Nexa Center for Internet & Society
Politecnico di Torino

Data di pubblicazione: 24 settembre 2014

Trasparenza della Pubblica Amministrazione: i primi risultati di un'analisi di dettaglio

Il presente documento è disponibile all'indirizzo:
http://nexa.polito.it/lunch-24

Studying the Internet, exploring its potential & experimenting new ideas

Via Pier Carlo Boggio 65/A, 10129 Torino, Italia
(dove siamo: http://nexa.polito.it/contact)
+39 011 090 7217 (Telefono)
+39 011 090 7216 (Fax)
info@nexa.polito.it

Indirizzo postale:
Centro Nexa su Internet & Società
Politecnico di Torino - DAUIN
Corso Duca degli Abruzzi, 24
10129 TORINO

Il Centro Nexa su Internet & Società è un centro di ricerca del Dipartimento di Automatica e Informatica del Politecnico di Torino (http://dauin.polito.it).

Gli autori di questo lavoro sono: Lorenzo Canova, Raimondo Iemma e Federico Morando. Si ringraziano per il contributo: Alessandro Ferrarin, Giuseppe Futia, Paola Chiesa, Andrea Violato, Pasquale Pellegrino.

Il presente documento è rilasciato dal Centro Nexa su Internet & Società secondo i termini della licenza Creative Commons Attribuzione 4.0 International.

Table of Contents

Executive summary **1**
Introduzione **2**
L'hackathon e gli obiettivi **2**
I risultati del monitoraggio dei capoluoghi di Regione **3**
Analisi dei formati 3
Errori di pubblicazione 6
I risultati per capoluogo di Regione 7
I risultati dei capoluoghi di provincia in Piemonte **10**
Limiti e sviluppi futuri **12**
Conclusioni **12**
Nota metodologica **13**
Beni immobili e gestione del patrimonio 13
Sovvenzioni, contributi, sussidi, vantaggi economici 14
Raccolta dei dati e risultati 14
Bibliografia essenziale **15**

EXECUTIVE SUMMARY

Il Decreto Trasparenza (d.lgs. n.33, 14 marzo 2013) disciplina gli obblighi di pubblicità, trasparenza e diffusione delle informazioni da parte delle Pubbliche Amministrazioni (PA) ed attua la legge anticorruzione (190/2012). Dall'approvazione del decreto molti dataset sono stati pubblicati dagli enti pubblici, tuttavia spesso capita che questi non forniscano una quantità di informazioni sufficiente (o che siano sufficientemente comprensibili), oppure che non siano presenti nella sezione in cui dovrebbero essere consultabili.

L'obiettivo di questo progetto è di fornire uno strumento flessibile per verificare il rispetto degli obblighi di pubblicazione da parte delle PA, individuando le principali lacune ed evidenziando le migliori pratiche, presentando evidenze confrontabili tra loro.

A tal fine sono stati selezionati dal Decreto Trasparenza due obblighi di pubblicazione, Beni immobili e gestione del patrimonio (Art. 30, d.lgs. n. 33/2013) e Sovvenzioni, sussidi e contributi (Art.27 d.lgs. n. 33/2013), che prescrivono in tutto la pubblicazione di 5 dataset, su cui si possano fare confronti tra i vari enti pubblici. Per ogni obbligo di pubblicazione, dove non specificato per legge, si è scelto un insieme di attributi ritenuti ragionevolmente necessari a rendere sufficientemente informativa per i terzi la pubblicazione di quel dataset. Per registrare le informazioni riguardanti i dataset scrutinati, è stata creata un'ontologia del Decreto Trasparenza e successivamente una Wiki semantica in cui è possibile eseguire delle query.

Sono state formulate varie interrogazioni sulla Wiki semantica che forniscono indicazioni sulla qualità dell'informazione pubblicata. Gli aspetti della qualità analizzata riguardano: (i) il formato del file (aperto o non aperto, è processabile automaticamente o no) (ii) la pubblicazione del file e i relativi errori (il file è stato pubblicato? È comprensibile?) (iii) la quantità di informazione fornita (ossia il numero di attributi presenti nel dataset sul totale di quelli che dovrebbero essere presenti per legge).

Su 100 dataset analizzati, 27 non sono stati pubblicati, 11 contengono errori gravi di pubblicazione (dati troppo aggregati, non comprensibili o non in formato tabellare), il 38% dei dati è di fatto inutilizzabile. Inoltre dei dataset pubblicati soltanto 26 su 73 (il 36%) sono in formato aperto e processabile automaticamente così come prescritto dall'art. 7 del d.lgs. n. 33/2013. Il capoluogo di Regione più virtuoso, secondo le dimensioni prese in considerazione in questa analisi, è Genova, a seguire ci sono Torino e Perugia che però ottengono risultati positivi solo in (i) il formato dei file e (ii) la pubblicazione senza errori del dataset. Inoltre si nota una lieve tendenza a pubblicare dati migliori al Nord/Nord Ovest.

È infine da notare come per il dataset riguardante gli atti di concessione, per cui la legge (ai sensi dell'Art. 27) è decisamente più prescrittiva rispetto agli altri obblighi analizzati (vengono infatti elencati i singoli attributi che dovrebbero essere presenti nel dataset pubblicato), i dati siano di qualità superiore: nella nostra analisi è maggiormente frequente che siano pubblicati e che siano in formato aperto. Le possibili motivazioni potrebbero essere: (i) una norma più dettagliata fornisce una sorta di "linea guida" per la pubblicazione dei dati (ii) a questa specifica norma (l'Art. 27) sono associate sanzioni particolarmente ingenti, in caso di violazione.

