



UNIVERSITÀ DI PISA

Corso di Laurea in Informatica Umanistica

RELAZIONE

Studio sugli Open Data relativi alle attrazioni turistiche italiane

Candidato: Elisabetta Triolo

Relatore: *prof. Andrea Marchetti, ing. Angelica Lo Duca*

Correlatore: *prof. Paolo Macchia*

Anno Accademico 2016-2017

INDICE

| | |
|--|-----------|
| CAPITOLO 1 | |
| INTRODUZIONE | 4 |
| CAPITOLO 2 | |
| TURISMO: CONCETTI E DEFINIZIONI DI BASE | 6 |
| 2.1 Forme di turismo. | 7 |
| 2.2 Definizione di destinazione turistica. | 8 |
| 2.2.1 Il prodotto turistico. | 11 |
| 2.3 Definizione del concetto di attrazione turistica. | 11 |
| CAPITOLO 3 | |
| IL CONCETTO DI ATTRAZIONE TURISTICA NEL CONTESTO DEL WEB SEMANTICO | 14 |
| 3.1 Stato dell'arte sulle ontologie relative alle attrazioni turistiche. | 15 |
| 3.2 L'attività di Schema.org. | 22 |
| 3.3 La proposta di modifica di TouristAttraction. | 26 |
| 3.3.1 Una nuova definizione e nuove proprietà per TouristAttraction. | 26 |
| 3.3.2 Sviluppi futuri. | 30 |
| CAPITOLO 4 | |
| GLI OPEN DATA RELATIVI ALLE ATTRAZIONI TURISTICHE ITALIANE | 32 |
| 4.1 Gli Open Data in Italia. | 34 |
| 4.2 Selezione degli Open Data sulle attrazioni turistiche italiane. | 35 |
| 4.3 Qualità degli Open Data sulle attrazioni turistiche italiane. | 37 |
| 4.3.1 L'assenza di uno standard. | 39 |
| 4.3.2 Il formato dei dati. | 39 |
| 4.3.3 Problemi redazionali. | 41 |
| 4.4 Raccolta automatica dei dati e integrazione nella piattaforma Tourpedia. | 42 |
| 4.4.1 Mapping per le attrazioni turistiche. | 44 |
| 4.4.2 Estensione del Tourpedia Mapping Language. | 46 |
| 4.4.3 Arricchimento dei dati. | 48 |
| CAPITOLO 5 | |
| COSTRUZIONE DI UN'ONTOLOGIA SULLE ATTRAZIONI TURISTICHE A PARTIRE DAGLI OPEN DATA | 49 |
| 5.1 Criteri per la definizione dell'ontologia. | 50 |
| 5.2 Primo Caso d'uso: un'ontologia per la regione Marche. | 52 |
| 5.3 Secondo Caso d'uso: un'ontologia per la regione Sardegna. | 53 |
| CONCLUSIONI | 56 |

| | |
|---|-----------|
| Appendice | 58 |
| Tabella 1A: Dataset raccolti per ogni regione italiana. | 58 |
| Bibliografia | 71 |
| Sitografia | 74 |

CAPITOLO 1

INTRODUZIONE

L'obiettivo di questo lavoro di tesi consiste nello studio degli Open Data, forniti dalle varie regioni e province italiane, relativi alle attrazioni turistiche. Al fine di condurre uno studio approfondito, in primo luogo è stata data una definizione di turismo ed attrazione turistica. Questa attività è consistita principalmente nella collaborazione con il gruppo di lavoro *W3C Tourism Structured Web Data Community*, il cui obiettivo consiste nell'estendere l'ontologia definita dal portale *Schema.org* con concetti relativi al turismo. Il primo risultato ottenuto è stato proprio la proposta di estensione di *Schema.org* riguardo al concetto di *TouristAttraction*. Tale proposta è stata accettata dalla community di *Schema.org* e molto probabilmente sarà rilasciata nella versione 3.3 dello stesso portale.

Alla luce del concetto di attrazione turistica così definito, il lavoro di tesi si è poi concentrato nella selezione dei dataset open, forniti dalle varie regioni e province, che rientrassero nella definizione di attrazione turistica così data. I dati raccolti sono stati integrati nella piattaforma *Tourpedia*, che già conteneva più di 70.000 dati relativi alle strutture ricettive. In totale sono state raccolte 40.599 attrazioni turistiche.

Come ultimo passo, il lavoro si è spostato sulla definizione di un'ontologia relativa alle attrazioni turistiche, specifica per le regioni Marche e Sardegna.

La metodologia adottata nel corso del lavoro di tesi si è basata su un approccio combinato: l'elaborazione dei concetti teorici è stata sempre accompagnata dall'implementazione di software appropriato, sviluppato nel linguaggio *PHP*. Tutti i dataset sono stati memorizzati in un database No-SQL (*MongoDB*).

Le principali fonti utilizzate per lo sviluppo del presente lavoro di tesi sono state le seguenti. Relativamente alla definizione dei concetti teorici riguardanti il turismo, sono state usate le pubblicazioni della *United Nations World Tourism Organization*.

Per approfondimenti sul web semantico e sui linguaggi *OWL* e *RDF*, sono state usate le informazioni fornite online dal *W3C Semantic Web*. La documentazione fornita da *Schema.org* è stata utile per lo studio di *Microdata* e la realizzazione della nuova

definizione di *TouristAttraction*. Per la definizione del concetto di Open Data sono state usate le linee guida fornite dalla *Open Knowledge Foundation*. Per l'inserimento dei dati nella piattaforma *Tourpedia* è stato fondamentale l'uso del manuale *PHP* online e, in particolare, della parte di documentazione relativa a *MongoDB*.

Il presente lavoro è organizzato come segue. Nel capitolo 2 sono delineati i concetti base relativi al turismo e, in particolare, vengono descritte le diverse forme di turismo (2.1) e vengono date una definizione di destinazione turistica (2.2) e del concetto di attrazione turistica (2.3).

Il capitolo 3 si occupa della definizione del concetto di *TouristAttraction* all'interno del vocabolario di *Schema.org* (3.3), passando per una descrizione di alcune ontologie sulle attrazioni turistiche già esistenti (3.1) e degli strumenti e linguaggi specifici usati da *Schema.org* (3.2).

Nel capitolo 4, dopo un breve quadro sulle principali iniziative italiane nel campo degli *Open Data* (4.1), vengono descritti i criteri per la raccolta degli *Open Data* relativi alle attrazioni turistiche italiane (4.2). Di essi viene fornita una valutazione qualitativa (4.3) che riguarda principalmente l'assenza di uno standard (4.3.1), il formato dei dati (4.3.2) e l'eventuale presenza di problemi redazionali (4.3.3). Infine, viene descritto il processo di raccolta e integrazione dei dati nella piattaforma *Tourpedia* (4.4), ponendo l'attenzione sulle sue principali componenti, ovvero il *Tourpedia Data Model* (4.4.1), il *Tourpedia Mapping Language* (4.4.2) e l'eventuale arricchimento dei dati con informazioni aggiuntive (4.4.3).

Il capitolo 5, dopo la definizione di specifici criteri per la creazione di un'ontologia relativa alle attrazioni turistiche italiane (5.1), descrive le due ontologie realizzate per la regione Marche (5.2) e per la regione Sardegna (5.3).

CAPITOLO 2

TURISMO: CONCETTI E DEFINIZIONI DI BASE

Tentare di dare una definizione di attrazione turistica è un processo che si è rivelato, in realtà, più difficile di quello che poteva sembrare, almeno in un primo momento.

In primo luogo, infatti, esso necessita almeno di una basilare conoscenza del contesto in cui si colloca un'attrazione turistica, ovvero il complesso mondo del turismo. È necessario cercare di definire cosa può essere effettivamente considerato turismo, chi è il turista, quali sono gli altri elementi che insieme alle attrazioni formano il mondo del turismo, se è possibile classificare le attrazioni e se esistono già delle classificazioni più o meno ufficiali. È a questo scopo che è dedicata questa sezione per la quale sono state usate principalmente le linee guida fornite nelle sue svariate pubblicazioni dalla *United Nations World Tourist Organization (UNWTO)*, agenzia specializzata delle Nazioni Unite, fondata nel 1974, che si occupa del coordinamento delle politiche turistiche¹.

Nel documento *Glossary of tourism terms*, redatto dalla *United Nation World Tourist Organization (UNWTO)*, si definisce turismo “un fenomeno sociale, culturale ed economico che implica il movimento di persone fuori dal loro *ambiente quotidiano* per scopi personali o economico/professionali (AA. VV. 2014)”.

Il documento *International Recommendations for Tourism Statistics 2008*, redatto sempre dalla *UNWTO* con lo scopo di fornire le linee guida e uno standard internazionale per la realizzazione di ricerche statistiche sul turismo, fornisce una ancora più specifica definizione, ponendo l'attenzione su due aspetti fondamentali senza i quali è impossibile parlare di turismo. Il primo è la permanenza nel nuovo ambiente per meno di un anno, mentre il secondo è lo spostamento fuori dal cosiddetto *usual environment* (in italiano, *ambiente quotidiano*), termine con il quale si intende l'area geografica, non necessariamente contigua, all'interno della quale un individuo conduce la sua vita quotidiana.

A questi due fattori-chiave se ne unisce un altro di natura economica, ovvero *l'assenza di un rapporto di lavoro dipendente* presso un datore di lavoro residente

¹ <http://www2.unwto.org/>

nel luogo oggetto dello spostamento, mentre sono ammessi semplici rapporti di negoziazione, come la compravendita (AA. VV. 2010, p. 10).

Sono questi gli elementi fondamentali che permettono di distinguere i concetti di *viaggio* e *viaggiatore* da quelli di *turismo* e *visitatore*. Il viaggio, infatti, è un semplice spostamento da un luogo a un altro differente, per qualsiasi scopo e di qualsiasi durata, e il turismo non ne è che una sottocategoria, così come il visitatore non è che un particolare tipo di viaggiatore o meglio, un viaggiatore che si sposta verso un luogo che si trova fuori dal suo *ambiente quotidiano*, per meno di un anno e con uno scopo personale o economico/professionale che non sia, tuttavia, instaurare un rapporto di lavoro di tipo dipendente con un datore di lavoro che risiede nel luogo oggetto della visita. I visitatori, inoltre, vengono a loro volta classificati come *turisti*, se il loro spostamento include almeno un pernottamento nell'area oggetto della visita, oppure come *escursionisti*, in caso contrario (AA. VV. 2010, p. 10).

Come si può notare, la definizione di turismo fornita dalla *UNWTO* è fatta in termini di domanda e non di offerta, cioè si basa sulle attività e sul tipo del viaggiatore che si sposta in un luogo per soddisfare uno o più scopi personali, o di piacere, o di tipo professionale, e non considera i vari prodotti e attività offerte dall'area oggetto della visita.

Tale tipo di definizione è tuttavia, necessaria poiché non sono le caratteristiche dei prodotti offerti che permettono di identificare uno spostamento come turismo, ma le caratteristiche del viaggiatore che si sposta dal suo ambiente di vita quotidiana per un periodo di tempo inferiore ad un anno e senza il cui scopo, di tipo personale e non, lo spostamento non avrebbe ragione di esistere.

2.1 Forme di turismo.

La *UNWTO* fornisce anche una classificazione delle diverse forme di turismo (AA. VV. 2010, p. 15), indicando tre forme-base di turismo, ovvero:

- *turismo domestico*, che consiste nello spostamento di un visitatore all'interno dell'area di residenza;
- *turismo in entrata*, che consiste nello spostamento in una determinata area di un visitatore non residente in quell'area;

- *turismo in uscita*, che consiste nello spostamento di un visitatore fuori dal paese di residenza.

