

# Relazione Progetto Open Data

## Martinello Pierfrancesco

### Introduzione

Lo scopo del progetto è stato quello di pulire dati aperti relativi a luoghi di interessi culturali presenti nelle regioni [Sardegna](#) e [Puglia](#) e la città di [Palermo](#), creare un' ontologia per questi dati e collegarla ad altre basi di conoscenza, andando a creare dati a 5 stelle.

### I dati

I dati sono stati presi dal portale dei dati aperti di Palermo, per quanto riguarda i dati della sopramenzionata città e dal portale italiano dei dati aperti per i dati sulle due regioni prese in esame. I dati scaricati dai portali sono eterogenei tra di loro ed è stata necessaria un'operazione di pulizia dei dati per eliminare colonne non necessarie e errori dovuti alla codifica. I file originali sono stati mantenuti intatti e condivisi assieme al progetto

### Le licenze

I dati presi in esame sono stati distribuiti con licenze:

- [CC BY 4.0](#) per i dati della città di Palermo;
- [IODL 2.0](#) per il resto dei dati.

Le licenze sono equivalenti ed entrambe permettono completa libertà sulla condivisione e sulla modifica dei dati e la creazione di un lavoro derivato, alla sola condizione di riconoscere la paternità dei dati ai rispettivi proprietari.

### Pulizia dei dati

Come menzionato prima si è ritenuto necessario un'opera di pulizia dei dati. Per fare ciò si è utilizzato il software di pulizia [OpenRefine](#), che permette di semplificare le operazioni di pulizia con i suoi strumenti per il riconoscimento di testo. Di fatto è stato così possibile uniformare i campi che non avevano alcuna informazione, sostituendoli con delle stringhe vuote. Si è preferito questo strumento a programmi di pulizia automatici creati con linguaggio Python poiché alcuni dei dati contenevano per errore di codifica il simbolo ¿. Esso rappresentava sia il simbolo – che il simbolo ´ e un programma automatico avrebbe potuto preferire erroneamente un simbolo all'altro. Più dettagli nel file chiamato *log\_modifiche.txt*.

### Creazione dei dati a quattro stelle

I dati puliti sono stati processati da un programma Python, utilizzando la libreria [RDFLib](#) per creare un grafo di conoscenza, inserire i dati come nodi del suddetto grafo e serializzare il risultato in un file *Turtle*. Vengono proposte due versioni del programma, una dove l'ontologia creata è stata esplicitamente inserita nel grafo di conoscenza prima che i dati vengano processati e l'altra dove l'ontologia è implicita (descritta nel file *implicit\_ontology.txt*).

L'ontologia presenta una sola classe, la classe *POI* con le relative proprietà *Typology*, *Link\_to*, *Address* e *Is\_Located\_in*.

Ognuno dei 3 dataset utilizzati viene esaminato singolarmente, Viene infatti aperto un solo file alla volta e chiuso quando non più necessario grazie all'istruzione *with*. Alla lettura di una singola riga, viene aggiunto al grafo di conoscenza una serie di nodi che sono stati creati attraverso la funzione *rdflib.Graph.add()*

## URI

Il dominio fittizio utilizzato è <http://poifinder.it> . Vi è una parte dedicata all'ontologia alla seguente URI <http://poifinder.it/ontology/> e una parte dedicata alle risorse alla URI <http://poifinder.it/resource/> . Quando viene aggiunta una serie di nodi, essa è reperibile alla URI [http://poifinder.it/resource/{nome\\_città}/{nome\\_punto\\_di\\_interesse}](http://poifinder.it/resource/{nome_città}/{nome_punto_di_interesse})

## Collegamento con altre basi di conoscenza e creazione di dati a 5 stelle

Vi è un collegamento con una base di conoscenza esterna e con una vocabolario RDF. Il vocabolario utilizzato è [Geo](#) ed è stato introdotto per gestire al meglio le coordinate geografiche presenti nel grafo delle conoscenze. La base di conoscenza esterna utilizzata è [DBPedia.org](http://dbpedia.org) ed è stata utilizzata per creare collegamenti tra l'ontologia e altre fonti di conoscenza. Per ogni punto di interesse, la località in cui è sito viene direttamente collegata al corrispettivo in DBPedia (per la città di Palermo, ad esempio, la URI è <http://dbpedia.org/resource/Palermo> ).

## Risultati

La base di conoscenza ottenuta può essere facilmente utilizzata per mostrare informazioni sui punti di interesse attraverso query SPARQL e poi successivamente usare servizi online come <https://www.openstreetmap.org/> per ottenere un percorso dalla posizione fornita dall'utente finale alle coordinate del Punto di Interesse.

Documentazione, i file originali, i programmi e i risultati sono forniti con licenza [CC BY 4.0](#), compatibile con le licenze dei dati originali.