Analizzando una parte esigua dei dati pubblicati si sono ottenuti dei primi significativi risultati, l'obiettivo finale del progetto è quello di riuscire a fornire uno strumento per mappare in futuro tutti gli obblighi di trasparenza e la loro applicazione in un sotto-insieme significativo degli enti pubblici italiani (ad esempio, tutti i Comuni capoluogo di Provincia, tutte le Regioni, i Ministeri, ecc.).

Introduzione

Le opportunità derivanti dai dati aperti (Open Data)[1] sono sinora colte solo in parte. Della grande mole di dati prodotti e gestiti dalle Pubbliche Amministrazioni, solo una piccola parte è attualmente pubblicata come Open Data. Eppure i potenziali benefici non sono trascurabili, e vanno da un incremento della trasparenza a maggiori opportunità di partecipazione alle decisioni pubbliche, fino alle possibilità di creazioni di nuovi servizi basati sui dati. Alcuni di questi benefici possono investire anche la Pubblica Amministrazione (PA) stessa in termini di maggiore efficienza e razionalizzazione dei processi interni. Al fine di ottenere effetti positivi tangibili, non è sufficiente che i dati vengano pubblicati; ciò deve anche avvenire entro standard definiti, e con un livello di qualità sufficiente a permettere il loro agevole riutilizzo.

In Italia, il cosiddetto Decreto Trasparenza (d.lgs. n.33, 14 marzo 2013) disciplina gli obblighi di pubblicità, trasparenza e diffusione delle informazioni da parte delle Pubbliche Amministrazioni (PA) ed attua la legge anti-corruzione (190/2012). Dall'approvazione del decreto molti dataset sono stati pubblicati dagli enti pubblici, tuttavia spesso capita che questi non forniscano una quantità di informazioni sufficiente, oppure, che non siano presenti nella sezione in cui dovrebbero essere consultabili, oppure, ancora, che siano pubblicati in formati non aperti, o non facilmente elaborabili.

Per riuscire ad avere una panoramica sullo stato della pubblicazione dei vari dataset da parte delle Pubbliche Amministrazioni, è stata creata un'ontologia del Decreto Trasparenza e, successivamente, una Wiki semantica come strumento per censire i dati pubblicati.

A tal fine, durante il 23° Nexa Lunch Seminar[2] del 23 luglio 2014 si è tenuto un hackathon in cui è stato effettuato un monitoraggio "distribuito", supportato dalla Wiki semantica, per la verifica del rispetto degli obblighi di trasparenza relativi all'impiego delle risorse pubbliche dei principali comuni italiani. L'hackathon si è concentrato su 2 ambiti specifici (i) beni immobili (ii) sovvenzioni, sussidi e contributi, valutando se e come in questi due contesti gli obiettivi del "Decreto Trasparenza" vengano adempiuti dalle amministrazioni pubbliche, verificando l'aderenza con le specifiche tecniche richieste (ad es. in termini di campi delle tabelle).

Nel presente Report vengono esposti e discussi i risultati emersi analizzando i dati ottenuti dal censimento dei dataset.

L'hackathon e gli obiettivi

Per l'intero progetto sono stati definiti 3 goal: (i) analizzare tutti i capoluoghi di Regione (ii) analizzare i Comuni con almeno 100 mila abitanti (iii) analizzare tutti i Comuni capoluogo di Provincia (con meno di 100 mila abitanti).

Per ogni ente sono stati censiti 5 tipi (o categorie) di dataset riguardanti due ambiti specifici[3]:

1. Beni immobili e la gestione del patrimonio (Art. 30, d.lgs. n. 33/2013)
   a) Patrimonio immobiliare
   b) Canoni di locazione attivi
   c) Canoni di locazione passivi
2. Sovvenzioni, sussidi e contributi (Art.26-27 d.lgs. n. 33/2013)
   a) Atti di concessione
   b) Albo dei beneficiari

1 Per la definizione di Open Data si faccia riferimento a questo sito: http://opendefinition.org/od/
2 La descrizione del lunch seminar con i documenti allegati è disponibile a questo indirizzo: http://nexa.polito.it/lunch-23
3 Per maggiori dettagli si faccia riferimento alla Nota Metodologica a pagina 13

L'hackaton ha permesso di portare a termine il primo goal, analizzando 5 dataset per ognuno dei 20 capoluoghi di Regione, per un totale di 100 dataset. Nella Wiki, inoltre, erano già stati inseriti in precedenza i dati relativi ai capoluoghi di Provincia del Piemonte e successivamente sono stati recensiti altri 3 dataset, per arrivare ad un totale di 145 dataset scrutinati.

I risultati del monitoraggio dei capoluoghi di Regione

Eseguendo delle interrogazioni sulla Wiki semantica, in cui sono stati inseriti i dati riguardanti i censimenti dei dataset, è stato possibile ricavare degli indici di qualità dei dati pubblicati.

In questa analisi sono stati presi in considerazione soltanto i dati relativi ai 100 dataset riguardanti i capoluoghi di Regione. Questa scelta è stata effettuata in modo da analizzare dataset che fossero recensiti contemporaneamente[4] e che fossero pubblicati tra enti del medesimo livello amministrativo, distribuiti in modo omogeneo sul territorio italiano.