Queste tre forme di turismo possono a loro volta combinarsi costituendo altre forme derivate tra le quali la *UNWTO* ne individua tre principali:

- *turismo interno*, che comprende il turismo domestico e il turismo in entrata, ovvero le attività di visitatori residenti e non residenti all'interno di una determinata area oggetto del viaggio;
- *turismo nazionale*, che comprende il turismo domestico e il turismo in uscita, ovvero le attività dei visitatori residenti in una data area all'interno di quell'area o al suo esterno;
- *turismo internazionale*, che comprende il turismo in entrata e quello in uscita, ovvero data una determinata area, le attività dei visitatori non residenti all'interno di quell'area e le attività all'esterno di quell'area dei visitatori residenti nell'area.

È importante notare, inoltre, il fatto che nell'individuazione delle diverse forme di turismo, la *UNWTO* tiene conto non della nazionalità del visitatore, ma della residenza e definisce residente in una data area un individuo il cui centro dei principali interessi economici si trova in quell'area (AA. VV. 2010, p. 95).

2.2 Definizione di destinazione turistica.

In generale, una destinazione turistica può essere definita come il luogo centrale nella decisione di intraprendere lo spostamento da parte del visitatore (AA. VV. 2010, p. 13). È per visitare un determinato luogo, infatti, che il visitatore attua uno spostamento che può essere di tipo turistico.

La *Committee on Tourism and Competitiveness (CTC)*, un organo secondario del Consiglio Esecutivo dell'*UNWTO* creato nel 2013, ha fornito una più dettagliata definizione di destinazione turistica, indicando come tale

uno spazio fisico con o senza confini amministrativi e/o analitici nel quale un visitatore può pernottare. È un *cluster* di *prodotti* e *servizi*, e di *attività* ed *esperienze* (...) Una destinazione incorpora diversi stakeholders e può ramificarsi in modo da formare destinazioni più estese (AA. VV. 2016, p. 13).

È bene notare, in primo luogo, il fatto che la definizione di destinazione turistica è molto ampia riguardo al tipo di luogo che può essere una destinazione, indicandolo genericamente come uno spazio fisico che può o non può essere sottoposto a confini amministrativi e può anche non avere confini precisi. Questa descrizione molto ampia corrisponde a una delle caratteristiche fondamentali di una destinazione turistica, ovvero il fatto che essa non è tale sulla base di un qualche confine amministrativo o stabilito a priori, ma diventa tale quando il mercato ne acquisisce consapevolezza, ovvero quando viene riconosciuta sul mercato, o per un processo di autodeterminazione da parte degli stakeholders di un determinato territorio che si propongono sul mercato come destinazione, o per un processo di riconoscimento da parte dei visitatori (Marchioro 2011, p. 7).

2.2.1 Il prodotto turistico.

È sempre la *CTC* a fornire anche una definizione di *prodotto turistico*, indicando con questo termine

una combinazione di elementi *tangibili* e *intangibili*, come *risorse naturali, culturali* e *artificiali, attrazioni, strutture, servizi* e *attività* attorno a uno specifico centro di interesse che (...) crea una complessiva esperienza di visita che include *aspetti emozionali* per il potenziale fruitore. Un prodotto turistico ha un prezzo, è venduto tramite i canali di distribuzione e ha un ciclo di vita (AA. VV. 2016, p. 13).

È necessario precisare che la definizione di prodotto turistico fatta dalla *CTC*, nata soprattutto per scopi di tipo statistico/economico, tiene conto solo di quei prodotti che sono venduti attraverso i canali del mercato, ma nelle *International Recommendations for Tourism Statistics 2008*, la *UNWTO* specifica che i prodotti turistici si possono dividere essenzialmente in due categorie, ovvero *prodotti di consumo* e *prodotti di non consumo* (AA. VV. 2010, p. 41), indicando con questi ultimi quelli la cui natura non è il consumo di beni e servizi, che non possono essere consumati come beni e servizi all'interno del mercato, estendendo, dunque, notevolmente, l'insieme di elementi che possono essere considerati un prodotto turistico.

È possibile notare che, così come la definizione di destinazione turistica, anche quella di prodotto turistico è una definizione abbastanza ampia che comprende una

vasta quantità di elementi eterogenei, come mezzi di trasporto, strutture ricettive, attrazioni turistiche, caratteristiche del posto visitato, uffici di informazione turistica e di accoglienza, infrastrutture, in grado, però, nel loro insieme di fornire al visitatore un'esperienza anche di tipo emozionale.

La stessa *UNWTO* specifica che non esiste uno standard per la classificazione dei prodotti turistici, ma suggerisce una lista dei prodotti turistici di consumo raggruppati per scopo della visita. Essa è costruita sulla base della *Classification of Individual Consumption by Purpose (COICOP)*, una classificazione internazionale dei consumi, legata alla *Central Product Classification (CPC)*, uno standard messo a punto dalle Nazioni Unite per la classificazione statistica dei prodotti, e della *International Standard Industrial Classification of All Economic Activities (ISIC)*, una classificazione delle attività economiche definita sempre dalle Nazioni Unite (AA. VV. 2010, pp. 35, 39).

La *COICOP* individua otto classi fondamentali in cui è possibile raggruppare i prodotti turistici, ovvero:

1. *Package travel, package holidays and package tours;*
2. *Accommodation;*
3. *Food and drink;*
4. *Local transport;*
5. *International transport;*
6. *Recreation, culture and sporting activities;*
7. *Shopping;*
8. *Others.*

La classe *Others* comprende tutti i prodotti di consumo che non rientrano nelle classi precedenti.

I prodotti sono classificati sulla base dello scopo, del fine della visita e la scelta di tale criterio è dettata dal fatto che, come specificato in 2, è proprio lo scopo a determinare lo spostamento del visitatore e, quindi, il flusso turistico e il consumo del prodotto turistico, tutte cose che non sussisterebbero in assenza di uno scopo principale che determini il movimento.

All'interno dell'ampissima lista, per ogni classe della *COICOP* sono specificate una o più sottoclassi di prodotti appartenenti alla più vasta classificazione *CPC* e per ogni

sottoclasse è indicata l'attività ad essa corrispondente, definita in termini di tipo di produzione di una data sottoclasse, di ciò che la sottoclasse produce, così come definita dal rispettivo codice *ISIC*.

Ad esempio, la classe *Food and drink* appare definita come segue:

| CPCVer.2 subclass | Description | Categorization (1) | Corresponding activity ISIC Rev.4 |
|--------------------------|---|---------------------------|--|
| Food and drink (c) | | | |
| 63310 | Meal serving services with full restaurant services | X | 5610 |
| 63320 | Meal serving services with limited services | X | 5610 |
| 63399 | Other food serving services | X | 5610 |
| 63400 | Beverage serving services | X | 5610 |

(Tabella 1. Lista dei prodotti turistici di consumo raggruppati per lo scopo della visita)

2.3 Definizione del concetto di attrazione turistica.

All'interno del quadro tracciato finora appare ora necessario fornire una più precisa definizione di cosa si intende per attrazione turistica. In 2.2, infatti, l'attrazione turistica è stata menzionata come un tipo di prodotto turistico, ma è necessario specificare quali sono le caratteristiche che permettono di distinguerla da altri tipi di prodotti come le strutture ricettive o le infrastrutture di una certa destinazione turistica.

Harris e Howard, nel loro *Dictionary of Travel & Tourism Hospitality Terms*, descrivono un'attrazione turistica come

una *caratteristica fisica* (cors. agg) o *culturale* (cors. agg) di un particolare luogo che viaggiatori individuali o turisti percepiscono come in grado di soddisfare uno o più dei loro specifici *bisogni* (cors. agg) ricreativi. Alcune caratteristiche possono trovarsi in natura (ad esempio, clima, cultura, vegetazione o paesaggio), oppure possono essere specifiche di una location, come uno spettacolo teatrale, un museo o una cascata (Harris e Howard 1996).

Medlik, nel suo *Dictionary of Travel, Tourism & Hospitality*, dà una definizione simile indicando come attrazioni turistiche “gli elementi del prodotto turistico che attraggono visitatori e determinano la scelta di visitare un posto piuttosto che un altro (Medlik 2003)”. Medlik fornisce una generica distinzione delle attrazioni in *site attractions*, quando il posto in sé, ad esempio per il suo clima, paesaggio o storia, è il principale motivo della visita, e *event attractions*, quando gli eventi organizzati in un dato luogo, come festival, performances, eventi sportivi, costituiscono il principale motivo della visita. Inoltre, fa anche una distinzione tra *attrazioni naturali*, come un vulcano o una foresta, e *attrazioni artificiali* come un museo, o una pittura rupestre. Sulla base di queste due definizioni è possibile individuare una caratteristica fondamentale che un elemento deve possedere per essere considerato un’attrazione turistica, ovvero deve riuscire a soddisfare la domanda del visitatore, il bisogno che ha determinato lo spostamento e, quindi, deve essere il *motivo principale dello spostamento* del visitatore in un particolare luogo/destinazione turistica.

Al di là di ciò, è davvero difficile individuare dei criteri che permettano di identificare un particolare elemento come attrazione turistica e, a parte la generica suddivisione in naturali e artificiali, non esiste uno standard per la classificazione delle attrazioni turistiche in grado di rappresentare la vastità e eterogeneità degli elementi che possono rientrare nell’insieme delle attrazioni. Si potrebbe pensare di utilizzare la lista dei prodotti turistici di consumo suggerita dall’*UNWTO* (v. 2.2.1), ma essa costituisce una classificazione di tutti i prodotti turistici e non solo delle attrazioni e, inoltre, tiene conto solo dei prodotti di consumo, escludendo un vasto insieme di elementi.

In sostanza, qualsiasi elemento di tipo naturale o artificiale, fisico o culturale che caratterizza un dato luogo può essere un’attrazione turistica se determina lo spostamento di un visitatore verso quel luogo al fine di soddisfare un suo bisogno di tipo personale e/o di altro tipo. Ciò implica che anche un’entità a cui di solito non si pensa come attrazione, come ad esempio un centro commerciale a Milano, potrebbe diventarlo nel momento in cui un determinato visitatore, uscendo dal suo ambiente di vita quotidiana (ad esempio, la sua città in Toscana), vi si rechi in cerca di un

articolo, come un oggetto di elettronica, in vendita in anteprima nazionale solo in quel centro, facendo del punto vendita il principale motivo del suo spostamento.

Se è quasi impossibile realizzare una classificazione standard delle attrazioni turistiche, è invece, possibile usare dei criteri per stabilire la qualità di un'attrazione turistica.

Varra (2012, p. 21) parla di tre criteri fondamentali, ovvero, la *replicabilità della risorsa*, cioè la possibilità che possa scorporarsi per innestarsi altrove, *l'unicità della risorsa* cioè la capacità di qualificare il territorio differenziandolo dagli altri, e *l'importanza riconosciuta* dagli stakeholder nella creazione di valore, cioè il contributo ad essa associato nella produzione globale di valore. In base a questi criteri, le attrazioni possono essere *embedded* o *located* (se si considera il criterio di replicabilità), *differenzianti* o *anonime* (se si considera il criterio di unicità), *centrali* o *periferiche* (se si considera il criterio di importanza riconosciuta).

