (Ontology Web Language), sviluppato dal Centro di Biomedical Informatics Research all'Università di Medicina di Stanford.



Processo di pubblicazione



LMF – Linked Media Framework



SPARQL

LMF, è un'applicazione server che offre diversi servizi relativi al Web semantico. Si compone di un server centrale e diversi moduli, tra cui il servizio di interrogazione SPARQL, Snorql.

Endpoint: <http://elisa.virtualbox:8080/LMF/sparql/select>

SPARQL:

```
PREFIX rdfs: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX http://wordnet.rkbexplorer.com/id/synset-town-noun-1/
PREFIX http://www.w3.org/2006/03/wn/wn20/schema/
PREFIX http://www.w3.org/2006/03/wn/wn20/instances/
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
```

Browse:

```
1 SELECT DISTINCT ?property ?hasValue ?isValueOf
2 WHERE {
3   { <http://wordnet.rkbexplorer.com/id/synset-town-noun-1> ?property ?hasValue }
4   UNION
5   { ?isValueOf ?property <http://wordnet.rkbexplorer.com/id/synset-town-noun-1> }
6 }
7 ORDER BY (!BOUND(?hasValue)) ?property ?hasValue ?isValueOf
```

Results:



Linked Media Framework

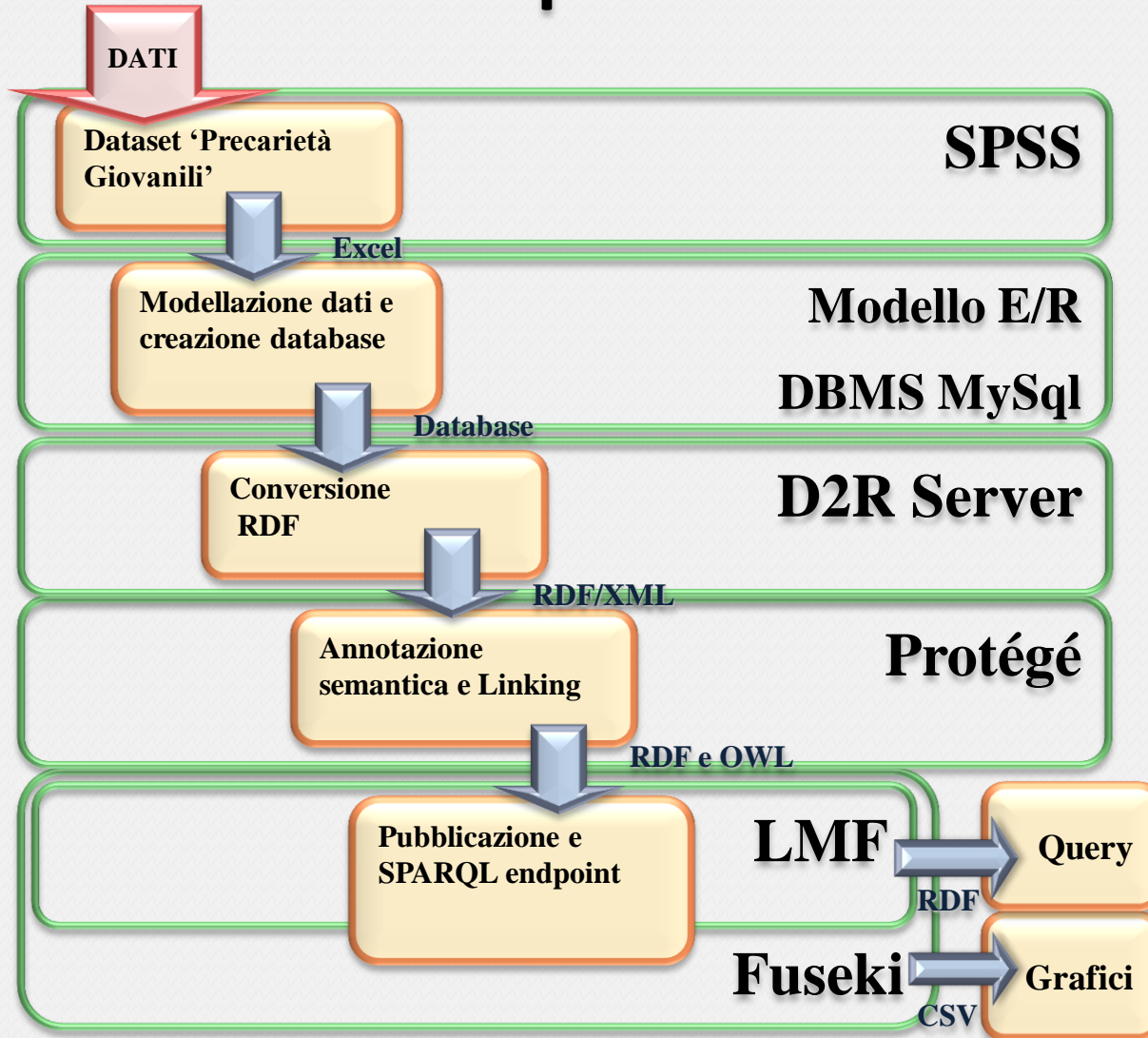
Imf

Linked Media Framework

Description of <http://wordnet.rkbexplorer.com/id/synset-town-noun-1/>

property	hasValue	isValueOf
dc:creator	"ReSIST Project"	-
