



UNIVERSITÀ  
DEGLI STUDI  
FIRENZE

# RELAZIONE DATA WAREHOUSING

Di

Lavinia De Divitiis

## Sommario

Scopo dell'elaborato.....	3
Strumenti usati .....	3
Realizzazione .....	4
Diagramma ER .....	4
Ontologia con Protégé.....	5
Numero di scuole per città, in questo caso Firenze .....	7
Numero di scuole per regione, in questo caso Toscana.....	7
Numero di scuole per Zona, in questo caso, Centro .....	8
Numero classi per Firenze .....	8
Numero classi per Toscana.....	9
Numero classi per Centro .....	9
Collegare Linked Open Data e Tabella Istat.....	10
Utilizzo di CONTRUCT .....	12
Ulteriore utilizzo di CONSTRUCT .....	13

## Scopo dell'elaborato

Lo scopo di questo elaborato è stato quello di permettere di raggiungere dati forniti da Istat, ma di cui Istat non fornisce il modo di ricavarlo tramite Linked Open Data (LOD).

In questo caso, ho tenuto in considerazione la tabella delle scuole di infanzia, presente sul sito ISTAT, che si può trovare navigando attraverso la categoria “Istruzione e Formazione” -> “Scuole” -> “Scuola dell'infanzia” -> “Scuole, classi e bambini”.

## Strumenti usati

- **Protege:** *“Protégé è una piattaforma gratuita e open-source che fornisce a una comunità di utenti in crescita una suite di strumenti per costruire modelli di dominio e applicazioni basate sulla conoscenza con ontologie.”* <https://protege.stanford.edu/>
- **Draw.io:** tool online per la realizzazione di modelli. È stato usato per realizzare il modello Entità Relazione (Modello ER)
- **Dati Istat** (<http://dati.istat.it/#>)
- **Linked Open Data Istat** (<http://datiopen.istat.it/sparql/istat.php>)

## Realizzazione

### Diagramma ER

Come prima cosa ho realizzato un modello Entità Relazione, come si vede in *Figura 1*.