SempreVarra (2012, p. 21-22) parla poi, della *funzione d'uso* delle attrazioni: un'attrazione può avere una o più funzioni d'uso (ad esempio, funzione di svago, di divertimento, di intrattenimento, ecc...), ma di solito i fruitori non si servono di una sola funzione ma di più funzioni integrate in *processi di fruizione* (ad esempio, processo di svago= parco+cinema). La predominanza di alcune funzioni sulle altre nei diversi processi di fruizione determina la *vocazione* di un territorio e la sintesi delle vocazioni costruisce l'*identità* di una destinazione, connotandola in maniera netta rispetto alle altre. Ad esempio, se la funzione di svago predomina nei diversi processi di fruizione del visitatore di una data destinazione, quest'ultima avrà la vocazione di svago e divertimento, e tale vocazione contribuirà, insieme alle altre vocazioni, a formare l'identità della destinazione.

CAPITOLO 3

IL CONCETTO DI ATTRAZIONE TURISTICA NEL CONTESTO DEL WEB SEMANTICO

Il concetto di *web semantico* (*semantic web*, in inglese) è un concetto relativamente recente elaborato dal *World Wide Web Consortium (W3C)* con lo scopo di fornire uno standard per la rappresentazione dei contenuti semantici sul web. L'obiettivo, sostanzialmente, è quello di creare un'estensione dell'attuale web, ossia un web fatto di dati², in cui, cioè, sia possibile associare ai dati un significato formale e costruire le relazioni semantiche che intercorrono tra di essi. Si tratta, dunque, di creare delle ontologie del web.

Un'ontologia, in senso informatico, è un modello formale per la concettualizzazione di un dominio di interesse. Più esplicitamente, si tratta di un modello che definisce, tipi, proprietà e interrelazioni delle entità che esistono in un particolare dominio di interesse.

Le tecnologie basate sul web semantico potrebbero avere numerose e utili applicazioni come l'integrazione di dati di diversa provenienza e formato, il miglioramento dei processi di ricerca dei dati, nella catalogazione per descrivere i dati e le relazioni tra i dati appartenenti ad esempio, a una libreria digitale, per raggruppare in collezioni pagine web che rappresentano semanticamente un unico documento.

Il web semantico si basa sull'uso di *RDF*, *Resource Description Framework*³, che è un modello standard per la rappresentazione dei contenuti semantici sul web e, in particolare, su *RDFs*, *Resource Description Framework Schema*, e su *OWL*, *Web Ontology Language*, due linguaggi che ne rappresentano le estensioni.

² <https://www.w3.org/2001/sw/SW-FAQ#WhatIsTheSW>

³ <https://www.w3.org/RDF/>

3.1 Stato dell'arte sulle ontologie relative alle attrazioni turistiche.

Negli ultimi anni, il fenomeno turistico ha suscitato un interesse crescente all'interno della comunità scientifica, dell'informatica e sul web. Basti pensare alle numerose applicazioni mobile o ai siti web di viaggi che sono sorti negli ultimi anni per guidare gli utenti nella scelta dell'itinerario da seguire, della struttura ricettiva in cui alloggiare, delle attrazioni da visitare e molto altro. Da non scordare è anche la nascita di portali per la visita virtuale di luoghi di attrazione come i musei.

A questo andamento corrisponde la crescente necessità di definire ontologie, ovvero modelli astratti relativi a uno specifico dominio di applicazione, in questo caso rappresentato dal turismo.

Un esempio di ontologia è l'*HarmoNET Tourism Ontology*, realizzata da *HarmoNET* (*Harmonisation Network for the Exchange of Travel and Tourism Information*), una comunità nata nel 2006 e aperta a tutti coloro, singole persone o organizzazioni, interessate al fenomeno turistico. L'ontologia descrive il concetto di attrazione turistica, indicata come *Attraction*, con una serie di attributi tra i quali *name*, *type*, *location* e *description*, e propone una lunghissima lista di valori da associare al tipo, tra i quali troviamo, ad esempio *church*, *castle*, *monument*, *museum*, *national park*, *pyramid* (Höpken e Ciaran 2008).

Un'altra ontologia specifica per il turismo è *OnTour*, realizzata nel 2004 dal *Digital Enterprise Research Institute (DERI)*. Essa contiene oltre alla definizione di usuali concetti turistici come *Accommodation* e *Infrastructure*, modelli per descrivere dati geografici come le coordinate GPS (Prantner 2004).

Diversi sono anche gli articoli scientifici dedicati alla definizione di ontologie turistiche. Tra quest'ultime si possono citare quella suggerita dalla facoltà di ingegneria dell'università di Tehran (Bahramian e Abbaspour 2015), oppure quella definita dall'Università Statale di San Paolo in Brasile, che si basa su una suddivisione tra *Natural Attraction* e *Cultural Attraction* (Tosqui-Lucks e Dias Da Silva 2012).

In campo italiano, importante è la *QUALL-ME Ontology*, un'ontologia del turismo per il *question answering (QA)* realizzata da un gruppo coordinato dalla *Fondazione Bruno Kessler (FBK)* in collaborazione con alcune aziende italiane impegnate nel settore informatico (*Comdata S.p.A.*, *Ubiest S.p.A.* e *Waycom S.r.l.*). Nell'ontologia sono individuate cinque sottocategorie di attrazione turistica, ovvero attrazioni culturali, naturali, religiose, parchi tematici, zoo e acquari⁴.

Un'altra ontologia, che, tuttavia, non riguarda solo il settore turistico ma è di più ampia scala, è *WordNet*, un database lessicale della lingua inglese elaborato presso l'Università di Princeton. All'interno del database, nomi, verbi, aggettivi e avverbi, sono raggruppati in *synsets*, ovvero gruppi di termini che hanno significato simile. I *synsets* sono legati tra di loro da relazioni di tipo semantico e lessicale, tra cui iperonimia ed iponimia⁵. Ad esempio, per il termine *church* è possibile trovare in *WordNet* *place of worship* tra i suoi iperonimi, ovvero termini di cui *church* rappresenta una particolare specie; tra gli iponimi, cioè termini che rappresentano una particolare specie di *church*, invece, si può trovare *cathedral*.

A fornire possibili modelli di classificazione degli elementi appartenenti al mondo del turismo sono anche i siti web e le applicazioni di viaggio, come *TripAdvisor*. Il portale presenta una lunghissima lista di sottocategorie (circa 217), organizzate in 20 categorie che comprendono non solo vere e proprie attrazioni ma anche servizi che arricchiscono la visita come *Bar e pub*, *Karaoke*, *Pianobar*, *Caffetterie*⁶. Tornando all'ambito italiano, esistono, riguardo ad alcune categorie di elementi che possono essere attrazioni turistiche, modelli di catalogazione e classificazione ufficiali.

Tra questi, ad esempio, si ha l'*Ateco 2007*, una classificazione alfa-numerica delle attività economiche adottata dall'*Istat* per le rilevazioni statistiche di carattere economico. Tra le varie sezioni, indicate da una lettera, è possibile trovare la sezione G, dedicata al commercio all'ingrosso e al dettaglio, che con le sue numerose divisioni fornisce una categorizzazione delle attività economiche relative allo shopping, che può costituire un elemento di attrazione⁷.

⁴ <http://qallme.fbk.eu/index.php?location=ontology>

⁵ <https://wordnet.princeton.edu/>

⁶ <https://developer-tripadvisor.com/content-api/business-content/categories-subcategories-and-types/>

⁷ <https://www.istat.it/it/strumenti/definizioni-e-classificazioni/ateco-2007>

Ancora l'*Istat*, a fini statistici, fornisce una classificazione dei musei e degli istituti simili in base alla tipologia individuando ben 28 categorie che comprendono *Archeologia, Storia, Scienza e tecnica, Etnografia e antropologia, Architettura fortificata o militare*⁸.

Anche per le biblioteche, l'*Istat* fornisce una classificazione in base alla tipologia funzionale: esse sono suddivise in *Nazionali, Istituzioni scolastiche, Istituti di istruzione superiore, Pubbliche, Speciali, Importanti non specializzate*⁹.

A fornire un modello per la catalogazione dei beni culturali italiani è l'*ICCD (Istituto Centrale per il Catalogo e la Documentazione)*. Nella descrizione del metodo di catalogazione del patrimonio culturale italiano, l'*ICCD* individua le categorie in cui possono suddividersi i beni culturali italiani. Essi si dividono in beni mobili, immobili e immateriali e possono appartenere a nove diversi settori disciplinari (beni archeologici, beni architettonici e paesaggistici, beni demoetnoantropologici, beni fotografici, beni musicali, beni naturalistici, beni numismatici, beni scientifici e tecnologici, beni storici e artistici) ulteriormente suddivisibili in sottocategorie (per un totale di 30) (Mancinelli 2017).

Modelli per la classificazione di altre entità che possono diventare attrazioni turistiche possono essere estratti da decreti e leggi che li riguardano e che, in genere, forniscono delle definizioni precise per gli elementi da essi regolati. Un esempio è il *Decreto Legislativo 22 gennaio 2004 n.2*¹⁰, che regola la disciplina in materia di attività cinematografiche. L'articolo 22, comma 2 fornisce le definizioni di sala cinematografica, cinema-teatro, multisala e arena. Altro esempio è il *Decreto Ministeriale 19 agosto 1996*¹¹, riguardante i metodi di prevenzione degli incendi nei luoghi di pubblico spettacolo. All'articolo 1 viene fornita una classificazione di tutti quei locali che costituiscono luoghi di pubblico spettacolo e a cui la normativa si applica.

Come si può notare, si tratta di un insieme di risorse eterogeneo poiché esse presentano diverse differenze le une dalle altre. In primo luogo, infatti, una parte di esse non è nata con lo scopo di descrivere il mondo turistico, ma per altri fini come

⁸ <https://www.istat.it/it/archivio/167566>

⁹ <https://www.istat.it/it/archivio/195678>

¹⁰ <http://www.camera.it/parlam/leggi/deleghe/04028dl.htm>

¹¹ <http://www.vigilfuoco.it/asp/ReturnDocument.aspx?IdDocumento=52>

quello legislativo o statistico. Si tratta dei decreti ministeriali e legislativi, della classificazione dei beni culturali fornita dall'*ICCD*, dell'*Ateco 2007* e delle altre classificazioni fornite dall'*Istat*.

Nelle altre ontologie, inoltre, anche se relative al mondo del turismo, il concetto di attrazione turistica viene trattato in maniera spesso troppo differente.

L'ontologia *OnTour*, ad esempio, non presenta addirittura la definizione di una classe specifica per le attrazioni turistiche. Si hanno le classi *Event*, *Activity*, *Location* e *Infrastructure*, che in qualche modo possono essere legate al concetto di attrazione turistica ma per le quali esso non sembra sia effettivamente applicabile. Anche l'ontologia proposta dall'Università di Teheran non definisce una specifica classe relativa alle attrazioni turistiche, ma parte dalla classe *Tourism* a cui appartengono le sottoclassi *Nature*, *Culture*, *Sports* e *Leisure*.

Importante è anche considerare come il concetto di attrazione viene sviluppato in quelle ontologie che, invece, presentano una specifica classe dedicata alle attrazioni. Riguardo a questo, pensando al concetto di attrazione turistica (2.3), quello che potrebbe essere importante valutare è quanto è eterogeneo l'insieme degli elementi che possono appartenere alla specifica classe dedicata alle attrazioni e se essi sono più o meno organizzati gerarchicamente.

In *HamoNET*, ad esempio, la classe *Attraction* ha la proprietà *type* che può assumere uno tra i circa 120 valori specificati. Si tratta di un buon numero di elementi, ma essi non sono organizzati in maniera gerarchica. Per quanto riguarda l'ontologia fornita dall'Università brasiliana di San Paolo, invece, il tipo di attrazione viene individuato attraverso le due sottoclassi *NATURAL ATTRACTION* e *CULTURAL ATTRACTION*, per le quali però, è definito un numero troppo scarso di sottoclassi (10 per le attrazioni naturali, 11 per quelle culturali) che rischia di essere troppo limitativo. A lasciare una maggiore libertà nella classificazione delle attrazioni sembra invece, essere la *QUALL-ME Ontology*, che individua come sottoclassi delle classe attrazioni pochi tipi ma più generici rispetto a quelli individuati dalle prime due ontologie.

