

## NOTA METODOLOGICA AL REPORT

### #WIKITRASPARENZA: CHI APRE DAVVERO I DATI SULLE RISORSE PUBBLICHE

Per questa analisi sono stati presi in considerazione la pubblicazione dei dati relativi alla sezione “beni immobili e la gestione del patrimonio” (Art. 30, d.lgs. n. 33/2013) e dei dati relativi alla sezione “sovvenzioni, sussidi e vantaggi economici”(Art.27 d.lgs. n. 33/2013). Per ciascuna tipologia di dataset, confrontando i vari dati pubblicati dai capoluoghi di Regione italiani, si è scelto un set di attributi standard che se rispettato fornirebbe una buona quantità di informazioni.

#### **Beni immobili e gestione del patrimonio**

Questa sezione è composta da due sottosezioni:

1. patrimonio immobiliare: dove deve essere pubblicato un dataset contenente le informazioni identificative degli immobili posseduti.
2. canoni di locazione o affitto: in cui devono essere pubblicati i dati relativi ai canoni di locazioni o di affitto versati (passivi) o percepiti (attivi.)

Per questa sezione dovrebbero quindi essere pubblicati tre dataset, cioè: patrimonio immobiliare, canoni locazioni attive e canoni locazioni passive.

#### *Dataset “patrimonio immobiliare” - attributi scelti*

Per il dataset patrimonio immobiliare sono stati scelti i seguenti attributi:

1. Ubicazione ed indirizzo completo
2. Dati catastali
3. Superficie
4. Natura del bene
5. Descrizione dell’immobile
6. Periodo di costruzione
7. Valore attuale

#### *Dataset “canoni locazioni attive” - attributi scelti*

Per il dataset canoni locazioni attive sono stati scelti i seguenti attributi:

1. Ubicazione ed indirizzo completo
2. Dati catastali
3. Superficie
4. Finalità utilizzo
5. Canone corrisposto
6. Periodicità canone
7. Canone iniziale
8. Data stipula del contratto
9. Locatario
10. Descrizione Immobile

#### *Dataset “canoni locazioni passive” - attributi scelti*

Per il dataset canoni locazioni passive sono stati scelti gli stessi attributi del dataset “canoni locazioni attive” sostituendo l’attributo “Locatario” con l’attributo “Locatore”.

#### **Sovvenzioni, contributi, sussidi, vantaggi economici**

Questa sezione è composta da due sottosezioni: “criteri e modalità” e “atti di concessione”.

Nella sottosezione “atti di concessione” deve essere pubblicato, per legge, un dataset con i seguenti attributi:

1. nome dell’impresa o dell’ente e i rispettivi dati fiscali o il nome di altro soggetto beneficiario (Art. 27, c. 1, lett. a), d.lgs. n. 33/2013).
2. importo del vantaggio economico corrisposto (Art. 27, c. 1, lett. b), d.lgs. n. 33/2013.)
3. norma o titolo a base dell’attribuzione (Art. 27, c. 1, lett. c), d.lgs. n. 33/2013).
4. ufficio e funzionario o dirigente responsabile del relativo procedimento amministrativo (Art.

27, c. 1, lett. d), d.lgs. n. 33/2013).

5. modalità seguita per l'individuazione del beneficiario (Art. 27, c. 1, lett. e), d.lgs. n. 33/2013).

6. link al progetto selezionato (Art. 27, c. 1, lett. f), d.lgs. n. 33/2013).

7. link al curriculum del soggetto incaricato (Art. 27, c. 1, lett. g), d.lgs. n. 33/2013)

Inoltre deve essere disponibile in *formato tabellare aperto* un “elenco dei soggetti beneficiari degli atti di concessione di sovvenzioni, contributi, sussidi ed ausili finanziari alle imprese e di attribuzione di vantaggi economici di qualunque genere a persone ed enti pubblici e privati di importo superiore a mille euro” (Art. 27, c. 2, d.lgs. n. 33/2013).