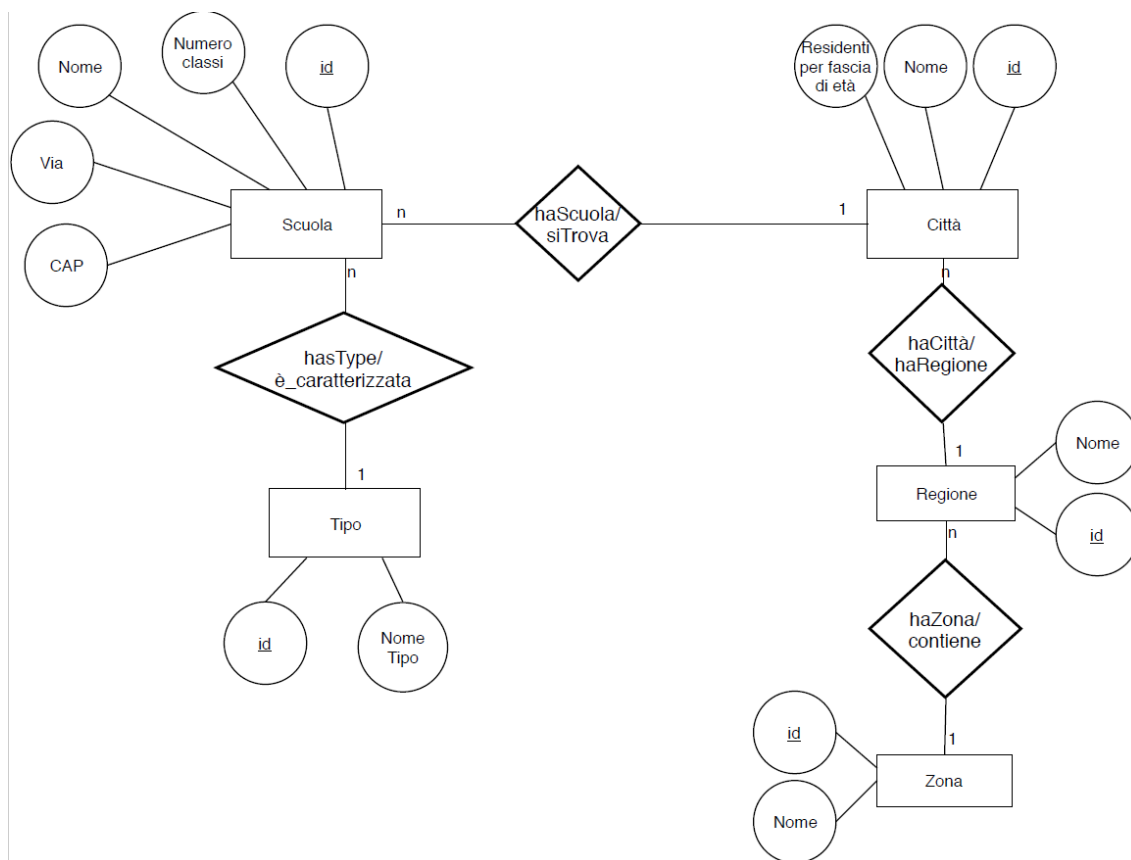


Figura 1:Modello ER

Ho realizzato questo modello basandomi sulla tabella presente nei dati di Istat relativa alla “Scuola dell’infanzia”, vedi *Figura 2*.

Esplora Temi

Cerca nei temi  Annulla

Tutti i temi

Censimento agricoltura 2010

Censimento industria, istituzioni pubbliche e non profit 2011, 2015

Censimento popolazione e abitazioni 2011

Ambiente ed energia

Caratteristiche del territorio

Popolazione e famiglie

Condizioni economiche delle famiglie e disuguaglianze

Salute e sanità

Assistenza e previdenza

Istruzione e formazione

Scuole

Scuola dell'infanzia

Scuole, classi e bambini

Insegnanti delle scuole statali

Scuola primaria

Scuola secondaria di primo grado

Scuola secondaria di secondo grado

Scuola dell'infanzia

Ordine scolastico: infanzia

Tipo dato: numero di scuole

Misura: valori assoluti

Seleziona anno (anno 1° anno sc. t-1 / t)

Gestione della scuola

	pubblica	statale	pubblica non statale pubblica	privata	totale
<b>Territorio</b>					
Italia	15 378	13 424	1 954	8 137	23 515
Nord-ovest	3 280	2 684	596	2 129	5 409
Piemonte	1 160	1 067	93	493	1 653
Valle d'Aosta / Vallée d'Aoste	86	–	86	8	94
Liguria	384	311	73	188	572
Lombardia	1 650	1 306	344	1 440	3 090
Nord-est	2 088	1 614	474	1 965	4 053
Trentino Alto Adige / Südtirol	118	–	118	158	276
Provincia Autonoma Bolzano / Bozen	–	–	–	–	–
Provincia Autonoma Trento	118	–	118	158	276
Veneto	677	593	84	1 082	1 759
Friuli-Venezia Giulia	334	300	34	146	480

Figura 2: Tabella Istat

La tabella mostra il numero di scuole per zona, regione e città, e anche il numero di classi, sempre per zona, regione e città.

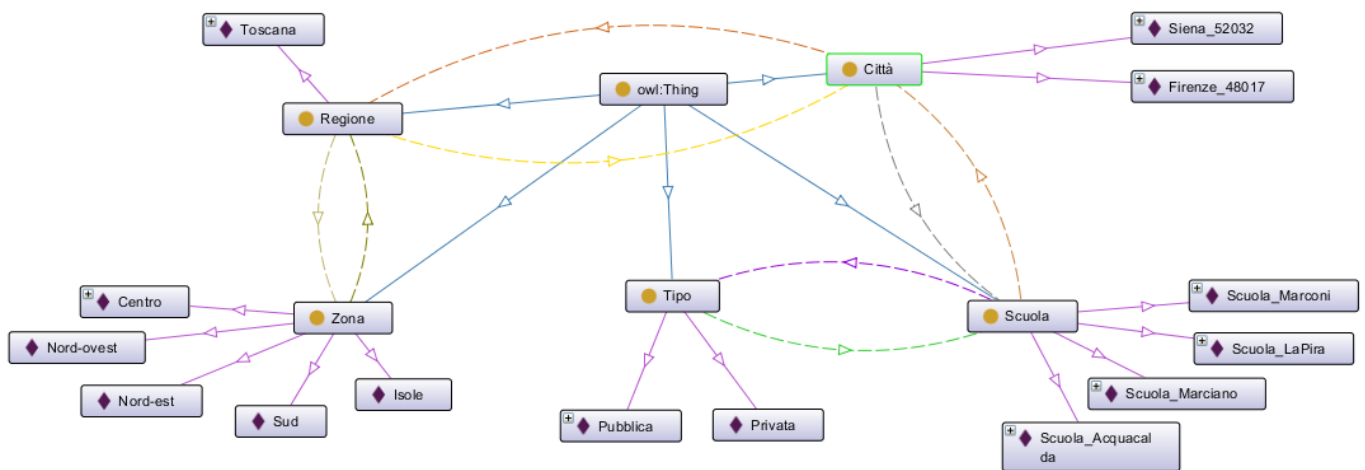
I dati presenti nella tabella di Istat (*Figura 2*), non sono attualmente reperibili attraverso i Linked Open Data di Istat.