È importante notare, inoltre, che non tutte le classificazioni più o meno legate al turismo sono effettivamente definite attraverso un linguaggio formale come *OWL*. I decreti ministeriali, le classificazioni *Istat* o la classificazione fornita da siti di viaggio come Trip-Advisor ad esempio, più che vere e proprie ontologie sono

descrizioni più o meno strutturate di un determinato dominio di interesse e la mancanza di una vera e propria definizione formale ne limita significativamente l'uso per scopi informatici.

La tabella seguente mostra, in sintesi, quali sono alcuni criteri utili per confrontare le diverse ontologie relative alle attrazioni turistiche e valutarne anche l'utilità.

| Ontologia | |
|---|---|
| Attinenza all'ambito turistico | Valutare se l'ontologia è stata creata con lo scopo di descrivere effettivamente il settore turistico, o se è stata realizzata per altri scopi, ad esempio statistici. |
| Formalizzazione | Valutare se l'ontologia è descritta formalmente attraverso un linguaggio per la definizione di ontologie come <i>OWL</i> . |
| Presenza di una classe dedicata alle attrazioni | Verificare se nell'ontologia è definita una classe specifica per la descrizione delle attrazioni turistiche. |
| Caratteristiche della classe che descrive le attrazioni | Valutare in che modo è definita la classe delle attrazioni, se esistono, ad esempio, delle sottoclassi, se esse sono organizzate in una struttura più o meno gerarchica e quanto la definizione di attrazione fornita limiti l'insieme degli elementi classificabili in base ad essa. |

(Tabella 2. Criteri di valutazione delle ontologie)

La tabella che segue, invece, applica ad ognuna delle ontologie descritte sopra, i criteri sopra definiti.

| Ontologia | Attinenza all'ambito turistico | Formalizzazione | Presenza di una classe dedicata alle attrazioni | Caratteristiche della classe che descrive le attrazioni |
|---------------------------|---------------------------------------|------------------------|--|---|
| HarmoNET Tourism Ontology | SI | SI (RDFS) | SI | La classe Attraction ha la proprietà type che può assumere uno tra i circa 120 valori specificati. Essi non sono organizzati in |

| | | | | |
|---|----|-------------|---|---|
| | | | | maniera gerarchica. |
| OnTour | SI | SI (OWL-DL) | NO | |
| Ontologia turistica dell'Università di Tehran | SI | SI | NO | |
| Ontologia dell'Università di San Paolo | SI | SI | SI | Si hanno le classi NATURAL ATTRACTION e CULTURAL ATTRACTION. Alla prima appartengono 10 sottoclassi di attrazioni, alla seconda 11, numeri troppo limitativi. |
| <i>QUALL-ME Ontology</i> | SI | SI (OWL) | SI | La classe Attraction ha le sottoclassi CulturalHeritage, NaturalHeritage, ReligiousHeritage, ThemePark, ZooAndAquarium. Meno limitativa per la classificazione delle attrazioni |
| WordNet | NO | SI | NO (è presente il synset attraction che | |

| | | | | |
|---|----|----|---|--|
| | | | non viene tuttavia, legato esplicitamente all'ambito turistico) | |
| TripAdvisor | SI | NO | SI | La classe Attraction comprende 20 categorie organizzate in circa 200 sottocategorie. |
| Ateco 2007 | NO | NO | NO | |
| Classificazion e Istat musei e istituti similari | NO | NO | NO | |
| Classificazion e Istat biblioteche | NO | NO | NO | |
| Catalogazione dei beni culturali dell'ICCD | NO | NO | NO | |
| Decreto Legislativo 22 gennaio 2004 n.2 (attività cinematografiche) | NO | NO | NO | |
| Decreto Ministeriale 19 agosto 1996 (luoghi di pubblico spettacolo) | NO | NO | NO | |

(Tabella 3. Confronto delle ontologie in base ai criteri definiti)

3.2 L'attività di Schema.org.

*Schema.org*¹² è una comunità nata con lo scopo di creare, mantenere e promuovere ontologie dei dati sul web, fornendo una strutturata classificazione semantica dei dati.

Schema.org, infatti, fornisce un *vocabolario* basato su un modello derivato da *RDFS* e che si presenta organizzato in:

- *types* (tipi), ovvero classi organizzate secondo una gerarchia in cui ognuna può essere sottoclasse dell'altra ereditandone tutte le proprietà, ma comunque tutte derivate dall'unico tipo iniziale *Thing*;
- *properties* (proprietà), ognuna della quali può avere come dominio e come range uno o più tipi, oppure un altro tipo di dato come una semplice stringa di testo, o un numero, o una data, ecc...

Il vocabolario presenta un *core*, cioè un insieme di tipi e proprietà di base, che può essere arricchito con diversi tipi di estensioni (*hosted*, *external*).

Ad esempio, il tipo *Accommodation* è una sottoclasse di *Place*, della quale eredita tutte le proprietà tra cui *address* che quindi ha come dominio *Accommodation*, mentre come range può avere o una semplice stringa di testo, oppure il tipo *PostalAddress*, come mostrato in figura 1.

¹² <http://schema.org/>

Accommodation
 Canonical URL: <http://schema.org/Accommodation>

Thing > Place > Accommodation

An accommodation is a place that can accommodate human beings, e.g. a hotel room, a camping pitch, or a meeting room. Many accommodations are for overnight stays, but this is not a mandatory requirement. For more specific types of accommodations not defined in schema.org, one can use `additionalType` with external vocabularies.

See also the [dedicated document on the use of schema.org for marking up hotels and other forms of accommodations](#).

[more...]

| Property | Expected Type | Description |
|--------------------------------------|---|--|
| Properties from Accommodation | | |
| amenityFeature | LocationFeatureSpecification | An amenity feature (e.g. a characteristic or service) of the Accommodation. This generic property does not make a statement about whether the feature is included in an offer for the main accommodation or available at extra costs. |
| floorSize | QuantitativeValue | The size of the accommodation, e.g. in square meter or squarefoot. Typical unit code(s): MTK for square meter, FTK for square foot, or YDK for square yard |
| numberOfRooms | Number or QuantitativeValue | The number of rooms (excluding bathrooms and closets) of the accommodation or lodging business. Typical unit code(s): ROM for room or C62 for no unit. The type of room can be put in the <code>unitText</code> property of the <code>QuantitativeValue</code> . |
| permittedUsage | Text | Indications regarding the permitted usage of the accommodation. |
| petsAllowed | Boolean or Text | Indicates whether pets are allowed to enter the accommodation or lodging business. More detailed information can be put in a text value. |
| Properties from Place | | |
| additionalProperty | PropertyValue | A property–value pair representing an additional characteristics of the entity, e.g. a product feature or another characteristic for which there is no matching property in schema.org. Note: Publishers should be aware that applications designed to use specific schema.org properties (e.g. http://schema.org/width , http://schema.org/color , http://schema.org/gtin13 , ...) will typically expect such data to be provided using those properties, rather than using the generic property/value mechanism. |
| address | PostalAddress or Text | Physical address of the item. |

(Figura 1. Il tipo *Accommodation* sul sito di *Schema.org*)

Chiunque desideri può proporre una modifica dei tipi e proprietà esistenti, oppure nuovi tipi e/o proprietà, iscrivendosi alla mailing-list dello *Schema.org Community Group* e/o inserendo una nuova *pull request* direttamente sulla pagina *Github* di *Schema*¹³.

I tempi di accettazione, eventuali revisione e integrazione delle nuove proposte sono di solito piuttosto lunghi e, comunque, non inferiori a un minimo di due settimane.

Sul sito web di *Schema.org* per molti dei tipi e delle proprietà sono disponibili uno o più esempi di markup semantico fatto utilizzandoli. Ogni esempio prevede del codice *HTML* senza markup e il corrispettivo markup semantico in *Microdata*, *RDFa* e *JSON-LD*.

Microdata è una specifica *HTML5* realizzata dal *Web Hypertext Application Technology Working Group (WHATWG)*, una comunità fondata da *Apple*, *Mozilla Foundation* e *Opera Software* nel 2004 con lo scopo di espandere il linguaggio *HTML* e le tecnologie ad esso legate.

¹³ <https://github.com/schemaorg/schemaorg>

Fondamentalmente, il vocabolario *Microdata* si basa sull'uso di tre attributi fondamentali che possono essere aggiunti agli elementi del codice *HTML5* per marcare i contenuti, ovvero:

- *itemscope*, che indica che il contenuto dell'elemento è un tipo;
- *itemtype*, il cui valore è un *URL* verso uno specifico tipo del vocabolario usato (in questo caso, quello di *Schema.org*) in modo da descrivere le caratteristiche del tipo indicato da *itemscope*;
- *itemprop*, che indica che il contenuto dell'elemento *HTML* è una proprietà di quel tipo e il cui valore è di solito il contenuto dell'elemento stesso, ma può essere anche il valore dell'attributo *href* di un tag `<a>`, oppure dell'attributo *src* di un tag ``, o dell'attributo *content* del tag `<meta>`. Il valore di un *itemprop* può a sua volta essere un tipo, e in questo caso, *itemprop* viene utilizzato in combinazione con *itemscope* e *itemtype*.

Il seguente esempio, tratto da *Schema.org*, mostra il markup di una porzione di una pagina *HTML* che parla di un film:

```
<div itemscope itemtype = "http://schema.org/Movie">
  <h1 itemprop="name">Avatar</h1>
  <div itemprop="director" itemscope
itemtype="http://schema.org/Person">
    Director: <span itemprop="name">James Cameron</span> (born
<span itemprop="birthDate">August 16, 1954</span>)
  </div>
  <span itemprop="genre">Science fiction</span>
  <a href=" ../movies/avatar-theatrical-trailer.html"
itemprop="trailer">Trailer</a>
</div>
```

Anche *RDFa* (*Resource Description Framework in Attributes*), come *Microdata*, è un'estensione di *HTML5*, basata su *RDF*, a cui un gruppo guidato dal *W3C* (*World Wide Web Consortium*) ha iniziato a lavorare circa dal 2004. Come *Microdata*, anche *RDFa* usa una serie di attributi che si aggiungono ai tag *HTML* per marcare i contenuti, tra cui i principali sono:

- *vocab*, il cui valore è l'*URL* verso un vocabolario (*Schema.org*, in questo caso);
- *typeof*, che indica che il contenuto dell'elemento è un tipo e il cui valore è il nome di uno specifico tipo appartenente al vocabolario;

- *property*, che indica che il contenuto dell'elemento è una *proprietà* di quel tipo e (come per *itemprop* in *Microdata*) il cui valore è di solito il contenuto dell'elemento stesso, ma può essere anche il valore dell'attributo *href* di tag `<a>`, oppure dell'attributo *src* di un tag ``, o dell'attributo *content* del tag `<meta>`. Il valore di un *property* può a sua volta essere un tipo, e in questo caso, l'attributo viene utilizzato in combinazione con *typeof*.

L'esempio seguente mostra la stessa porzione di pagina marcata in *Microdata* precedentemente, con un markup fatto utilizzando *RDFa*:

```
<div vocab = "http://schema.org/" typeof="Movie">
  <h1 itemprop="name">Avatar</h1>
  <div property="director" typeof="Person">
    Director: <span itemprop="name">James Cameron</span> (born
  <span property="birthdate">August 16, 1954</span>)
  </div>
  <span itemprop="genre">Science fiction</span>
  <a href=" ../movies/avatar-theatrical-trailer.html"
property="trailer">Trailer</a>
</div>
```

JSON-LD (JavaScript Object Notation for Linked Data) è un'estensione di *JSON* nata intorno al 2010 con l'obiettivo di usare *JSON* per creare *Linked Data*, ovvero un modello di dati strutturati collegati semanticamente tra loro.