Sempre nella stessa sezione deve essere presente un albo dei beneficiari che comprenda “soggetti, ivi comprese le persone fisiche, cui sono stati erogati in ogni esercizio finanziario contributi, sovvenzioni, crediti, sussidi e benefici di natura economica a carico dei rispettivi bilanci” (Art. 1, D.P.R. n. 118/2000).

Quindi, in sostanza, possono essere pubblicati 2 dataset, uno in formato tabellare aperto che contenga gli attributi sopra elencati per soddisfare i comma 1 e 2 dell’art. 27, l’altro, in qualsiasi formato che contenga i nomi di tutti i beneficiari.

Per quanto riguarda la sezione criteri e modalità non deve essere pubblicato nessun dataset, ma soltanto le informazioni necessarie a capire come venga erogato un sussidio, questa sezione non è stata presa in considerazione nell’analisi.

#### *Dataset “atti di concessione” - attributi scelti*

Per il dataset atti di concessione sono stati scelti gli attributi obbligatori per legge. Inoltre per il punto 6 si è scelto di aggiungere un attributo “Descrizione progetto finanziato” poiché presente nei dataset di molte amministrazioni che non hanno la possibilità di fornire un link al progetto finanziato.

#### *Dataset “albo dei beneficiari” - attributi scelti*

Per il dataset albo dei beneficiari sono stati scelti i seguenti attributi:

1. Importo del beneficio
2. Dati fiscali del beneficiario
3. Nome del beneficiario
4. Descrizione dell’oggetto del beneficio

#### **Raccolta dei dati e risultati**

Per rappresentare i dati sopra descritti e per rendere possibile la creazione di una Wiki semantica del Decreto Trasparenza e dei relativi dati pubblicati è stato necessario creare un’ontologia semantica<sup>1</sup>. Una volta creata la struttura della Wiki e implementata su Semantic Media Wiki, è stato possibile censire i dataset. Per facilitare la raccolta dei dati è stato creato un bot in grado di inserire in modo automatico le informazioni facilmente reperibili<sup>2</sup> nelle pagine della Wiki relative agli enti pubblici (I Comuni in questo caso) e ai dataset pubblicati. Per quanto riguarda le informazioni relative ai comuni sono state estratte da dbpedia<sup>3</sup> tramite delle query SPARQL. Per quanto riguarda i link alle sezioni della trasparenza sono state estratti dalla “bussola della trasparenza”<sup>4</sup>.

L’inserimento degli altri dati riguardanti i dataset pubblicati è stato invece manuale ha seguito la seguente procedura:

1. Selezionare un Comune
2. Selezionare un dataset relativo al Comune
3. Andare sul sito del Comune scelto e visualizzare il dataset da analizzare
4. Inserire le informazioni richieste nel modulo della Wiki:
  - formato del file
  - url del file
  - segnalare errore se file non è pubblicato

1 Per maggiori informazioni sull’ontologia e sulla creazione della Wiki si faccia riferimento all’Appendice A

2 Quali link alle sezioni dove vengono pubblicati i dataset, link ai siti dei Comuni e informazioni sui Comuni (es: popolazione).

3 Maggiori informazioni: <http://it.dbpedia.org/>

4 Reperibile all’indirizzo: <http://www.magellanopa.it/bussola/>

- riportare la presenza (o l'assenza) di ciascun attributo del dataset
- segnalare errori di pubblicazione (dati troppo aggregati, non comprensibili o in formato non tabellare)

Una volta raccolte le informazioni è stato possibile effettuare diverse interrogazioni sullo stato di pubblicazione dei dati, i cui risultati sono esposti in questo elaborato.<sup>5</sup>

Per quanto riguarda il calcolo degli indici di qualità esposti in questo elaborato, non si sono calcolate dimensioni di qualità intrinseca nel dato (quali l'attualità, la completezza e l'accuratezza) data l'impossibilità di effettuare certi tipi di controllo su dataset con schemi diversi tra loro e in formati per la maggior parte non processabili automaticamente. Tuttavia, facendo una precisazione sull'attualità dei dataset pubblicati: si è assunto, quando non specificato, che il dataset fosse riguardante l'anno 2013 (anno in cui è entrato in vigore il Decreto Trasparenza), quindi, di default, si è considerato che i dati pubblicati fossero attuali. In un caso si sono trovati dati riguardanti solo anni antecedenti al 2011, si è quindi considerato come se il dataset non fosse stato pubblicato.