dc:source	<http://wordnet.rkbexplorer.com/id/synset-town-noun-1>	-
<http://purl.org/dc/terms/created>	"2012-10-30T11:10:13.000+0000" ^{^^xsd:dateTime}	-
rdf:type	rdfs:Resource	-
rdf:type	owl:Ontology	-
rdf:type	owl:Thing	-
rdf:type	<http://www.w3.org/2006/03/wn/wn20/schema/NounSynset>	-
rdfs:label	"town"	-
owl:sameAs	<http://www.w3.org/2006/03/wn/wn20/instances/synset-town-noun-1>	-
<http://www.w3.org/2006/03/wn/wn20/schema/containsWordSense>	<http://wordnet.rkbexplorer.com/id/wordsense-town-noun-1>	-
<http://www.w3.org/2006/03/wn/wn20/schema/gloss>	"(an urban area with a fixed boundary that is smaller than a city; "they drive through town on their way to work")"	-
<http://www.w3.org/2006/03/wn/wn20/schema/hyponymOf>	<http://wordnet.rkbexplorer.com/id/synset-municipality-noun-1>	-
<http://www.w3.org/2006/03/wn/wn20/schema/synsetId>	"108135026"	-
<http://localhost:2020/vocab/resource/annotation>	-	<http://localhost:2020/vocab/resource/citta>
dc:source	-	<http://wordnet.rkbexplorer.com/id/synset-town-noun-1>

Processo di pubblicazione



Fuseki

Fuseki, è un SPARQL server che fornisce dati RDF sul protocollo HTTP, sviluppato nel progetto Apache Jena. Esso supporta l'utilizzo di operatori di raggruppamento (*GROUP BY*) e funzioni di conteggio (*COUNT()*) nel protocollo SPARQL.