Gli elementi fondamentali della sintassi specificata da *JSON-LD* sono:

- *@context*, che indica l'*URL* o una serie di definizioni di termini che corrispondono al vocabolario usato per la codifica dei contenuti;
- *@type*, che indica uno specifico tipo appartenente al vocabolario.

Il seguente esempio mostra gli stessi dati dei primi 2 codificati usando *JSON-LD*:

```
<script type="application/ld+json">{
  "@context": "http://schema.org",
  "@type": "Movie",
  "name": "Avatar",
  "trailer": " ../movies/avatar-theatrical-trailer.html",
  "director": {
    "@type": "Person",
    "name": "James Cameron ",
    "birthDate": "August 16, 1954"
  }
}
</script>
```

3.3 La proposta di modifica di *TouristAttraction*.

Il *core* del vocabolario di *Schema.org* contiene il tipo *TouristAttraction*, come sottoclasse di *Place*, definito in maniera tautologica come “a tourist attraction” e non avente alcuna sottoclasse.

È ovvio che, considerata la complessità e l’eterogeneità che caratterizzano il mondo del turismo (come mostrato nel capitolo precedente) e la tendenza anche da parte delle principali organizzazioni che si occupano di turismo come la *UNWTO* di fornire delle classificazioni e definizioni più precise e applicabili a livello internazionale dei diversi aspetti che costituiscono l’esperienza turistica, una definizione di questo tipo appare molto, troppo generica.

Per questo motivo, l’obiettivo della prima parte del lavoro di tesi è stato fornire una descrizione del concetto di attrazione turistica più precisa rispetto a quella data da *Schema* e di proporla in modo da contribuire ad allargare i contenuti semantici standardizzati dalla comunità.

3.3.1 Una nuova definizione e nuove proprietà per *TouristAttraction*.

Al fine di proporre una nuova definizione del concetto di *TouristAttraction* su *Schema.org* è stata stabilita una collaborazione con il gruppo *W3C Tourism Structured Web Data Community Group*¹⁴, una community nata per l’estensione di *Schema.org* con dati strutturati relativi a contenuti legati al mondo del turismo.

In base alla definizione di attrazione (2.3) e alle linee guida sul turismo fornite dall’*UNWTO* (cap. 2), qualsiasi entità di tipo fisico, culturale, artificiale, naturale, che sia in grado di determinare lo spostamento di uno o più viaggiatori verso un luogo diverso da quello in cui abitualmente vivono, è un’attrazione turistica.

Per questo motivo, al tipo *TouristAttraction* è stata fornita una *nuova e più precisa descrizione*:

A tourist attraction. In principle *any Thing can be a TouristAttraction*, from a *Mountain and LandmarksOrHistoricalBuildings* to a *LocalBusiness*. This Type can be used on its own to describe a general *TouristAttraction*, or be used as an *additionalType* to add tourist attraction properties to any other type.

¹⁴ <https://www.w3.org/community/tourismdata/>

Al tipo, inoltre, sono state date *due nuove proprietà* ovvero *availableLanguage*, che descrive la “lingua che qualcuno può usare con o presso un’entità, un servizio o un posto”, e *touristType*, che descrive il tipo principale di turista attratto, ad esempio bambini o turisti provenienti da un determinato paese.

Al tipo *Place* sono state aggiunte la già esistente proprietà *isAccessibleForFree* (che indica se un luogo è accessibile gratuitamente) e la *nuova proprietà publicAccess* che descrive sotto forma di booleano se un posto è accessibile dai visitatori. Essendo una sottoclasse di *Place*, queste proprietà possono essere usate anche con *TouristAttraction*.

Come si può notare, la realizzazione della nuova definizione di *TouristAttraction* non è consistita nell’aggiunta al tipo di nuove sottoclassi rappresentanti diversi tipi di entità che possono essere attrazioni turistiche, poiché esse non costituirebbero che un numero limitato ed esiguo di entità non in grado di abbracciare l’intero insieme di elementi che potrebbero essere un’attrazione. Piuttosto, per indicare che un determinato tipo (ad esempio un elemento del tipo *Museum*) è anche un’attrazione turistica, è possibile usare i cosiddetti *multiple types* o *multi-types (MTEs)*. Essi, infatti, permettono di associare a un’entità due o più tipi, permettendo, dunque, di definire uno stesso contenuto come appartenente a due o più tipi. Utilizzandoli, quindi, è possibile, ad esempio, definire la *Shell Beach* in Australia non solo come un’entità di tipo *Beach*, ma anche come entità di tipo *TouristAttraction*. In questo modo, in teoria, qualsiasi entità descrivibile tramite il vocabolario di *Schema* potrebbe essere definita anche come attrazione turistica, appartenente al tipo *TouristAttraction*.

Dal punto di vista implementativo, inoltre, l’uso dei multi-types non richiede alcuna particolare o complessa sintassi, infatti, per quanto riguarda, ad esempio, i *Microdata*, basta aggiungere l’*URL* che rimanda al tipo *TouristAttraction* come valore dell’attributo *itemtype*. Anche per quanto riguarda la sintassi *JSON-LD* e *RDFa* basta aggiungere il termine *TouristAttraction* come valore rispettivamente di *@type* e *typeof* (v. 3.2).

Nella stessa nuova descrizione di *TouristAttraction*, infatti, è ribadito che a qualsiasi altro tipo del vocabolario può essere aggiunto il tipo *TouristAttraction* (ad esempio,

Mountain o *LocalBusiness*), usando i multi-types. La proprietà *additionalType* a cui si fa riferimento è una proprietà usata, soprattutto in *Microdata* e *JSON-LD* (v. 3.2), ma sconsigliata per *RDFa* (v. 3.2), per aggiungere un tipo a un'entità.

Nel contesto dell'attività del gruppo *W3C* sul turismo, mi sono occupata personalmente di realizzare alcuni esempi che mostrano l'uso delle nuove proprietà e degli *MTEs* per descrivere un'attrazione turistica.

Il seguente esempio in *Microdata* mostra un possibile uso della proprietà *touristType* e *additionalType*:

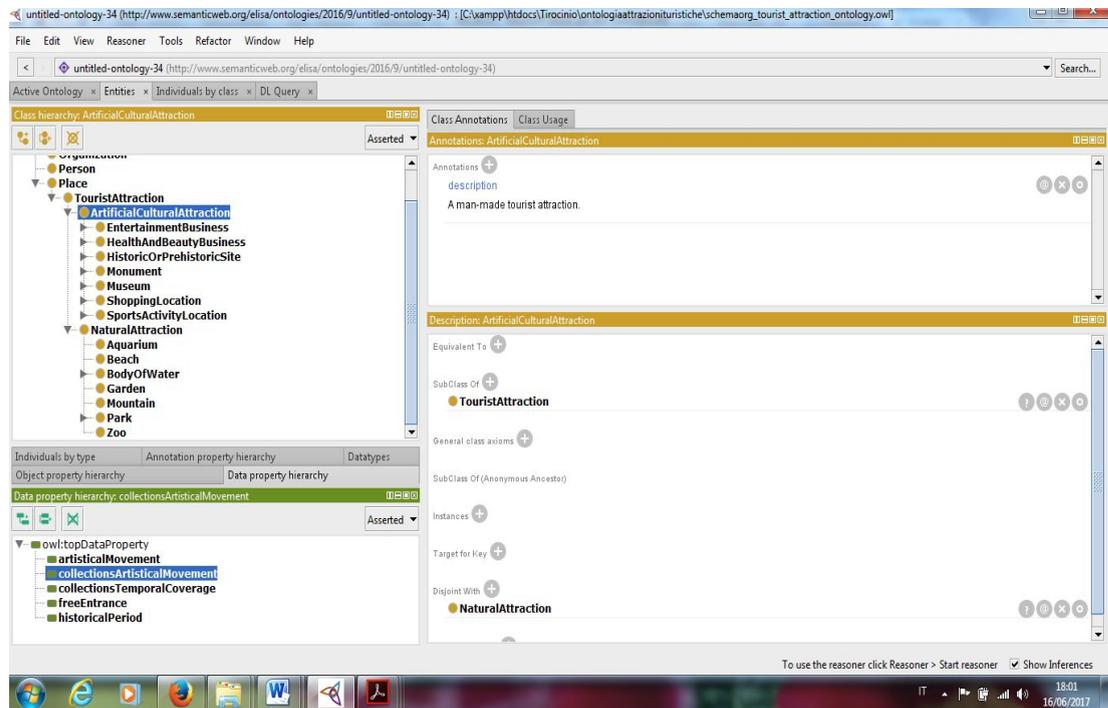
```
<div itemscope itemtype="http://schema.org/TouristAttraction">
  <h1><span itemprop="name">Please Touch Museum</span></h1>
  <div>It is a
    <div itemprop="touristType" itemscope
itemtype="http://schema.org/Audience">
      <span itemprop="audienceType">children</span>
    </div>'s museum located in
    <div itemprop="address" itemscope
itemtype="http://schema.org/PostalAddress">
      <div itemprop="addressLocality">Philadelphia</div>,
      <div itemprop="addressCountry">USA</div>.
    </div>
  <div>
    <span itemprop="description">The museum focuses on
      teaching children through interactive exhibits and special
      events.</span>
  </div>
</div>
```

Nell'esempio che segue, invece, in *RDFa*, viene mostrato l'uso della proprietà *availableLanguage* per definire, in questo caso, le lingue disponibili per i tour guidati:

```
<div vocab="http://schema.org" typeof="TouristAttraction">
  <h1><span property="name">Neuschwanstein Castle</span></h1>
  <div>It is a nineteenth-century Romanesque Revival palace in
    <div property="address" typeof="PostalAddress">
      <span
property="addressLocality">Schwangau</span>, in <span
property="addressCountry">Germany</span>.
    </div>
  </div>
  <div>Guided tours in
    <div property="availableLanguage" typeof="Language">
      <span property="name">German</span>
    </div> and
    <div property="availableLanguage" typeof="Language">
      <span property="name">English</span>.
  </div>
```

```
</div>
</div>
</div>
```

Dal punto di vista implementativo, per la costruzione della nuova definizione si è scelto di usare *Protege*, un software open source che permette di definire ontologie usando il linguaggio *OWL* (par. 3.2) . In figura 2 è mostrata una schermata-tipo del software.



(Figura 2. Schemata-tipo di *Protege*)

Scaricando il progetto *Schema.org* dalla piattaforma *Github*¹⁵ è stato possibile modificarlo aggiungendo la nuova proposta sotto forma di file *RDFa* (par. 3.2) e gli esempi costruiti. Grazie all'uso di *Google App Engine*¹⁶, piattaforma cloud gestita da *Google* per lo sviluppo e l'hosting di applicazioni web, è stato inoltre, possibile ottenere un *URL* che replicasse il vocabolario di *Schema* arricchito con l'estensione proposta.

La nostra proposta ha convinto la comunità di *Schema.org* e sarà inserita nella versione 3.3 di *Schema*. La nuova versione di *TouristAttraction* sarà, dunque, ufficiale e utenti di tutto il mondo potranno usarla per marcare semanticamente contenuti web relativi al settore turistico.

¹⁵ <https://github.com/schemaorg/schemaorg>

¹⁶ <https://cloud.google.com/appengine/>

Al momento la nuova versione 3.3 di *Schema.org*, che contiene la nuova definizione di *TouristAttraction*, è visibile su *webschemas.org*, dove vengono pubblicate le diverse versioni non ancora ufficiali ma in attesa di rilascio.