---

5 Esempi delle query effettuate sono disponibili nell'Appendice A

## Appendice A

### CREAZIONE DELL'ONTOLOGIA

Per rappresentare i dati sopra descritti e per rendere possibile la creazione di una Wiki semantica del Decreto Trasparenza e dei relativi dati pubblicati è stato necessario creare un'ontologia semantica. A tale scopo è stato usato il programma "Protegé"<sup>6</sup>.

Prima di tutto sono state definite le classi (o categorie), si è deciso di creare tre macro categorie:

1. **ObbligoPubblicazione**: contiene tutte le sotto categorie che descrivono gli obblighi di pubblicazione
2. **FormatoFile**: contiene i formati in cui i file possono essere pubblicati
3. **PubblicaAmministrazione**: contiene gli enti che pubblicano i dataset.

Successivamente sono state create le "Object Properties" che stabiliscono le relazioni tra le varie classi:

1. **HaPubblicato**: con dominio la classe "PubblicaAmministrazione" e range la classe "ObbligoPubblicazione"
2. **HaFormatoFile**: che ha come dominio "ObbligoPubblicazione" e range "FormatoFile"
3. **RiguardaEnte**: che ha come dominio "ObbligoPubblicazione" e range "PubblicaAmministrazione". Quest'ultima è la proprietà inversa di "HaPubblicato"

Sono state definite poi le "Data Properties". Tutte le data properties rappresentanti un'attributo di un dataset hanno formato "Boolean". Per ogni attributo analizzato che rientri nel set di attributi scelti nell'analisi precedente, sarà quindi possibile scegliere un valore "vero" se l'attributo è presente nel dataset, oppure "falso" nel caso in cui non sia presente.

Oltre alle proprietà rappresentanti gli attributi dei dataset, sono state create altre tre "Data Properties":

1. **tracciaRecordAttributiDataset**: che contiene la traccia record di tutti gli attributi del dataset (quindi anche quelli che non vengono rappresentati dalle data properties scelte). Ha formato "String".
2. **URLdataset**: contiene il link al dataset analizzato. Ha formato "URL".
3. **URLsezione**: contiene il link alla sezione in cui è disponibile il file da scaricare. Ha formato "URL"

Come ultimo passaggio sono state create le entità per ciascuna classe. Per la classe **ObbligoPubblicazione** le entità sono i dataset analizzati, per la classe "PubblicaAmministrazione" sono le Pubbliche Amministrazioni di cui si sono analizzati i file, per la classe "FormatoFile" sono i formati dei file usati per pubblicare i dataset.

Il risultato finale, in cui sono stati inseriti i dati pubblicati dal Comune di Asti, è visibile nella figura 1. La didascalia degli archi (che rappresentano le Object Properties) è presente nella figura 2. Le sottoclassi di **ObbligoTrasparenza** includono, oltre alle due sezioni analizzate (cioè "Beni Immobili" e "Sovvenzioni Sussidi e contributi"), anche le sezioni "personale" e "performance" per mostrare che lo studio è ampliabile anche alle altre sezioni degli obblighi di pubblicazione.

---

6 [http://en.Wikipedia.org/Wiki/Prot%C3%A9g%C3%A9\\_\(software\)](http://en.Wikipedia.org/Wiki/Prot%C3%A9g%C3%A9_(software))