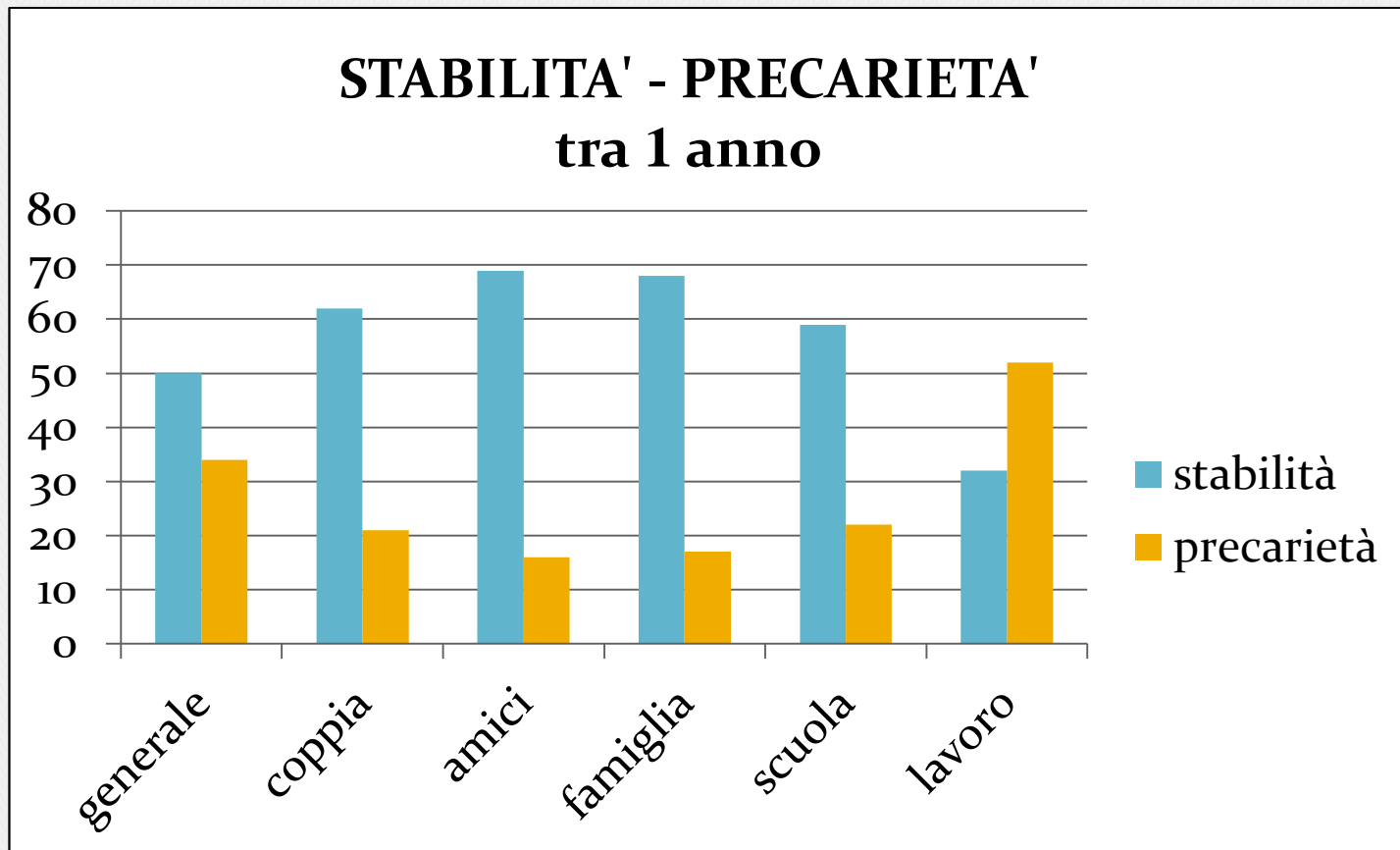
“Quanti intervistati hanno indicato stabilità tra un anno, nei diversi ambienti sociali? E quanti precarietà?”