Perciò il mio primo obiettivo è stato quello di creare una ontologia basandomi sulla *Figura 2*.

### Ontologia con Protégé

Per costruire l'ontologia ho utilizzato il software Protégé.

Nella *Figura 3* è mostrata la struttura della mia ontologia e in basso (*Figura 4*) il significato degli archi.



*Figura 3: Ontologia*

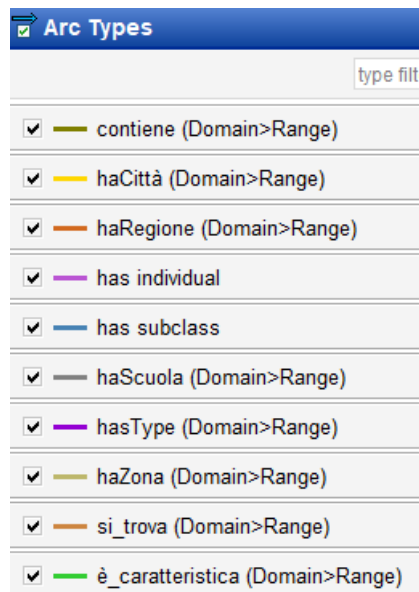
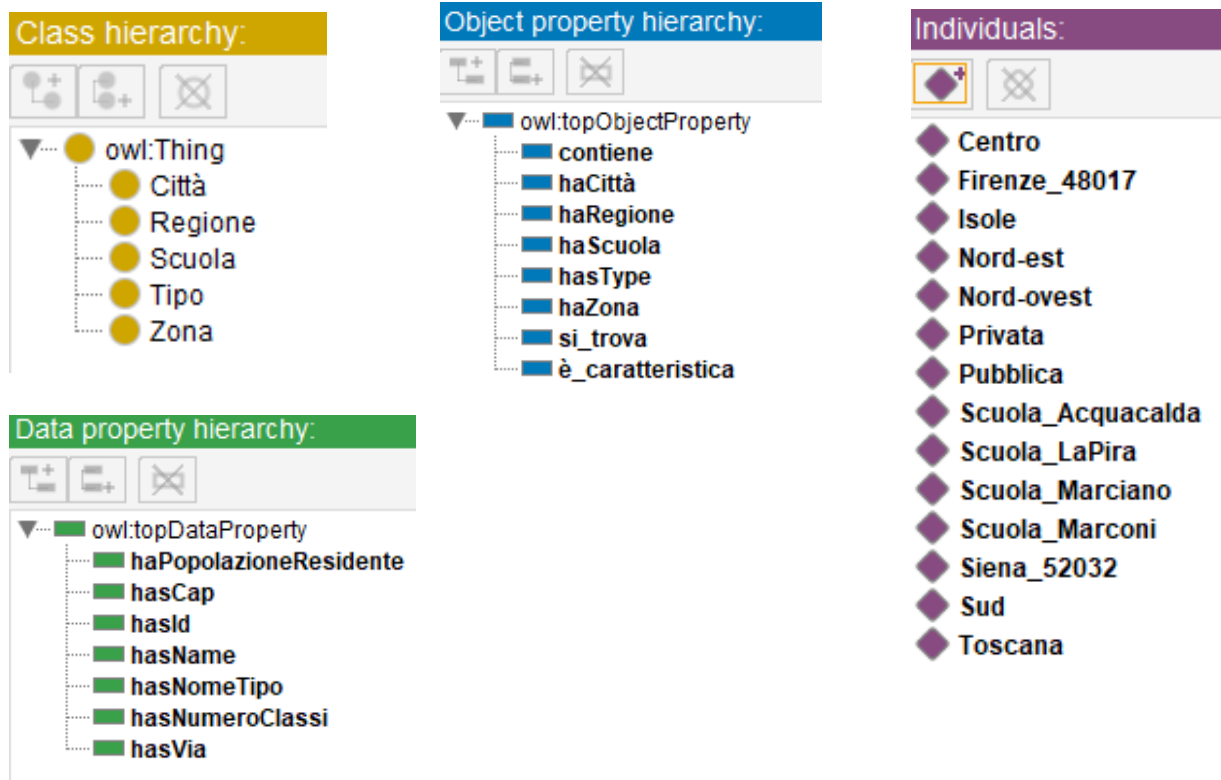