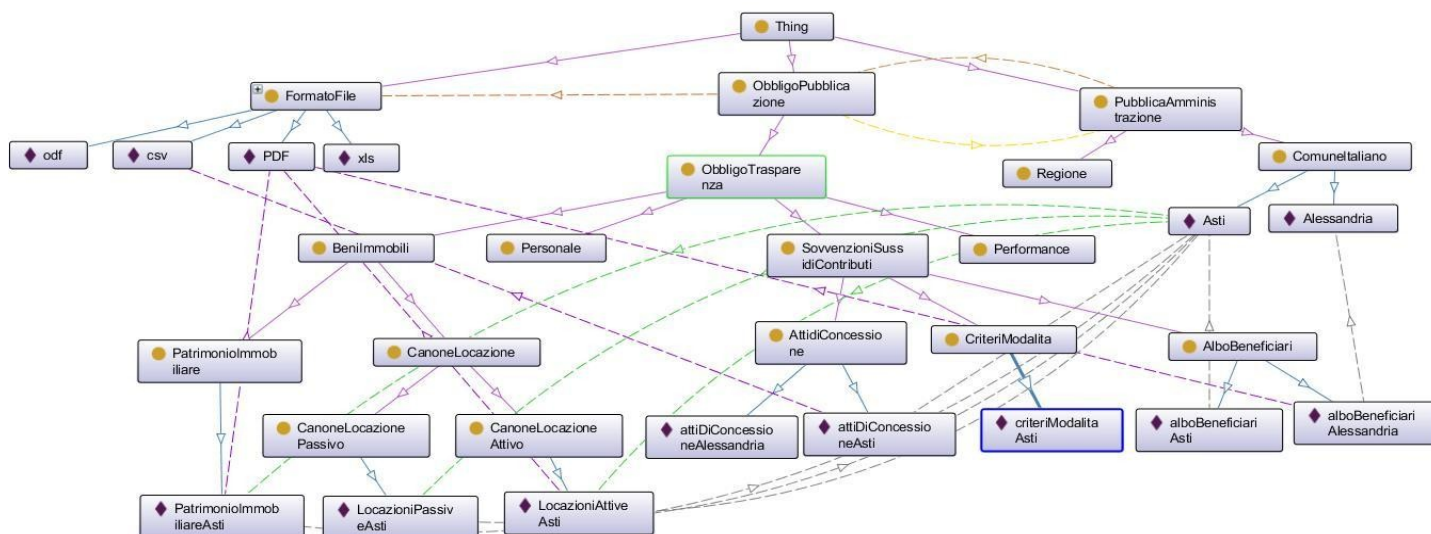


figura 1



figura 2

## CREAZIONE DELLA WIKI SEMANTICA

La Wiki è stata creata scegliendo un approccio bottom-up: sono state create prima le pagine (entità) che descrivono i dataset analizzati con le varie proprietà (data properties e object properties) e successivamente le categorie (classi) a cui esse appartengono come descritto nell'ontologia semantica<sup>7</sup>.

### templates

Per facilitare l'input dei dati nella Wiki e creare dei moduli *user friendly* sono stati creati diversi template:

1. LocazioniAttive
2. LocazioniPassive
3. ProprietaGenerali
4. alboBeneficiari

<sup>7</sup> La Wiki è disponibile a questo indirizzo: [https://trasparenza.nexacenter.org/w/index.php/Pagina\\_principale](https://trasparenza.nexacenter.org/w/index.php/Pagina_principale)

5. attiConcessione
6. erroriFilePubblicato
7. erroriPubblicazione
8. patrimonioImmobiliare

I template “LocazioniAttive”, “LocazioniPassive”, “patrimonioImmobiliare”, “alboBeneficiari” e “attiConcessione” sono specifici ai gruppi di dataset analizzati e contengono al loro interno le data properties scelte per ciascun gruppo di dataset (ad esempio il gruppo dei dataset “LocazioniAttive”). I template “ProprietàGenerali”, “erroriPubblicazione”, “erroriFilePubblicato” contengono al loro interno proprietà comuni a ciascun dataset.

Il template “erroriPubblicazione” contiene le proprietà che descrivono errori nella pubblicazione del file, cioè quando il file non è stato pubblicato. Il template “erroriFilePubblicato” contiene le proprietà che descrivono gli errori nel file pubblicato, ad esempio il fatto che non sia in formato tabellare come prescritto dalla legge. Entrambi i template di errore se presenti nella pagina relativa al dataset associano la pagina alla categoria “Errore”.

Il template “ProprietàGenerali” come da nome, contiene le proprietà che ciascun dataset possiede, cioè:

- da che ente è stato pubblicato (o non pubblicato)
- l’URL del dataset
- l’URL della sezione e il formato del file.