The screenshot shows the Fuseki Query interface. At the top, it says "Fuseki Query" and "Dataset: rstata". Below that, there is a "SPARQL Query" section with a text area containing the following query:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
SELECT ?s ?ambiente ?valore
WHERE {
  ?s rdf:type <http://localhost:2020/vocab/resource/ID_ambiente>.
  ?s <http://localhost:2020/vocab/resource/ID_ambiente>
  <http://localhost:2020/precarieta/futuro/2>.
  ?s <http://localhost:2020/vocab/resource/ID_ambiente>
  ?ambiente.
  ?s <http://localhost:2020/vocab/resource/stabilita-precarieta> ?valore }
GROUP BY ?ambiente ?valore
Output: CSV
```

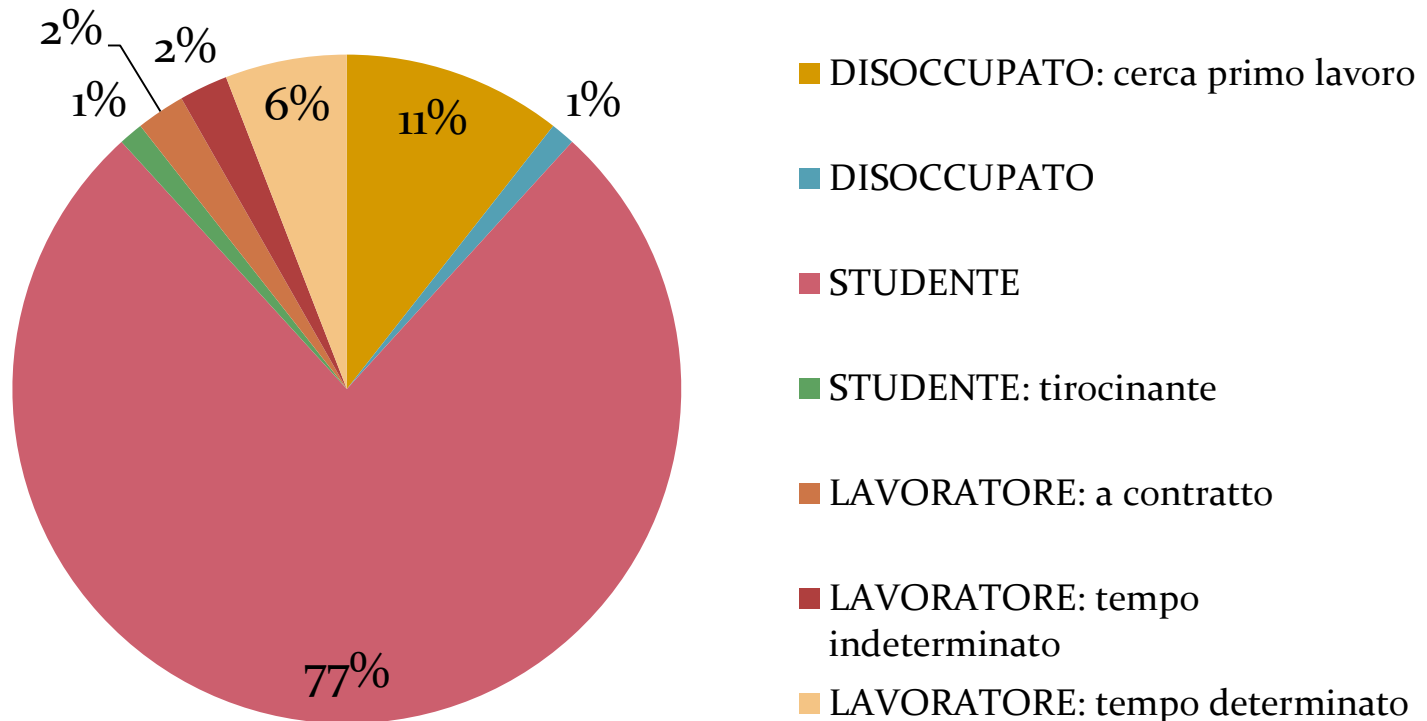
Below the query, there is a "Get Results" button. Underneath, there is a "SPARQL Update" section with a text area and a "Perform update" button. At the bottom, there is a "File upload" section with a "File:" input field, a "Sfoggia..." button, a "Graph:" dropdown menu set to "default", and an "Upload" button.