La figura 3 mostra come apparirà il tipo *TouristAttraction* dopo il rilascio ufficiale.

TouristAttraction
 Canonical URL: <http://schema.org/TouristAttraction>

Thing > Place > **TouristAttraction**

A tourist attraction. In principle any Thing can be a **TouristAttraction**, from a **Mountain** and **LandmarksOrHistoricalBuildings** to a **LocalBusiness**. This Type can be used on its own to describe a general **TouristAttraction**, or be used as an **additionalType** to add tourist attraction properties to any other type. (See examples below)

Usage: Between 1000 and 10,000 domains [more...]

| Property | Expected Type | Description |
|---|------------------|--|
| Properties from TouristAttraction | | |
| availableLanguage | Language or Text | A language someone may use with or at the item, service or place. Please use one of the language codes from the IETF BCP 47 standard . See also inLanguage |
| touristType | Audience or Text | Attraction suitable for type(s) of tourist. eg. Children, visitors from a particular country, etc. |
| Properties from Place | | |
| additionalProperty | PropertyValue | A property-value pair representing an additional characteristics of the entity, e.g. a product feature or another characteristic for which there is no matching property in schema.org. Note: Publishers should be aware that applications designed to use specific schema.org properties (e.g. http://schema.org/width , http://schema.org/color , http://schema.org/gtin13 , ...) will typically expect such data to be provided using those properties, rather than using the generic property/value mechanism. |

(Figura 3. Il tipo *TouristAttraction* nella versione 3.3 di Schema)

3.3.2 Sviluppi futuri.

L'approdo al rilascio ufficiale della nuova definizione di *TouristAttraction* su *Schema.org* ha costituito un ottimo risultato non solo per me, in quanto studentessa, ma anche per il *Web Applications for the Future Internet Lab* del *CNR* di Pisa.

Senza alcun dubbio si tratta di un buon lavoro, in quanto accettato per essere rilasciato ufficialmente, ma è anche vero che può sempre essere migliorato concentrandosi, ad esempio, su due aspetti.

Il primo riguarda la descrizione stessa di *TouristAttraction*. Essa si incentra giustamente, sul fatto che qualsiasi entità può, in linea di principio, essere un'attrazione turistica. È vero, anche un negozietto di antiquariato alla fine di una piccola viuzza di un quartiere periferico può diventare un'attrazione turistica; ma è anche vero che è necessario indicare almeno un altro criterio per il quale un'entità, di qualsiasi natura essa sia, può essere considerata un'attrazione turistica. Si tratta del fatto che essa, come indicato anche nelle linee guida del *UNWTO* (v. 2.3) deve causare lo spostamento di un viaggiatore dall'ambiente in cui vive abitualmente: se un individuo va al supermercato sotto casa è ovvio che quest'ultimo non costituisce

in alcun modo un'attrazione turistica.

Per questo motivo, sarebbe auspicabile proporre in futuro, l'aggiunta di questo aspetto nella descrizione di cosa si intende per *TouristAttraction*, in modo da cercare di limitare un insieme di possibili attrazioni che è già di per sé molto ampio.

Il secondo aspetto da considerare per un miglioramento della descrizione semantica di attrazione turistica e del mondo del turismo in generale è la possibilità di intraprendere un lavoro di aggiunta a *Schema.org* di nuovi tipi che rappresentano elementi che possono essere attrazioni turistiche come *Castle*, *ArchaeologicalSite*, *AncientRuin*. Si tratta, infatti, di tipi che con più frequenza possono considerarsi attrazioni turistiche e la cui aggiunta nel vocabolario favorirebbe una migliore e meno generica marcatura semantica dei contenuti web riguardanti il turismo.

CAPITOLO 4

GLI OPEN DATA RELATIVI ALLE ATTRAZIONI TURISTICHE ITALIANE

La *Open Definition* redatta dalla *Open Knowledge Foundation* stabilisce che si definiscono *aperti dati e contenuti* a cui chiunque può accedere liberamente, che possono essere *usati, modificati e ridistribuiti per qualsiasi scopo* e le cui eventuali restrizioni possono essere esclusivamente l'obbligo di citare la fonte e di mantenerli aperti¹⁷.

In particolare, sempre secondo la *Open Definition*, un dato si può definire aperto se rispetta quattro criteri fondamentali¹⁸, ovvero:

- la sua distribuzione non è sottoposta a Copyright o simili restrizioni, oppure è accompagnata da una *licenza aperta*, ovvero una licenza che permetta liberi uso, ridistribuzione e modifica del dato per qualsiasi scopo;
- è disponibile nella *sua interezza* preferibilmente tramite il download gratuito via *Internet*, accompagnandolo anche con informazioni aggiuntive riguardanti la licenza;
- è disponibile in forma *machine-readable*, in modo che sia leggibile e modificabile da un computer;
- è distribuito in un *formato aperto*, cioè un formato che non sia soggetto a restrizioni, anche di tipo monetario e che può essere facilmente processato da un programma open source o libero (ad esempio il formato *csv*).

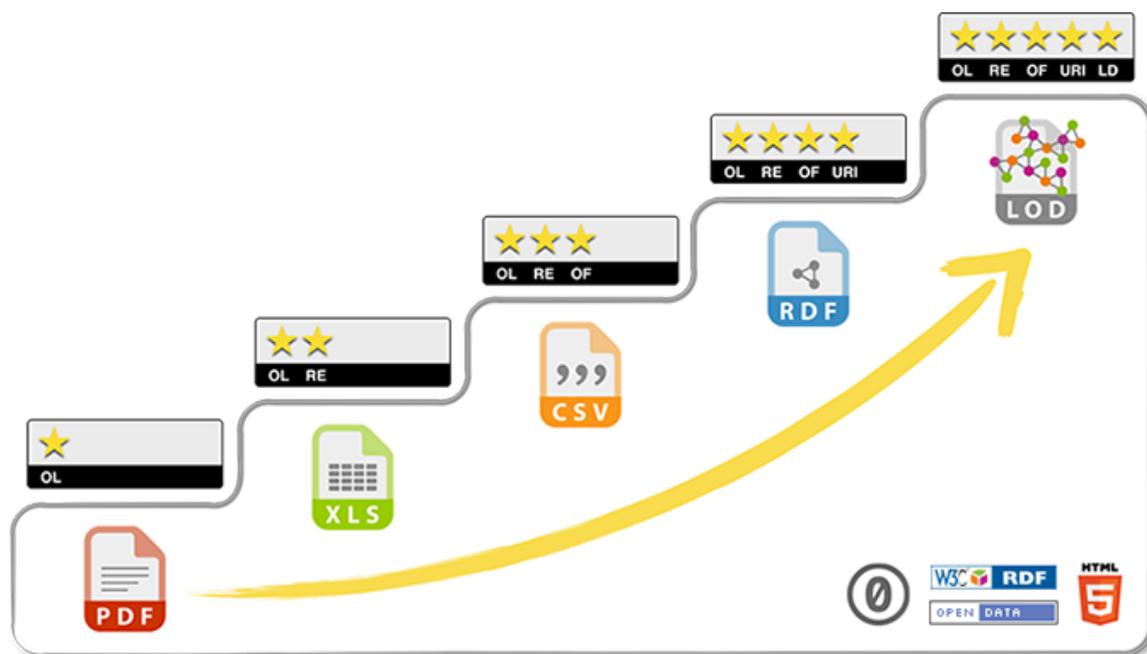
I software non sono inclusi nella definizione di Open Data, poiché di essi si occupa la *Open Source Definition* dalla quale, tra l'altro, la *Open Definition* deriva.

A fornire un'ulteriore definizione di Open Data, molto vicina alla *Open Definition* è stato Tim Berners-Lee, inventore del *World Wide Web* e fondatore del *World Wide Web Consortium (W3C)*. Nella sua definizione egli ordina i dati secondo una gerarchia che va da uno a cinque stelle, in cui una stella corrisponde al minimo grado

¹⁷ <http://opendefinition.org/>

¹⁸ <http://opendefinition.org/od/2.1/en/>

di apertura del dato e cinque rappresentano, invece, il massimo grado di apertura¹⁹, come mostra la figura 4. Al livello più basso vi sono dati che sono disponibili sul web sotto una licenza aperta (OL= Open License), ma che non sono strutturati e sono codificati in un formato proprietario come *PDF*. Al secondo livello si hanno i dati che sono strutturati (RE= machine REadable), ma sono codificati in un formato proprietario, ad esempio, i dati in formato *Excel*. Ad aggiudicarsi tre stelle sono, invece, i dati strutturati e codificati in un formato non proprietario o aperto (OF= Open Format), ad esempio i dati in formato *CSV* ed è proprio a partire da questo livello che i dati possono effettivamente considerarsi aperti. Al quarto livello si trovano, invece, i dati che non solo sono strutturati e codificati in un formato non proprietario, ma in cui ogni dato è identificato da un *URI (Uniform Resource Identifier)* che gli dà un significato condiviso universalmente e questo è il caso dei dati in formato *RDF*. Sul quinto e ultimo gradino vi sono, invece, i *Linked Open Data (LD= Linked Data)*, ovvero insiemi di dati aperti collegati tra loro.



(Figura 4. La gerarchia di Berners-Lee)

È evidente il fatto che la creazione e la distribuzione di Open Data rappresentano una risorsa importantissima non solo per una maggiore diffusione della conoscenza di dati che altrimenti rimarrebbero accessibili solo a determinate tipologie di utenti, ma

¹⁹ <http://5stardata.info/en/>

anche per lo sviluppo dell'ambito dell'informazione, ad esempio attraverso il *data journalism* o *giornalismo di precisione*, ovvero l'insieme di inchieste giornalistiche basate sull'uso della matematica, della statistica e delle scienze sociali che traggono profondo beneficio dalla possibilità di attingere liberamente a banche dati di tipo open. Non bisogna, infine, dimenticare che la possibilità di ottenere e poter utilizzare liberamente una grande quantità di dati sarebbe di grande importanza per lo sviluppo di una grande varietà di applicazioni web che manipolino tali dati.

4.1 Gli Open Data in Italia.

Per quanto riguarda l'Italia, con la *Legge 17 dicembre 2012, n. 221*, è stato modificato il comma 3 dell'articolo 68 del *Codice dell'amministrazione digitale (Decreto Legislativo 7 marzo 2005, n. 82)* fornendo una definizione formale di dati aperti che rispecchia quella data dall'*Open Knowledge Foundation*: si tratta di dati accessibili liberamente, machine-readable, distribuiti in formati aperti e sottoposti a una licenza che ne consenta l'uso da parte di chiunque e per qualunque scopo, anche di tipo commerciale²⁰.

In linea con questa definizione, già nel 2011 nasce il portale *dati.gov.it* che raccoglie, sotto varie categorie, i dati open della pubblica amministrazione ed è gestito dall'*Agenzia per l'Italia digitale* e da *Formez PA* che si occupano della digitalizzazione del sistema informativo della pubblica amministrazione.

Il portale raccoglie gli open data delle pubbliche amministrazioni riguardanti argomenti come *Energia, Scienza e tecnologia, Salute, Ambiente, Istruzione, cultura e sport* e la *Formez PA* ha rilasciato l'*Italian Open Data License v2.0*, un contratto di licenza aperto per l'uso e la modifica degli open data delle pubbliche amministrazioni²¹.