L'analisi della qualità dei dati pubblicati si è concentrata su tre specifiche dimensioni:

1. Tipo di formati in cui sono stati pubblicati i file, ossia la percentuale di dataset in formato aperto[5] pubblicati.
2. Presenza di errori di pubblicazione, ossia la percentuale di dataset non pubblicati, oppure inutilizzabili perché affetti da problemi quali scarsa leggibilità, aggregazione eccessiva delle informazioni, dati in formato non tabellare (quando prescritto per legge).
3. Quantità di informazioni presenti nei dataset: per ognuna delle 5 categorie di dataset analizzati, è stato definito un set di attributi standard che, se rispettato, fornirebbe una buona quantità di informazioni[6]. Per ora l'analisi dei risultati si concentra solo sulla pubblicazione degli attributi appartenenti ai dataset riguardanti la categoria "Atti di concessione". È stata effettuata questa scelta poiché è l'unico obbligo di pubblicazione (tra quelli analizzati) che prevede per legge che nel dataset sia presente un set di attributi predefinito. La dimensione della qualità del dato, in questo caso, è il numero di attributi pubblicati sul totale degli attributi prescritti per legge.

In questa analisi, i risultati vengono esposti dapprima aggregati e per categoria di dataset pubblicato, successivamente per singolo capoluogo di Regione.

Analisi dei formati

In prima istanza sono stati analizzati i formati dei dataset pubblicati. Nel Grafico 1 (dove in verde sono rappresentati i formati aperti e facilmente processabili) e in Figura 1, rispettivamente a pagina 4 e 5, si può notare come i dati siano nel 46% dei casi pubblicati in formato pdf, di questi il 17% è un pdf raster[7].

Il 34% dei dati è pubblicato in un formato processabile automaticamente e di questi il 76% è anche un formato open, quindi tra i 100 dataset analizzati soltanto 26 rispettano le indicazioni fornite dall'art. 7 del d.lgs. n. 33/2013 e otterrebbero tre stelle nella categorizzazione Five Star Open Data (FSOD)[8].

4 Tutte le informazioni riguardanti i dati pubblicati dai capoluoghi di Regione sono stati raccolti durante l'hackathon il 23 luglio 2014

5 Per la definizione di formato aperto si faccia riferimento all'Art 68 del d.lgs 82/2005 (Codice Amministrazione Digitale)

6 Nella Nota Metodologica a pagina 13 è disponibile una spiegazione in dettaglio degli attributi scelti.

7 I pdf raster sono file pdf che contengono al loro interno immagini bitmap, ad esempio immagini scannerizzate o simili.

8 Five Star Open Data: scala elaborata da Tim Berners Lee per valutare il grado di apertura di un determinato dataset. Per maggiori informazioni: http://5stardata.info/

Formato del file pubblicato	Occorrenze
PDF vettoriale	28
CSV	17
XLS	7
ODS	7
PDF raster	6
HTML	5
DOC	1
HTML tabellare	1
XML	1
Totale	**73**
Dataset non pubblicati	27
Totale dataset analizzati	**100**