Risultati query

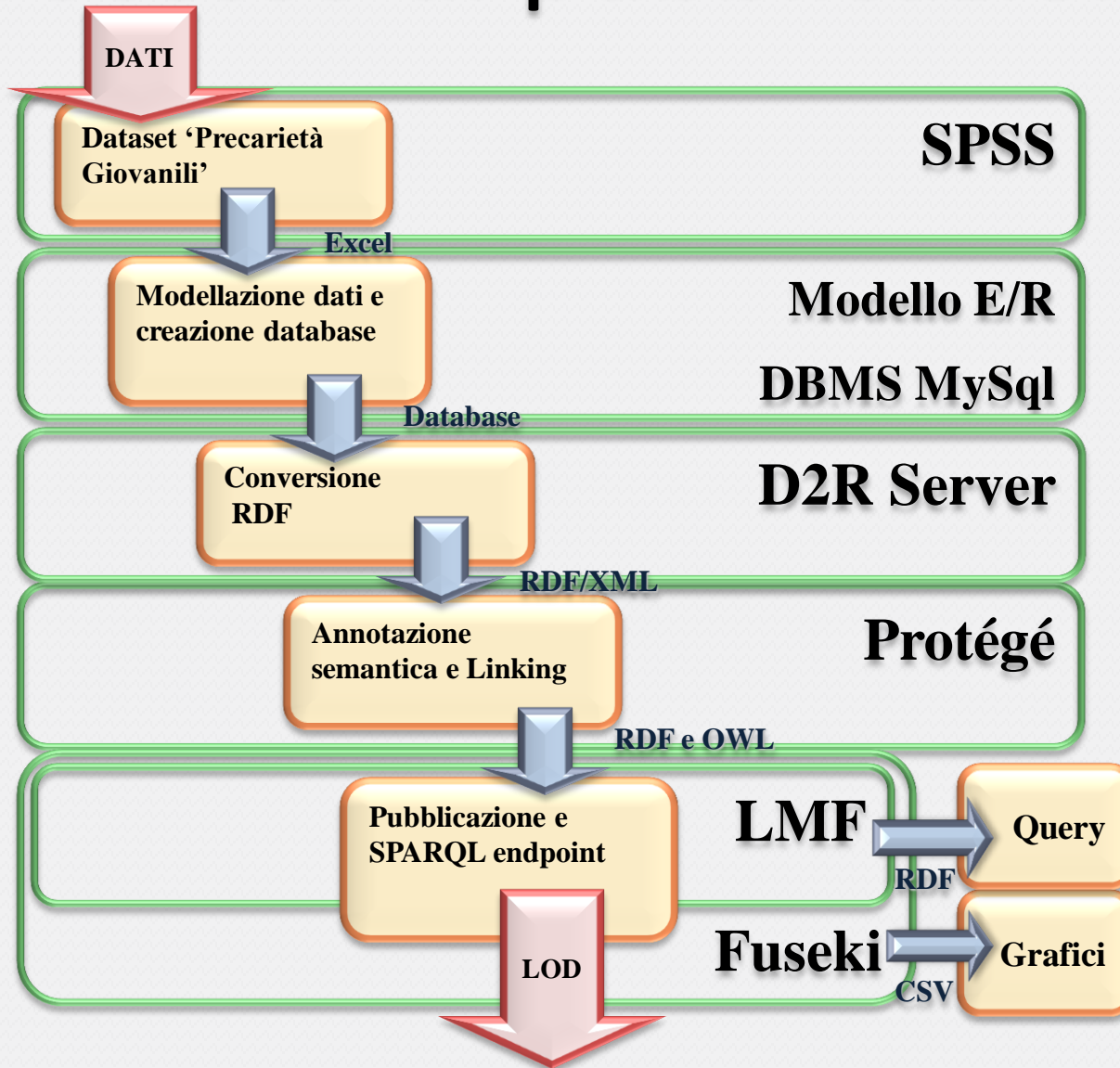


Altri Risultati

Intervistati - Stato occupazionale



Processo di pubblicazione



Conclusioni

- Panoramica sugli Open Data, sui Linked Data e sul progetto LOD
- Panoramica sullo sugli strumenti a disposizione per l'elaborazione dei Linked Open Data
- Sviluppo di una metodologia generale per la pubblicazione dei Linked Open Data
 - Applicazione della metodologia ad un caso concreto
- Inserimento di un dataset pubblico nella LOD Cloud: ***il dataset è ora confrontabile ed incrociabile con altri miliardi di dati !***

Sviluppi Futuri

- Automatizzare le fasi del processo di pubblicazione dei Linked Open Data
- Estendere il sistema MOMIS per l'utilizzo nella sfera dei Linked Open Data:
 - Creazione di un wrapper per dati RDF
 - Integrazione del thesaurus MultiWordNet per l'annotazione di sorgenti in italiano
 - Implementare l'esportazione in RDF dei dati integrati

Graxie per l'attenzione