Figura 4: Arc Types

Inoltre, in basso riporto gli screenshot da Protégé per mostrare come ho strutturato le classi, le data properties, le object properties e gli individual nella mia ontologia:



Nello specifico degli individual, per Toscana, Siena e Firenze ho utilizzato gli IRI di Linked Open Data.

Dopo aver definito le classi e le relazioni tra di esse, ho realizzato le query che permettono, attraverso la mia ontologia, di trovare i dati presenti nella tabella Istat (Figura 2). Le query permettono di trovare:

- Il numero di scuole per:
  - Città
  - Regione
  - Zona
- Il numero di classi per:
  - Città
  - Regione
  - Zona

Queste query sono state eseguite all'interno di Protégé.

Numero di scuole per città, in questo caso Firenze

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX reg: <http://datiopen.istat.it/odi/risorsa/territorio/regioni/>
PREFIX sw:<http://www.semanticweb.org/lavinia/ontologies/2019/6/untitled-ontology-7#>
PREFIX cit:<http://datiopen.istat.it/odi/risorsa/territorio/comuni/>
SELECT (count(*) as ?numeroScuole)
WHERE {
cit:Firenze_48017 sw:haScuola ?scuola
}

```

Numero di scuole per regione, in questo caso Toscana

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX reg: <http://datiopen.istat.it/odi/risorsa/territorio/regioni/>
PREFIX sw:<http://www.semanticweb.org/lavinia/ontologies/2019/6/untitled-ontology-7#>
PREFIX cit:<http://datiopen.istat.it/odi/risorsa/territorio/comuni/>
SELECT (count(*) as ?numeroScuole)

```

```

WHERE {
  reg:Toscana sw:haCittà ?citta.
  ?citta sw:haScuola ?scuola.
}

```

Numero di scuole per Zona, in questo caso, Centro

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX reg: <http://datiopen.istat.it/odi/risorsa/territorio/regioni/>
PREFIX sw: <http://www.semanticweb.org/lavinia/ontologies/2019/6/untitled-ontology-7#>
PREFIX cit: <http://datiopen.istat.it/odi/risorsa/territorio/comuni/>
PREFIX cen: <http://datiopen.istat.it/odi/ontologia/censimento/>
SELECT (count(*) as ?numeroScuole)
WHERE
{
  sw:Centro sw:contiene reg:Toscana.
  reg:Toscana sw:haCittà ?città.
  ?città sw:haScuola ?scuola
}

```

Dopodichè ho cercato il numero di classi:

Numero classi per Firenze

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

```



PREFIX reg: <<http://datiopen.istat.it/odi/risorsa/territorio/regioni/>>

PREFIX sw:<<http://www.semanticweb.org/lavinia/ontologies/2019/6/untitled-ontology-7#>>

PREFIX cit:<<http://datiopen.istat.it/odi/risorsa/territorio/comuni/>>

SELECT (SUM(?nclassi) AS ?numeroClassi)

WHERE {

cit:Firenze\_48017 sw:haScuola ?scuola.

?scuola sw:hasNumeroClassi ?nclassi.

}

Numero classi per Toscana

SELECT (SUM(?nclassi) AS ?numeroClassi)

WHERE {

reg:Toscana sw:haCittà ?citta.

?citta sw:haScuola ?scuola.

?scuola sw:hasNumeroClassi ?nclassi.

}

Numero classi per Centro

SELECT (SUM(?nclassi) AS ?numeroClassi)

WHERE {

sw:Centro sw:contiene reg:Toscana.

reg:Toscana sw:haCittà ?citta.

?citta sw:haScuola ?scuola.

?scuola sw:hasNumeroClassi ?nclassi.

}

Se la mia ontologia venisse unita ai Linked Open Data di Istat, i dati della tabella Istat (*Figura 2*) che attualmente non si trovano nei LOD, entrerebbero a far parte dei LOD e potrebbero essere reperibili attraverso queste query.

## Collegare Linked Open Data e Tabella Istat

A questo punto, vogliamo riuscire a fare un paragone tra il numero di scuole, il cui dato è presente solo nella tabella Istat e la popolazione residente, il cui dato è presente nei LOD.