Tabella 1: Formati dei file pubblicati - Occorenze per ciascun formato

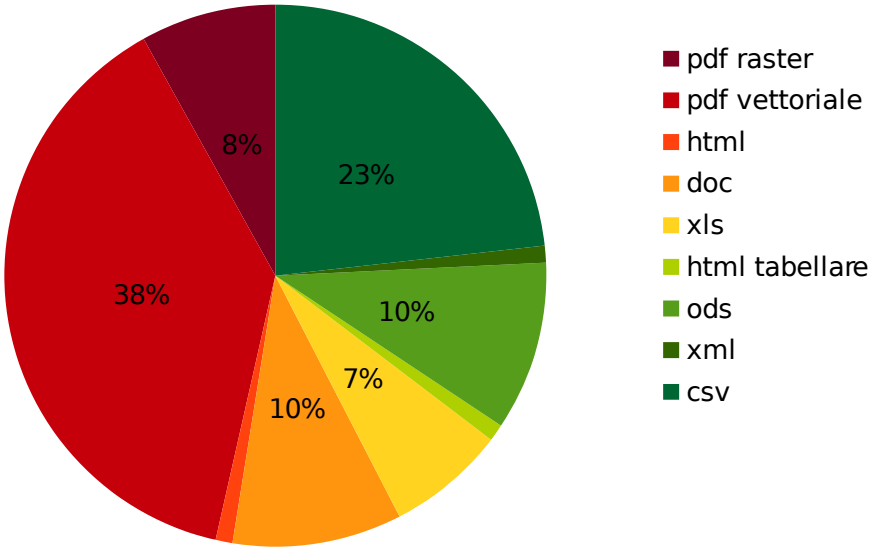

Grafico 1: Formato dei file pubblicati

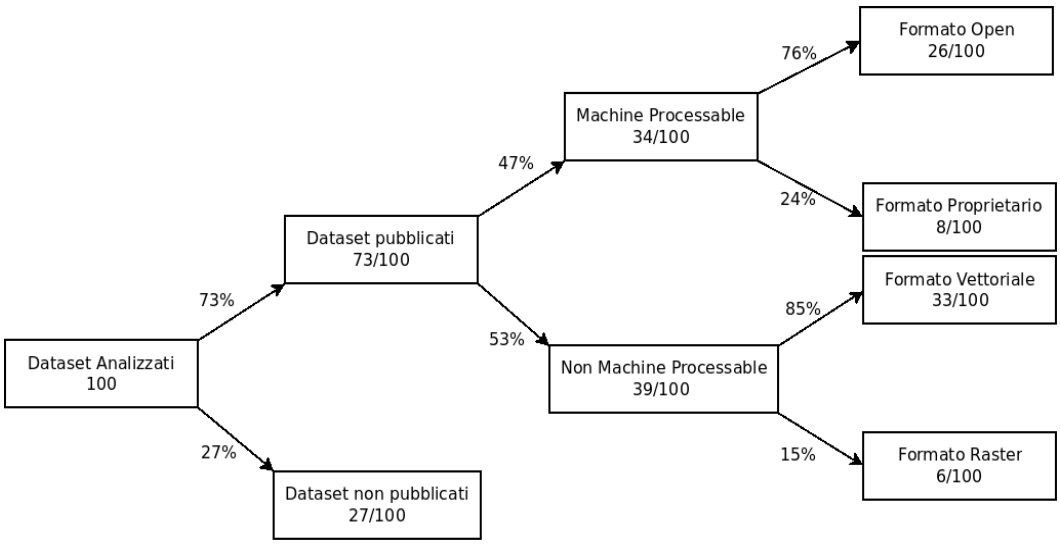

Figura 1: Albero dei formati - capoluoghi di Regione

Analisi dei formati per categoria di dataset

Si nota dal Grafico 2 inoltre, come ci siano delle differenze sostanziali nei formati pubblicati per ciascuna sezione. Per la sezione "Atti di concessione", per cui la legge (ai sensi dell'Art. 27) è decisamente più prescrittiva rispetto agli altri obblighi analizzati, i dati sono generalmente pubblicati in formati aperti e più facilmente elaborabili.

Per sezione analizzata, il numero di dataset che hanno almeno 3 stelle secondo il FSOD (ossia sono in formato: csv, ods e xml)[9] sono:

- Canone Locazioni attive: 5
- Canone Locazioni passive: 3
- Patrimonio Immobiliare: 6
- Atti di concessione: 9
- Albo Beneficiari: 2

9 Nell'ablero in Figura 1 sono state contate le tabelle html come aperti e processabili automaticamente, nella tabella invece sono presenti per sezione solo i csv, gli ods e i csv

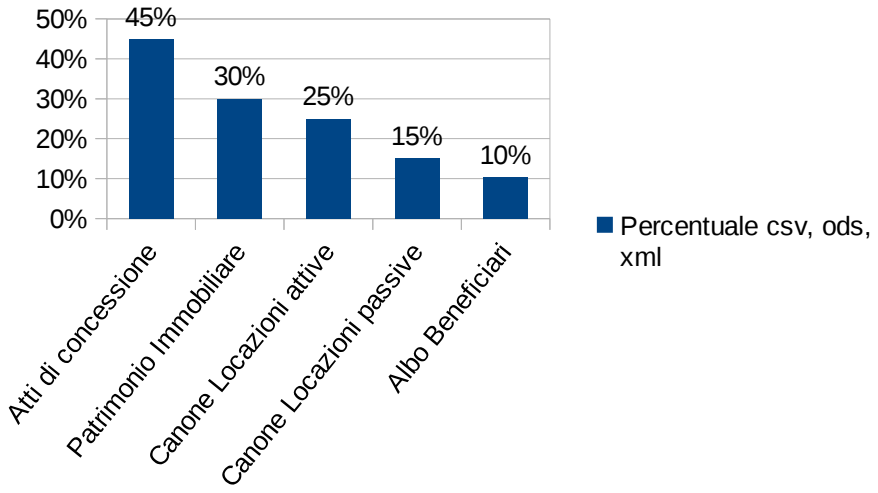

Grafico 2: Percentuale di dataset in formato csv, ods o xml (ossia con almeno 3 stelle FSOD) sul totale (20)

Errori di pubblicazione

Dei 100 dataset presi in considerazione 27 non sono stati pubblicati. Nei restanti 73 dataset, inoltre, sono stati trovati 11 dataset con carenze gravi, quali: scarsa leggibilità, eccessivo livello di aggregazione, dati non in formato tabellare (prescritto dalla legge). In conclusione solo 62 dataset sono stati pubblicati in modo accettabile, nell'ottica di un effettivo riutilizzo, mentre 38 dataset su 100 sono stati considerati come dataset "con errori", ossia, di fatto inutilizzabili (o scarsamente utilizzabili e comunque non a norma di legge) o addirittura assenti.

I dataset con errori si dividono in due categorie:

- File non pubblicati = 27
- File con errori di pubblicazione (es: non tabellare, non comprensibile, troppo aggregato) = 11

Analisi degli errori di pubblicazione per categoria di dataset

Anche in questo caso si nota come nella sezione "Atti di concessione" siano presenti meno errori di pubblicazione rispetto ai dataset riguardanti gli altri obblighi di pubblicazione analizzati. La distribuzione degli errori per sezione è infatti la seguente:

- Canone Locazioni attive: 9, di cui 5 non pubblicati
- Albo Beneficiari: 9, di cui 7 non pubblicati
- Canone Locazioni passive: 8, di cui 6 non pubblicati
- Patrimonio Immobiliare: 7, di cui 6 non pubblicati
- Atti di concessione: 5, di cui 2 non pubblicati

Infine, come riportato nel Grafico 3, ci sono sezioni, quelle riguardanti gli atti di concessione e il patrimonio immobiliare, che sono meno soggette ad errori di pubblicazione e in cui i dati vengono

pubblicati in formati che rispettano maggiormente il paradigma Open Data. Inoltre, la sezione "Atti di concessione" è stata maggiormente penalizzata in questo tipo di analisi poiché, avendo associate prescrizioni di legge più specifiche in termini di formati e campi, il dataset incorreva nella condizione di "errore di pubblicazione" (ad esempio formato non aperto o non tabellare) con maggiore frequenza rispetto ai dati riguardanti le altre sezioni.