Quasi tutte le regioni italiane, inoltre, hanno realizzato portali web che raccolgono dati aperti ad esse relativi. All'appello mancano ancora la Sicilia e il Molise che però, sembra si stiano lentamente muovendo in questa direzione. Per quanto riguarda la Sicilia, in particolare, a giugno 2017 si è svolto a Caltanissetta l'*Opendatafest*, un'iniziativa civica per la promozione degli Open Data, ed è online il *NON-Portale*

²⁰ http://www.agid.gov.it/sites/default/files/leggi_decreti_direttive/dl-18-ottobre-2012-n.179_0.pdf

²¹ <http://www.dati.gov.it/content/italian-open-data-license-v20>

Open Data della Regione Siciliana, che raccoglie dati aperti riguardanti la Sicilia ma non è il portale ufficiale della regione.

4.2 Selezione degli Open Data sulle attrazioni turistiche italiane.

La ricerca dei dati aperti riguardanti le attrazioni turistiche italiane è stata fatta a partire dai principali portali italiani per la diffusione di questo tipo di dati, ovvero *dati.gov.it*, *datiopen.it* e i portali ufficiali delle singole regioni. Delle venti regioni italiane, solo Sicilia e Molise non hanno ancora un portale dedicato agli Open Data.

Provando a cercare le attrazioni turistiche sui portali di dati open ci si rende subito conto della complessità di dati a cui è possibile accedere. Si tratta, infatti, di un insieme altamente eterogeneo di dati codificati secondo diversi formati e privi di uno standard di categorizzazione adeguato che permetta di raccogliarli e selezionarli facilmente.

In primo luogo, infatti, ciò che si nota è il fatto che non tutti i portali hanno una sezione dedicata esplicitamente al turismo e solo questo fattore basta a complicare la ricerca. Ad aggiungersi a questo, vi è il fatto che sono davvero pochi i dataset che contengono nel loro nome i termini “attrazioni”, “attrazioni turistiche”, “turismo”. Questo complica di molto la ricerca di possibili attrazioni turistiche poiché, considerando proprio il concetto di attrazione turistica così come definito in 2.3, qualsiasi elemento che sia in grado di determinare un flusso turistico è un’attrazione turistica. Il risultato è quindi che, laddove non sia presente una specifica sezione *Turismo* o nel nome del dataset non si trovino termini come “attrazioni turistiche”, è necessario armarsi di buona volontà e cercare tra le altre categorie di dati quelli che hanno una qualche probabilità di essere un’attrazione turistica. Tale probabilità, inoltre, non è tanto da intendersi in senso statistico, ma più come derivante dall’esperienza e dalla percezione della realtà da parte di chi estrae i dati. Essa non è effettivamente fondata su un insieme di dati significativamente rappresentativo. I dati su cui potrebbe fondarsi, infatti, sono gli stessi che in formato aperto non sono adeguatamente catalogati e possono essere difficili da recuperare da fonti non aperte. Escludendo, dunque, dataset come *Produzione editoriale della regione Toscana nel*

2015, oppure *Superfici agricole e seminativi*, o ancora *Bilanci – Bilancio preventivo e consuntivo 2015*, i cui dati hanno una potenzialità pari a zero di essere effettivamente attrazioni turistiche, rimane comunque un grandissimo e molto eterogeneo insieme di dati che potenzialmente possono essere attrazioni turistiche. Tra questi ci sono elementi che con maggiore probabilità possono essere attrazioni turistiche, come i musei, i siti archeologici, gli stabilimenti balneari, i teatri, i beni culturali. Ce ne sono poi, molti altri che pur avendo una minore probabilità di diventare attrazioni, possono, in certi casi, esserlo ugualmente, come le fontane (basta pensare alla Fontana di Trevi), i cinema, i centri commerciali, i giardini (il Giardino di Boboli a Firenze, ad esempio), i locali e le botteghe storiche, gli stadi (nel caso, ad esempio, di concerti o di eventi sportivi), le biblioteche. Ci sono, infine, elementi che hanno una probabilità ancora più bassa di quelli precedenti di essere attrazioni turistiche, come le farmacie, le strutture sanitarie, le università, le medie strutture di vendita. Anche se a una prima occhiata potrebbe sembrare strano, esiste la possibilità, anche se minima rispetto ad altri tipi di attrazione (come i musei, ad esempio), che elementi di questo tipo diventino un'attrazione turistica. Secondo la definizione di turismo e attrazione turistica (v. cap. 2), infatti, non si parla di turismo solo quando lo spostamento avviene per divertimento o svago, ma il motivo del movimento può essere anche di altro tipo: esiste anche il turismo sanitario, per esempio.

Nella prima fase della ricerca sono stati raccolti i dataset i cui elementi appartengono a uno dei tre insiemi descritti sopra, ovvero quelli che hanno un qualche probabilità, minima, media o alta, di essere un'attrazione turistica.

Solamente 13 regioni forniscono risorse significative per l'ambito turistico. A fornire il numero più elevato di dataset sono Piemonte, Umbria, Lombardia, Toscana, Sardegna, Emilia-Romagna e Lazio. Alcuni dataset, tuttavia, sono relativi non a tutto il territorio regionale ma a una specifica città o provincia che ne fa parte. È il caso della Basilicata, per la quale sono disponibili solo dati relativi alla città di Matera, dell'Emilia-Romagna, per la quale molti dei dati sono disponibili solo per la città di Bologna, della Lombardia, per la quale molti sono i dati disponibili solo per la provincia di Monza-Brianza, del Trentino, per il quale i dati riguardano solo la provincia di Trento ed escludono quella di Bolzano, e della Toscana, per la quale

alcuni dati riguardano solo Firenze e i comuni limitrofi. Per quanto riguarda i dati della regione Liguria, inoltre, tutti tranne uno sono accessibili tramite link che restano attivi per soli cinque giorni dopo i quali è necessario inoltrare nuovamente una richiesta per il download dei dati.

Di fronte a un insieme così vasto di elementi e all'assenza di una categorizzazione preventiva da parte del distributore del dato che suddivida le attrazioni da altri tipi di elementi, è stata operata un'ulteriore selezione che includesse alcuni tipi di elementi e ne escludesse altri. In questo caso, il criterio scelto per la selezione è proprio quello della probabilità con la quale un'entità può essere un'attrazione turistica.

Dall'insieme di dati raccolti nella prima parte della ricerca, infatti, sono stati esclusi quelli con la probabilità più bassa di essere attrazioni turistiche, ovvero elementi come farmacie, strutture sanitarie, medie strutture di vendita.

È bene notare che questo tipo di selezione si è basato soprattutto sul nome del dataset e, eventualmente, sulle categorie di elementi da esso contenuti. È sufficiente solo che all'interno del dataset vi sia una categoria di elementi che abbia probabilità media o alta di essere un'attrazione turistica (come una fontana o un museo) per far rientrare tutti i dati ad esso appartenenti nell'insieme delle attrazioni turistiche. Proprio per quest'ultimo motivo, tuttavia, non è escluso che all'interno della raccolta siano presenti anche elementi che, in realtà, hanno una bassissima probabilità di essere attrazioni e che, diversamente, sarebbero stati esclusi.

La tabella 1A in Appendice mostra i dataset raccolti per ogni regione indicando anche se sono stati o meno selezionati come insiemi di attrazioni turistiche.

È evidente che una catalogazione di questo tipo non solo è dispendiosa in termini di ricerca e selezione dei dati, ma rischia di allontanarsi dalla realtà, includendo anche elementi che, in realtà, non appartengono al mondo del turismo o, addirittura, non sono attrazioni turistiche.

4.3 Qualità degli Open Data sulle attrazioni turistiche italiane.

Come scrive Rufus Pollock, fondatore e presidente di *Open Knowledge*, sull'*Open Knowledge International Blog*, la *Open Definition* mira alla creazione di uno

standard per i dati aperti che legghi indissolubilmente il concetto di Open Data con tre elementi chiave, ovvero *qualità, compatibilità e semplicità*²².

Con qualità si intende il fatto che non basta che un insieme di dati sia semplicemente disponibile per essere considerato open. I dati open, in quanto tali, come stabilisce la definizione stessa, devono essere anche accessibili liberamente, utilizzabili e modificabili da chiunque e per qualsiasi scopo. Se non si ha chiaro questo concetto, è facile confondere la natura dei dati e attribuire l'attributo "open" anche a tipi di dati che effettivamente non lo sono.

La compatibilità consiste nel fatto che i dati aperti dovrebbero essere distribuiti secondo formati e con licenze d'uso che siano compatibili tra loro, in modo da permettere di combinarli tra di loro.

Con semplicità non si intende solo il fatto che l'accesso ai dati sia garantito senza costi, ma anche e soprattutto il fatto che sia reso esplicito, attraverso le licenze, cosa è possibile fare con i dati, eliminando qualsiasi eventuale timore riguardante il loro uso.

A queste tre parole chiave si può aggiungere la scala sul grado di "apertura" dei dati fornita da Tim Berners-Lee (fig. 4) e considerare come migliori tipi di dati aperti quelli che si possono collocare almeno sul terzo gradino della scala, ovvero i dati strutturati che sono distribuiti secondo una licenza aperta in formati non proprietari.

Giudicare la qualità di un insieme di dati aperti, quindi, non significa solo valutare la quantità dei dati, ad esempio in relazione a un particolare argomento, ma vuol dire anche controllare che siano effettivamente utilizzabili e accessibili liberamente e soprattutto, verificare su quale gradino della gerarchia di Lee essi si collocano, quindi, individuare se si tratta o meno di dati strutturati e in che formati essi sono disponibili.

Per quanto riguarda gli Open Data relativi alle attrazioni turistiche italiane, quindi, la loro qualità può essere valutata in relazione a fattori come la presenza o meno di uno standard nella loro categorizzazione, il formato in cui i dati sono codificati, e la presenza di eventuali problemi redazionali come errori ortografici, errori nella struttura informativa. A questo tipo di analisi sono destinati rispettivamente i

²²

<https://blog.okfn.org/2014/09/30/why-the-open-definition-matters-for-open-data-quality-compatibility-and-simplicity/>

paragrafi 4.3.1, 4.3.2 e 4.3.3.

4.3.1 L'assenza di uno standard.

A rendere complesso l'insieme dei dati raccolti come attrazioni turistiche italiane non è solo l'assenza di uno standard adeguato che permetta di distinguerle nettamente da altri tipi di elementi (v. 4.2), ma è anche un'altra importante caratteristica che li riguarda, ovvero l'assenza di uno standard per la categorizzazione interna dei diversi tipi di dataset che li renda compatibili gli uni con gli altri in modo da facilitarne l'integrazione e l'uso integrato. Ciò significa che dataset che raccolgono lo stesso tipo di dato presentano al loro interno una suddivisione in categorie/tipologie diversa l'uno dall'altro (ad esempio, per quanto riguarda i musei la regione Liguria presenta la categoria *Specializzato-Scienza e tecnica*, che unisce due tipologie trattate separatamente dalle altre regioni). Si verifica anche che alcuni tipi di dati considerati macro-categorie in certi dataset, in altri siano invece indicati come sotto-categorie di un'altra categoria più grande (ad esempio, gli ecomusei costituiscono una sottocategoria dei musei per alcune regioni, mentre per altre come l'Umbria, sono una categoria a parte). Altro problema legato alla mancanza di uno standard è il fatto che alcuni dataset presentano categorie troppo generiche o estremamente eterogenee, mentre altri risultano troppo specifici (ad esempio, la Puglia presenta il dataset *Luoghi di interesse turistico, culturale, naturalistico*, mentre il Lazio considera le multisale non come una sottocategoria dei cinema ma come una categoria a parte).