Per trovare la popolazione residente, questo dato è reperibile solo tramite i LOD di Istat. Per questo per aggiungere questo dato alla mia ontologia è stato necessario eseguire diverse query manuali per ottenere dai LOD il dato richiesto.

Date le URI dei LOD per Firenze, Siena e Toscana, ho trovato nei LOD la popolazione residente per Firenze e Siena e l'ho inserito a mano nella mia ontologia.

Di seguito riporto le query per trovare la popolazione residente a Firenze:

PREFIX rdf: <<http://www.w3.org/1999/02/22-rdf-syntax-ns#>>

PREFIX rdfs: <<http://www.w3.org/2000/01/rdf-schema#>>

PREFIX ter: <<http://datiopen.istat.it/odi/ontologia/territorio/>>

PREFIX cen: <<http://datiopen.istat.it/odi/ontologia/censimento/>>

PREFIX qb: <<http://purl.org/linked-data/cube#>>

PREFIX xsd: <<http://www.w3.org/2001/XMLSchema#>>

SELECT ?comune

WHERE {

?comune ?p "Firenze"@it.

}

comune
< <a href="http://datiopen.istat.it/odi/risorsa/territorio/centriabitati/Firenze_4801710002">http://datiopen.istat.it/odi/risorsa/territorio/centriabitati/Firenze_4801710002</a> >
< <a href="http://datiopen.istat.it/odi/risorsa/territorio/comuni/Firenze_48017">http://datiopen.istat.it/odi/risorsa/territorio/comuni/Firenze_48017</a> >
< <a href="http://datiopen.istat.it/odi/risorsa/territorio/centriabitati/Firenze_4801710002">http://datiopen.istat.it/odi/risorsa/territorio/centriabitati/Firenze_4801710002</a> >
< <a href="http://datiopen.istat.it/odi/risorsa/territorio/comuni/Firenze_48017">http://datiopen.istat.it/odi/risorsa/territorio/comuni/Firenze_48017</a> >
< <a href="http://datiopen.istat.it/odi/risorsa/territorio/province/Firenze">http://datiopen.istat.it/odi/risorsa/territorio/province/Firenze</a> >
< <a href="http://datiopen.istat.it/odi/risorsa/territorio/province/Firenze">http://datiopen.istat.it/odi/risorsa/territorio/province/Firenze</a> >

SELECT ?p

WHERE {

<[http://datiopen.istat.it/odi/risorsa/territorio/comuni/Firenze\\_48017](http://datiopen.istat.it/odi/risorsa/territorio/comuni/Firenze_48017)> ?p ?o .

}

<b>p</b>
<http://datiopen.istat.it/odi/ontologia/territorio/haIndicatoreCensimento>
<http://datiopen.istat.it/odi/ontologia/territorio/sezione_di_COM>
<http://datiopen.istat.it/odi/ontologia/territorio/haIndicatoreCensimento>
<http://datiopen.istat.it/odi/ontologia/territorio/sezione_di_COM>
<http://datiopen.istat.it/odi/ontologia/territorio/sezione_di_COM>
<http://datiopen.istat.it/odi/ontologia/territorio/haIndicatoreCensimento>
<http://datiopen.istat.it/odi/ontologia/territorio/sezione_di_COM>
<http://datiopen.istat.it/odi/ontologia/territorio/sezione_di_COM>
<http://datiopen.istat.it/odi/ontologia/territorio/sezione_di_COM>

SELECT ?p

WHERE {

<http://datiopen.istat.it/odi/risorsa/territorio/comuni/Firenze\_48017>

<http://datiopen.istat.it/odi/ontologia/territorio/haIndicatoreCensimento> ?o .

?o ?p ?popolazioneResidente

}

<b>p</b>
<http://datiopen.istat.it/odi/ontologia/censimento/haInterni>
<http://datiopen.istat.it/odi/ontologia/censimento/haClassiEta16Categorie>
<http://datiopen.istat.it/odi/ontologia/censimento/haPopolazioneResidente>
<http://purl.org/linked-data/cube#dataSet>
<http://datiopen.istat.it/odi/ontologia/censimento/haPeriodoCostruzione>
<http://datiopen.istat.it/odi/ontologia/censimento/haAnno>
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://datiopen.istat.it/odi/ontologia/censimento/haAnno>
<http://datiopen.istat.it/odi/ontologia/censimento/haGradoIstruzione>
<http://purl.org/linked-data/cube#dataSet>
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://datiopen.istat.it/odi/ontologia/censimento/haClasseEta15Anni>