Queste osservazioni potrebbero suggerire una relazione tra la specificità delle prescrizioni di legge sulla pubblicazione dei singoli dati e la qualità con cui vengono pubblicati dalle PA[10].

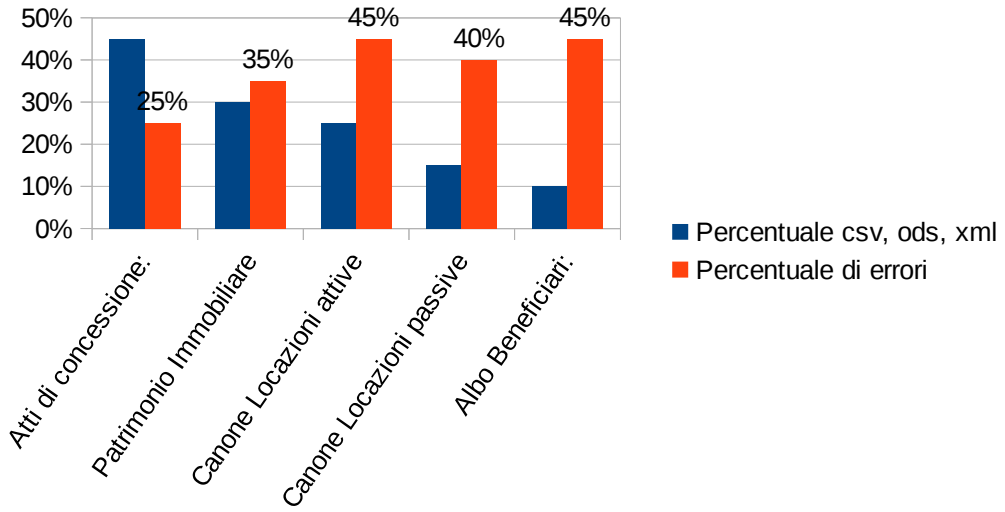

Grafico 3: Percentuale di dataset con errori[11] per sezione e dataset pubblicati in formato csv ods e xml (ossia con almeno 3 stelle FSOD) sul totale dei dataset analizzati (20 per sezione)

I risultati per capoluogo di Regione

Per ogni capoluogo di Regione sono presenti in Tabella 2 i risultati delle query sulla qualità dei dati pubblicati.

La prima colonna della tabella ("Proprietà omesse su 10") rappresenta il numero di attributi prescritti per legge (Art. 27, c. 1, d.lgs. n. 33/2013) che non sono stati inclusi nel dataset pubblicato da ciascun capoluogo nella sezione "Atti di concessione, sovvenzioni e contributi". Nella seconda colonna, denominata "Numero di dataset con errori" sono indicati il numero di file (su 5 analizzati) che presentano un errore[11]. Nella terza colonna infine è indicato il numero di file aperti e processabili automaticamente sui 5 analizzati.

Per avere un'idea migliore sui risultati ottenuti da ciascun capoluogo di provincia, i dati esposti nella tabella sono stati illustrati nelle infografiche in Figura 2, Figura 3 e Figura 4.

10 A pagina 12 vengono ipotizzate probabili cause.

11 Con errore si intende, o il fatto che il file non è stato pubblicato, oppure il fatto che il file è stato pubblicato, ma risulta di fatto inutilizzabile perché troppo aggregato, illeggibile o non in formato tabellare (quando prescritto dalla legge).

	Proprietà omesse su 10 (Atti di concessione)	Numero di dataset con errori	Numero di file open e processabili automaticamente
Ancona	5	3	1
Aosta	5	1	0
Bari	10	3	0
Bologna	6	1	0
Cagliari	3	3	2
Campobasso	1	4	0
Catanzaro	10	4	0
Firenze	5	1	0
Genova	1	0	3
L'Aquila	4	4	1
Milano	2	1	1
Napoli	3	1	2
Palermo	7	1	0
Perugia	5	0	3
Potenza	10	5	0
Roma	3	1	0
Torino	6	1	4
Trento	3	2	0
Trieste	0	2	0
Venezia	5	0	0

Tabella 2: Risultati capoluoghi di Regione italiani

In Figura 2 sono rappresentate le proprietà obbligatorie omesse (su 10). Si nota come soltanto Trieste pubblichi tutti i dati prescritti per legge. Altri esempi virtuosi sono rappresentati da Genova, Campobasso e Milano. Per quanto riguarda Bari, Catanzaro, Palermo e Potenza invece non è stato pubblicato alcun dataset, perciò gli attributi non pubblicati sono 10 su 10.