### **I moduli**

Una volta creati i templates, sono stati creati dei moduli (forms) per rendere l’input dei dati meno difficoltoso. Ogni modulo creato serve per modificare uno specifico tipo di dataset analizzato, i moduli creati sono quindi:

1. AlboBeneficiari
2. AttiConcessione
3. LocazioniAttive
4. LocazioniPassive
5. PatrimonioImmobiliare

Ogni modulo è formato dai template:

1. “ProprietàGenerali”
2. il relativo template specifico (ad esempio il template “LocazioniAttive” per il modulo “LocazioniAttive”).
3. Gli opzionali template di Errore, necessari in caso si riscontrino errori. Sono disponibili cliccando sul pulsante “aggiungi un altro” nella fase di compilazione del modulo e faranno rientrare la pagina creata, oltre che nella categoria specifica per il dataset, anche nella categoria “Errore”.

## **ESEMPI DI INLINE QUERIES NELLA WIKI**

Una volta creata la Wiki con una base di dati abbastanza consistente, si è deciso di trovare le risposte alle seguenti domande:

1. Quali sono i dataset non pubblicati
2. Quale ente ha commesso più errori
3. Quale categoria di dataset contiene più errori
4. Chi ha pubblicato più file in un formato aperto e processabile automaticamente (csv, ods, xml)
5. Quali sono i formati più usati per pubblicare i dataset
6. Quale ente non ha pubblicato i dataset con tutti gli attributi previsti per legge

### **Quali sono i dataset non pubblicati**

```
Query:{{#ask: [[Categoria:ObbligoTrasparenza]][[fileNonPubblicato::true]]
| mainlabel=Scheda
| ?riguardaEnte = Amministrazione
| ?presenzaMessaggioFileNonPubblicato=Messaggio
```

```
}}
```

### **Quale ente ha commesso più errori**

```
Query: <nome ente>: {{#ask: [[Category:Errore]] [[riguardaEnte::<nome ente>]]  
| format = count  
}}
```

### **Quale categoria di dataset contiene più errori**

```
Query: <nome dataset> {{#ask: [[Category:Errore]] [[Category:<categoria dataset>]]  
| format = count  
}}
```

### **Chi ha pubblicato più file in un formato aperto e processabile automaticamente (csv, ods, xml)**

```
Query: {{#ask: [[Category:CanoneLocazionePassivo]] [[haFormatoFile::csv]] OR  
[[Category:CanoneLocazionePassivo]] [[haFormatoFile::ods]] OR [[Category:CanoneLocazionePassivo]]  
[[haFormatoFile::xml]]  
| format = count  
}}
```

### **Quali sono i formati più usati per pubblicare i dataset**

```
Query: <formato>: {{#ask: [[Category:ObbligoTrasparenza]] [haFormatoFile::<formato>]]  
| format = count  
}}
```

### **Quale ente non ha pubblicato i dataset con tutti gli attributi previsti per legge**

```
Query: {{#ask: [[Category:ViolazionePropObbligatorie]]  
| mainlabel=scheda  
| ?riguardaEnte  
| ?falseSuPropObbligatorie}}
```

## **ESEMPI DI QUERY SPARQL SULL'ENDPOINT DELLA WIKI**

All'indirizzo <https://trasparenza.nexacenter.org/sparql> è stato montato un endpoint per interrogare la wiki tramite SPARQL.

Esempi di query sono presenti su [https://trasparenza.nexacenter.org/wiki/Query\\_sparql](https://trasparenza.nexacenter.org/wiki/Query_sparql).

Esempio:

quale ente ha commesso più errori

```
PREFIX categoria: <http://trasparenza.nexacenter.org/id/Categoria-3A>  
PREFIX property: <http://trasparenza.nexacenter.org/id/Propriet-C3-A0-3A>  
PREFIX entity: <http://trasparenza.nexacenter.org/id/>  
PREFIX categoria: <http://trasparenza.nexacenter.org/id/Categoria-3A>  
SELECT ?ente count (?ente)  
WHERE {  
  ?dataset a categoria:Errore;  
           property:RiguardaEnte ?ente;  
           property:FileNonPubblicato 1.  
}
```