SELECT DISTINCT ?o ?popolazioneResidente ?label

WHERE {

```

<http://datiopen.istat.it/odi/risorsa/territorio/comuni/Firenze_48017>
<http://datiopen.istat.it/odi/ontologia/territorio/haIndicatoreCensimento> ?o .

?o <http://datiopen.istat.it/odi/ontologia/censimento/haPopolazioneResidente>
?popolazioneResidente.

?o rdfs:label ?label .

}

```

<http://datiopen.istat.it/odi/risorsa/censimento/indicatori/popolazione/comune/P40_48017>	"11598"	"Male resident population with age between 50 and 54 years in Municipality Firenze" @en
<http://datiopen.istat.it/odi/risorsa/censimento/indicatori/popolazione/comune/P14_48017>	"14200"	"Popolazione residente - Età < 5 anni per Comune di Firenze" @it
<http://datiopen.istat.it/odi/risorsa/censimento/indicatori/lavoro/comune/P132_48017>	"10143"	"Popolazione residente totale maschi di 15 anni e più studenti per Comune di Firenze" @it
<http://datiopen.istat.it/odi/risorsa/censimento/indicatori/popolazione/comune/P16_48017>	"14012"	"Resident population with age between 10 and 14 years in Municipality Firenze" @en
<http://datiopen.istat.it/odi/risorsa/censimento/indicatori/studio/comune/P53_48017>	"157131"	"Popolazione residente maschi di 6 anni o più per Comune di Firenze" @it

SELECT DISTINCT ?popolazioneResidente

WHERE {

```

<http://datiopen.istat.it/odi/risorsa/territorio/comuni/Firenze_48017>
<http://datiopen.istat.it/odi/ontologia/territorio/haIndicatoreCensimento>
<http://datiopen.istat.it/odi/risorsa/censimento/indicatori/popolazione/comune/P14_48017>
.

<http://datiopen.istat.it/odi/risorsa/censimento/indicatori/popolazione/comune/P14_48017>
<http://datiopen.istat.it/odi/ontologia/censimento/haPopolazioneResidente>
?popolazioneResidente.

}

```

<b>popolazioneResidente</b>
"14200" ^^<http://www.w3.org/2001/XMLSchema#decimal>

Lo stesso è stato fatto anche per la città di Siena e anche questa volta ho aggiunto manualmente il dato all'ontologia.

Se si unisse la mia ontologia con i LOD di Istat, attraverso la query CONSTRUCT, si potrebbe popolare il Data Property, *haPopolazioneResidente*, in modo automatico.

Utilizzo di CONSTRUCT

CONSTRUCT mi permette di costruire una triade che non esiste già, per trovare un dato di Istat, ma di cui Istat non fornisce il modo di ricavarlo tramite LOD.

Grazie a questa nuova triade, generata da CONSTRUCT, permetto di raggiungere il dato tramite i LOD e senza dover fare ulteriori query.

La query CONSTRUCT si realizza inserendo il WHERE della query che trova il dato finale, nel WHERE del CONSTRUCT:

CONSTRUCT

```
{  
<http://datiopen.istat.it/odi/risorsa/territorio/comuni/Firenze_48017> sw:haPopolazioneResidente  
?popolazioneResidente.  
}
```

WHERE

```
{  
  
<http://datiopen.istat.it/odi/risorsa/territorio/comuni/Firenze_48017>  
<http://datiopen.istat.it/odi/ontologia/territorio/haIndicatoreCensimento>  
<http://datiopen.istat.it/odi/risorsa/censimento/indicatori/popolazione/comune/P14_48017> .  
  
<http://datiopen.istat.it/odi/risorsa/censimento/indicatori/popolazione/comune/P14_48017>  
<http://datiopen.istat.it/odi/ontologia/censimento/haPopolazioneResidente> ?popolazioneResidente.  
}
```

In questo modo posso aggiornare il dato popolazione residente per Firenze in modo automatico tramite questa query.

Ulteriore utilizzo di CONSTRUCT

Sempre grazie a CONSTRUCT è possibile, inoltre, creare nuovi predicati che permettano di ottenere i dati presenti nella tabella Istat, semplicemente riusando il WHERE delle query che ho realizzato in precedenza sulla mia ontologia. Questi nuovi predicati permettono di ricavare i dati della tabella tramite LOD.