Figura 2: Numero di Proprietà obbligatorie omesse su 10 (in "Atti di concessione")

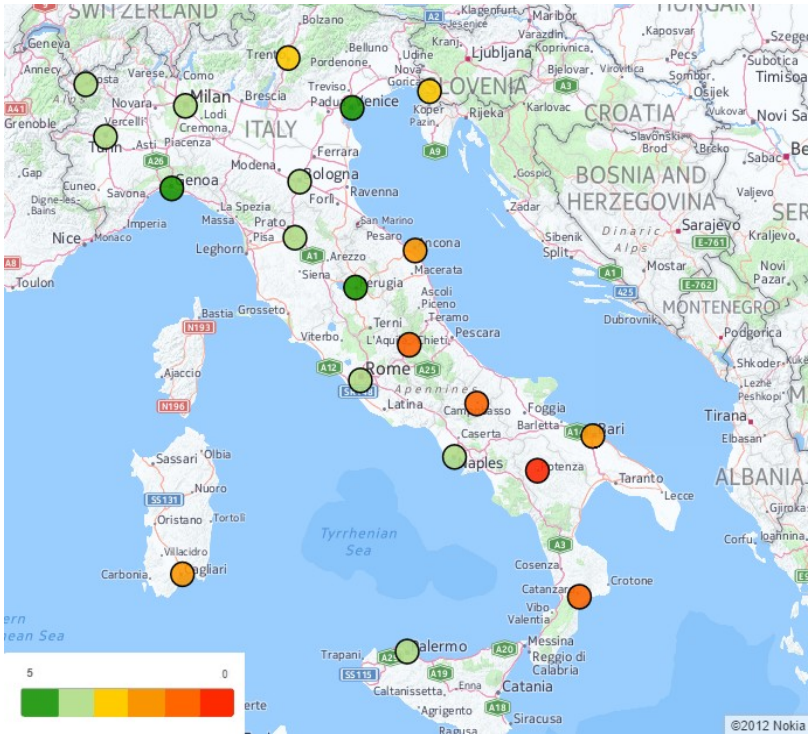

Figura 3: Numero di file pubblicati correttamente (cioè esenti da errori) su 5 analizzati

In Figura 3, vengono rappresentati il numero di file esenti da errori[11] su 5 analizzati. Tali evidenze suggeriscono un potenziale divario tra nord e sud, da confermare includendo nell'analisi almeno tutti i capoluoghi di Provincia. I Comuni più virtuosi, cioè quelli che hanno pubblicato 5 dataset su 5 analizzati, in questo caso sono Genova, Perugia e Venezia.

In Figura 4 è rappresentato il formato in cui ciascun capoluogo di Regione ha pubblicato i dati. Come già evidenziato all'inizio della presentazione dei risultati, la situazione è tutt'altro che rosea. La maggior parte dei dataset in formato aperto è stato pubblicato da un insieme ristretto di Comuni, tra questi spiccano Torino (ha pubblicato 4 dataset su 5 in formato aperto), Genova e Perugia (hanno pubblicato 3 dataset su 5 in formato aperto), tuttavia nessuno di questi ha pubblicato 5 dataset su 5 in formato aperto e processabile automaticamente. Bisogna considerare tuttavia che Torino non ha pubblicato solamente un dataset, quindi tutto ciò che ha pubblicato è in formato aperto e processabile automaticamente.

Figura 4: Numero di file pubblicati in formato open su 5 analizzati

I risultati dei capoluoghi di provincia in Piemonte

L'analisi dei capoluoghi di Provincia piemontesi non è stata oggetto dell'hackathon: essi sono stati analizzati in precedenza per testare l'efficacia della Wiki. In questa analisi sono stati presi in considerazione 40 dataset[12].

Come si nota dalla Tabella 3, rappresentata nell'infografica in Figura 5, certi enti hanno pubblicato dati di qualità maggiore rispetto ad altri. Un caso positivo è rappresentato dal comune di Novara che ha pubblicato tutti e cinque i dataset, quattro di questi sono in formato aperto, inoltre nel dataset "Atti di concessione" ha omesso un solo attributo su 10. I risultati non sono positivi invece per i Comuni di Cuneo, Biella e Alessandria che ottengono un punteggio sotto i valori medi delle scale per tutti e tre gli aspetti presi in considerazione.

12 per ognuno degli 8 capoluoghi di provincia sono stati analizzati 5 dataset appartenenti alle 5 categorie esposte a pagina 2.

	Proprietà omesse su 10 (Atti di concessione)	Numero di dataset con errori	Numero di file open e processabili automaticamente
Alessandria	3	3	1
Asti	0	1	1
Biella	4	3	1
Cuneo	2	3	1
Novara	1	0	4
Torino	6	1	4
Verbania	2	2	1
Vercelli	2	0	1

Tabella 3: Risultati capoluoghi di Provincia piemontesi

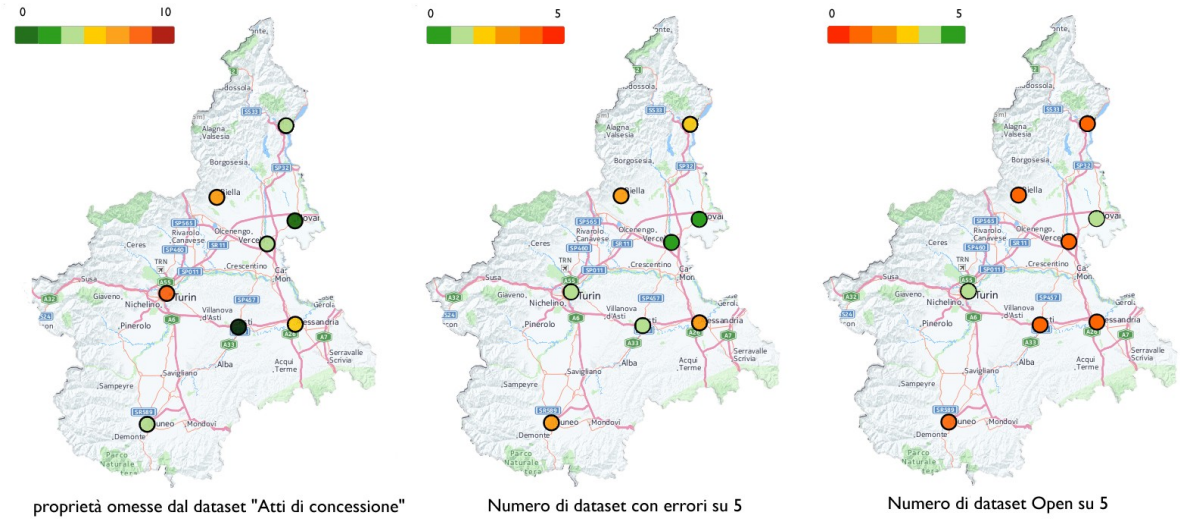

Figura 5: Capoluoghi di Provincia piemontesi - Analisi dei risultati

Limiti e sviluppi futuri

Il limite maggiore di questa analisi è la dimensione piuttosto ridotta della base dati su cui è fondata. Si tratta tuttavia, di un problema che sarà alleviato dedicando ulteriori risorse per immettere dati nella Wiki.

Inoltre, non è stata eseguita nessuna verifica sull'attualità, sulla completezza e sull'accuratezza delle informazioni presenti nei dataset. Per quanto riguarda l'attualità delle informazioni pubblicate si è considerato di default che i dati fossero attuali[13]. Per le dimensioni della qualità riguardanti l'accuratezza e la completezza invece, sarà possibile effettuare dei controlli soltanto una volta che le informazioni saranno pubblicate in formato processabile automaticamente e che, possibilmente, posseggano una struttura standard dello schema con cui sono rappresentate. Le valutazioni riguardo le dimensioni della qualità dei dataset[14] prese in considerazione in questa analisi inoltre, essendo state fatte da persone diverse durante l'hackathon, è possibile che siano soggette a distorsioni dovute ai diversi metri di giudizio adottati da ogni singolo partecipante.

Un altro fattore da tenere sotto controllo relativo alla Wiki riguarda l'aggiornamento degli URL registrati al suo interno che puntano ai dataset pubblicati. Si è notato empiricamente che, dopo pochi mesi, alcuni di questi scadono e non puntano più a nessuna risorsa. Un modo per ovviare a questo problema sarebbe individuare gruppi di persone che siano interessati a tenere aggiornata la Wiki per determinati enti e che, sotto notifica di un bot[15], aggiornino il link presente.

Inoltre coinvolgendo la comunità per tenere aggiornata la base dati, sarebbe possibile ottenere ed analizzare delle serie storiche, che documentino l'evoluzione dei dati aperti in Italia.

Conclusioni

Nonostante il numero di enti analizzati sia relativamente ristretto, questo studio consente di saggiare empiricamente lo stato di attuazione delle prescrizioni di legge tese a incrementare la trasparenza delle Pubbliche Amministrazioni, mediante la pubblicazione di dati in loro possesso.

Pare trovare qualche prima conferma empirica l'ipotesi che una legge più prescrittiva, in questo caso, porti a dati pubblicati in modo migliore, ossia più completi, con più informazioni e pubblicati in formato aperto. La causa di ciò si potrebbe trovare in due giustificazioni: (i) una norma più dettagliata fornisce una sorta di "linea guida" per la pubblicazione dei dati (ii) alle norme più prescrittive in esame si associano anche sanzioni maggiori in caso di violazione[16].

Si può concludere che allo stato attuale ci siano grandi margini di miglioramento rispetto alla pubblicazione di dati relativi all'utilizzo delle risorse pubbliche veramente aperti, utili e riutilizzabili. Ciò non dipende solo dalle inadempienze o mancanze delle singole Pubbliche Amministrazioni. Infatti i dati così come pubblicati ad oggi, in assenza di uno schema standard, in formati per la maggior parte non processabili automaticamente sono in pratica poco riutilizzabili perdendo il loro valore potenziale. Per far sì che l'apertura dei dati delle Pubbliche Amministrazioni non risulti uno sforzo inutile, pare necessario adottare delle buone pratiche quali almeno: (i) pubblicazione dei dati in un formato aperto (ii) pubblicazione di dataset appartenenti alla stessa categoria con schemi simili, in modo che i dati di diverse PA siano confrontabili tra loro. La definizione di tali schemi è, a nostro avviso, un dovere delle amministrazioni centrali dello Stato (ad esempio, l'Agenzia per l'Italia Digitale, in concerto con l'Autorità Nazionale Anticorruzione), ma l'analisi delle buone pratiche già in essere (cui contribuisce anche il presente lavoro) può concorrere a questa definizione.

13 Ulteriori informazioni a riguardo sono disponibili a pagina 15 nella Nota Metodologica
14 Soprattutto per quanto riguarda la comprensibilità delle informazioni pubblicate
15 Uno script automatico che una volta ogni mese, per esempio, controllasse che i link funzionano e in caso contrario mandasse una notifica via mail a chi di interesse.
16 Ulteriori analisi dei dati raccolti, oltre alla raccolta di dati qualitativi – ad esempio tramite interviste dei funzionari pubblici coinvolti nella pubblicazione – sarebbero necessarie per discriminare tra (i) ed (ii).

Per ora il progetto è nella sua fase pilota, tuttavia ci pare sia già riuscito a fornire alcuni risultati significativi. L'obiettivo finale, raggiungibile con la collaborazione della comunità, è quello di fornire uno strumento sempre aggiornato, che mappi tutti gli obblighi di trasparenza e la loro applicazione con riferimento ad un insieme rilevante di enti pubblici. Entro la fine del 2014, il Centro Nexa si propone di condurre questa attività, almeno con riferimento agli obblighi di pubblicazione già censiti, ma allargando il monitoraggio a tutti i capoluoghi di Provincia, alle Regioni (e Province Autonome), ed ai Ministeri.

Nota metodologica

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[17].

Una volta creata la struttura della Wiki e implementata su Semantic Media Wiki, è stato possibile

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

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[18] 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[19] tramite delle query SPARQL. Per quanto riguarda i link alle sezioni della trasparenza sono state estratti dalla "bussola della trasparenza"[20].

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
    - 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)

Il censimento è aggiornato al 23 luglio 2014.

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.[21]

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.

Bibliografia essenziale

Governo Italiano (2013). Decreto Trasparenza, *Decreto Legislativo 14 marzo 2013, n. 33, Gazzetta ufficiale.*

***Benedetto Ponti (2013)**. La trasparenza amministrativa in trasformazione: istituti, opportunità e rischi del nuovo regime dei dati pubblici, seminario al Centro Nexa su internet e società, 16 ottobre 2013.* http://nexa.polito.it/2013-ponti

18 Quali link alle sezioni dove vengono pubblicati i dataset, link ai siti dei Comuni e informazioni sui Comuni (es: popolazione).
19 Maggiori informazioni: http://it.dbpedia.org/
20 Reperibile all'indirizzo: http://www.magellanopa.it/bussola/
21 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"[22].

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 sate create le "Object Properties" che stabiliscono le relazioni tra le varie classi:

1. HaPubblicato: con dominio la classe "PubblicaAmministrazione" e range la classse "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 visibilie 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.

22 http://en.Wikipedia.org/Wiki/Prot%C3%A9g%C3%A9_(software)

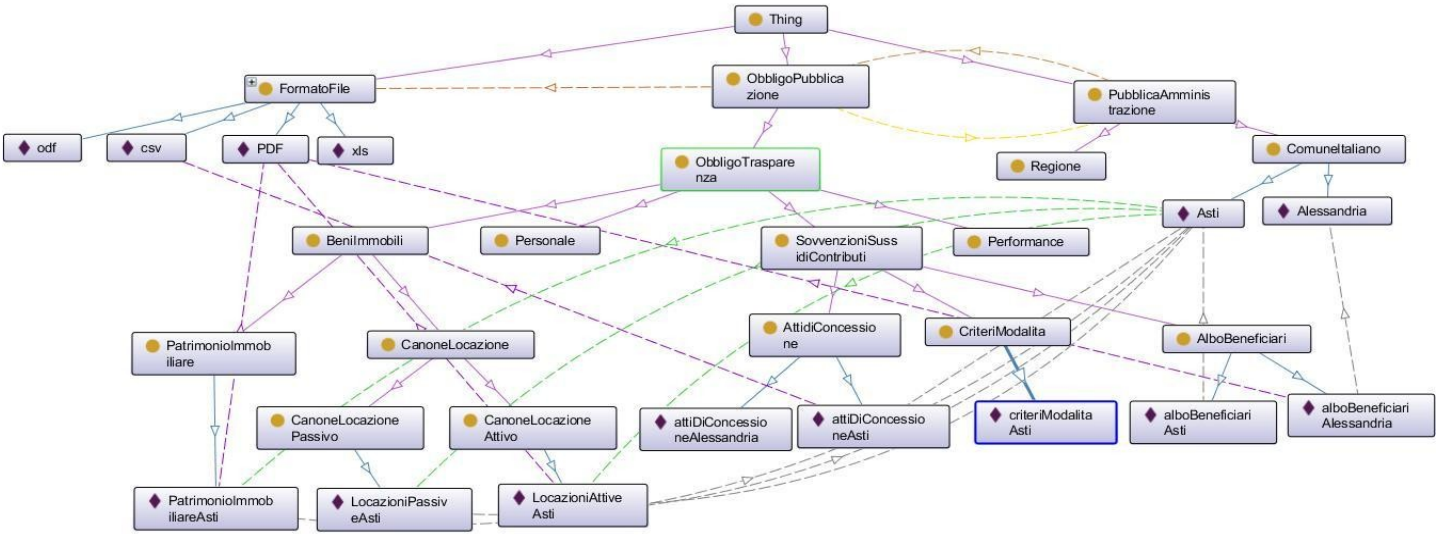

figura 1

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 propeties) e successivamente le categorie (classi) a cui esse appartengono come descritto nell'ontologia semantica[23].

23 La Wiki è disponibile a questo indirizzo: https://trasparenza.nexacenter.org/w/index.php/Pagina_principale

I 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
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 "erroriFilePbblicato" 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 "ProprietaGenerali" 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 Queries

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 atributi 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 automaticmante (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 atributi previsti per legge

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