4.3.2 Il formato dei dati.

La qualità dei dati open può essere misurata anche in relazione al formato in cui sono codificati e distribuiti. Considerando la gerarchia elaborata da Tim Berners-Lee (fig. 4), infatti, il grado di apertura di un dato è più alto se questo, oltre a essere strutturato e sottoposto a una licenza aperta, è codificato in un formato aperto, ovvero che non necessita di software proprietari per essere aperto e usato. Anzi, basandosi sulla definizione stessa di Open Data che li descrive come dati accessibili liberamente da chiunque, si potrebbero considerare come effettivamente open solo i dati che sono codificati proprio in formati non proprietari.

Gli Open Data relativi alle attrazioni turistiche italiane presentano in linea di massima quest'ultimo grado di apertura, ovvero sono per la stragrande maggioranza

codificati attraverso formati aperti.

Tra i formati più usati vi è il *CSV (Comma-separated values)*, un formato aperto che si basa su un file di testo in cui i vari campi, ognuno rappresentante di un tipo diverso di informazione, sono separati da una virgola o da altri delimitatori come il punto e virgola o lo spazio.

Numerosissimi sono anche i dataset codificati in *JSON (JavaScript Object Notation)*, un formato basato sul linguaggio *Javascript* e molto usato nell'interscambio di dati fra applicazioni client-server.

A seguire, si hanno anche dati codificati in *XML (eXtensible Markup Language)*, un metalinguaggio usato per la definizione di linguaggi di markup per la marcatura di contenuti e informazioni anche di tipo semantico.

Altro formato aperto usato, seppur in maniera minore rispetto ai precedenti, è *KML (Keyhole Markup Language)*, basato su *XML* e usato soprattutto per la codifica e la mappatura di dati geografici.

Molto pochi rispetto alla maggioranza sono anche i dati codificati in *RDF (Resource Description Framework) (v. 3.2)*, uno standard per la rappresentazione dei contenuti semantici che associa ad ogni entità codificata un *URI (Uniform Resource Identifier)* che lo identifica e che permette, inoltre, di aumentare il grado di apertura dei dati collocandoli sul quarto gradino della gerarchia di Lee (fig. 4).

Usato è anche il formato aperto *GeoJSON*, un'estensione di *JSON* per la codifica di dati geografici sotto forma di geometrie spaziali.

Un altro formato aperto usato, anche se in maniera esigua rispetto agli altri, per la codifica dei dati relativi alle attrazioni turistiche italiane è *Shapefile*, un formato di dati vettoriale per la codifica di informazioni geografiche usato soprattutto per la creazione di *Geographic Information Systems (GIS)*.

Davvero pochi rispetto ai dati codificati con formati aperti, sono i dati distribuiti in formati proprietari, ovvero di cui non sono liberamente fruibili le specifiche tecniche e che in genere necessitano di un programma proprietario, come *Microsoft Excel*, per essere usati. Tra questi, appunto, la stragrande maggioranza sono dati in formato *Excel (XLS)*, ma si possono trovare anche dati in *PDF (Portable Document Format)*, formato sviluppato da *Adobe Systems*.

4.3.3 Problemi redazionali.

Il fatto che un dato strutturato sia reso accessibile liberamente da chiunque accompagnato da una licenza aperta e in un formato non proprietario non sempre significa che sia semplice manipolarlo. A porre ostacoli all'analisi e all'integrazione dei dati aperti, infatti, sono fattori come l'assenza di uno standard nella loro nella loro categorizzazione interna che ne aumentano la complessità diminuendone la qualità, come si è visto in 4.3.1.

A complicare la manipolazione e l'analisi dei dati open sono anche eventuali problemi di tipo redazionale, cioè errori e imprecisioni che riguardano il processo di stesura e di codifica dei dataset da parte del loro distributore.

I problemi che si incontrano nei dataset riguardanti le attrazioni turistiche italiane sono di diverso tipo e, in particolare, mentre alcuni riguardano i linguaggi e i formati usati per la codifica dei dati, altri sono attinenti al contenuto informativo stesso.

Riguardo al primo tipo di errori, si possono trovare dataset in cui non è rispettata la sintassi del linguaggio utilizzato per la codifica e ci si può trovare, ad esempio, ad avere a che fare con file di modesta dimensione codificati in un *JSON* non ben formato che, quindi, è molto difficile se non impossibile da processare per manipolarne i dati.

Altro errore di questo tipo è, per quanto riguarda i file in formato *CSV*, la presenza di spazi o altri segni grafici che, alterando la struttura creata dai delimitatori effettivi, rendono i dati difficilmente processabili.

Anche il secondo tipo di errori, riguardante il contenuto informativo che viene codificato, complica la manipolazione dei dati stessi. A questo insieme appartengono un numero significativo di errori tra i più svariati.

Alcuni errori sono veri e propri errori ortografici che riguardano i nomi di comuni, città, province, gli indirizzi. Sarebbe possibile, ad esempio, trovarsi ad avere a che fare con la città di *Lece* al posto di *Lecce*, con la provincia di *Treni* invece che di *Terni*, con *Via Antonio Gramisci* invece di *Gramsci*. Errori di questo tipo, che potrebbero ad una prima analisi sembrare poco importanti, assumono un certo peso se si immagina non solo che essi sono più numerosi di quanto si possa pensare, ma anche che dati di questo tipo, come l'indirizzo e la città, potrebbero essere inseriti, ad

esempio, in sistemi di calcolo delle coordinate ottenendo, con molta probabilità, risultati meno precisi del previsto.

Altri errori riguardano, invece, la strutturazione stessa dell'informazione all'interno del dataset. Può avvenire che nel campo destinato a un certo tipo di informazione, ad esempio numero di telefono, venga inserito un altro tipo di informazione, come l'indirizzo e-mail. Anche questi errori non sono da sottovalutare, poiché creano un effettivo squilibrio sul piano semantico tra i dati.

Altri errori riscontrabili poi, riguardano la mancata revisione dei dati prima di essere distribuiti. È possibile, ad esempio, trovare in un dataset, annotazioni e commenti fatti dal redattore durante la stesura e poi non eliminati dal file definitivo. Si tratta di informazioni che non dovrebbero riguardare il contenuto informativo da rendere effettivamente accessibile e che non dovrebbero interessare l'utente che scarica i dati.

È evidente, dunque, il fatto che la presenza di errori redazionali tra i dati aperti delle attrazioni turistiche italiane, soprattutto se si considera l'eterogeneità di tali errori, contribuisce a far diminuire la qualità dei dati stessi complicando la loro manipolazione da parte del fruitore.

4.4 Raccolta automatica dei dati e integrazione nella piattaforma Tourpedia.

I dati relativi alle attrazioni turistiche sono stati raccolti all'interno della piattaforma *Tourpedia*. *Tourpedia* è un portale web sviluppato dall'*Istituto di Informatica e Telematica* del *CNR* di Pisa. Esso si occupa di raccogliere e manipolare dati open relativi al turismo. *Tourpedia* definisce una procedura automatica per la raccolta dei dati da diverse sorgenti.

Il presente lavoro di tesi ha arricchito *Tourpedia* con i dati relativi alle attrazioni turistiche estratte dai portali delle regioni e province italiane. In particolare, il contributo specifico di questo lavoro di tesi all'interno di *Tourpedia* consiste nei seguenti elementi:

- aggiunta a *Tourpedia* delle attrazioni turistiche relative alle regioni e province italiane;

- definizione del *Tourpedia Data Model* (descritto in seguito) relativo alle attrazioni turistiche;
- estensione del *Tourpedia Mapping Language* (descritto in seguito) in modo da supportare i formati *XSL*, *XML* e *KML*.

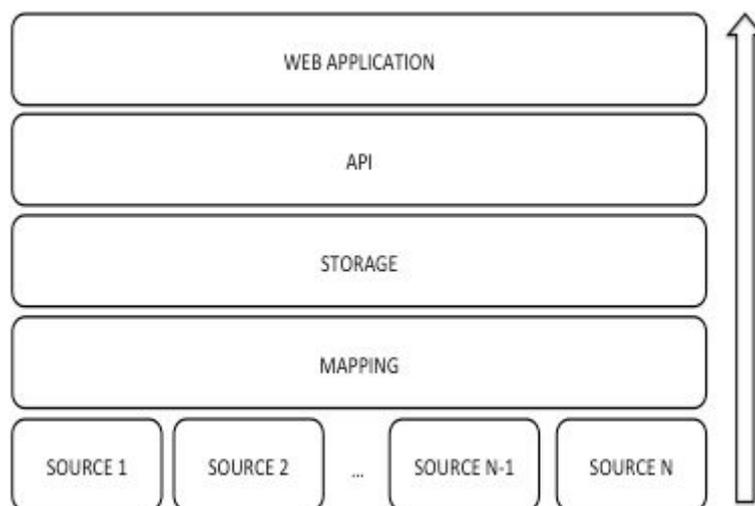
La tabella seguente mostra il numero di record raccolti per ogni regione italiana.

| Regione | Numero di record |
|-----------------------|-------------------------|
| Abruzzo | 0 |
| Basilicata | 169 |
| Calabria | 0 |
| Campania | 0 |
| Emilia-Romagna | 1531 |
| Friuli-Venezia Giulia | 0 |
| Lazio | 1990 |
| Liguria | 352 |
| Lombardia | 5537 |
| Marche | 7178 |
| Molise | 0 |
| Piemonte | 1471 |
| Puglia | 788 |
| Sardegna | 13289 |
| Sicilia | 0 |
| Toscana | 4668 |
| Trentino-Alto Adige | 1780 |
| Umbria | 1009 |
| Valle D'Aosta | 0 |
| Veneto | 837 |

(Tabella 4. Numero di record raccolti per ogni regione italiana)

La figura 5 mostra l'architettura di *Tourpedia*. Partendo dal basso verso l'alto, si

hanno N sorgenti, che corrispondono alle fonti dei dati (ad esempio i portali delle regioni italiane). Da ogni sorgente è estratto uno o più dataset, che successivamente è mappato in un formato comune, detto *Tourpedia Data Model (TDM)*, attraverso un linguaggio di mapping, detto *Tourpedia Mapping Language (TML)*. Una volta convertiti nel formato comune, i dataset sono memorizzati (*Storage*) e poi resi accessibili attraverso una *Web API*. Infine, a partire dai dati è possibile costruire delle applicazioni Web (*Web Application*)



(Figura 5. L'architettura di *Tourpedia*)

4.4.1 Mapping per le attrazioni turistiche.

Poiché ogni dataset è fornito in un formato differente, è definita una mappatura formale che non solo indica le caratteristiche del dataset da scaricare, ma che fornisce anche indicazioni su quali campi all'interno del dataset contengono le informazioni più significative per la raccolta. Si tratta del *Tourpedia Data Model (TDM)*. Il *TDM* rinomina i campi comuni relativi ai diversi dataset in un unico tag, espresso nella lingua inglese.

In generale, tenendo conto di tutti i dataset, sono individuati 19 tipi diversi di informazione da estrarre e da inserire come campi nel database, a cui si aggiunge un campo dedicato all'id che permette di identificare univocamente una specifica attrazione. Esso indica la regione, il numero della risorsa e il numero dell'elemento all'interno del dataset originario. La tabella seguente mostra per ogni campo del *TDM* il tipo di dato corrispondente e ne fornisce una descrizione.

| Campo | Tipo | Descrizione |
|---------------|-------------|--|
| _id | string | ID |
| name | string | nome dell'attrazione |
| description | string | descrizione dell'attrazione |
| category | string | categoria dell'attrazione |
| address | string | indirizzo relativo all'attrazione |
| region | string | regione in cui si trova l'attrazione |
| province | string | provincia in cui si trova l'attrazione |
| city | string | nome della città in cui si trova l'attrazione |
| codistat | integer | codice assegnato dall'Istat al comune in cui si trova l'attrazione |
| postal-code | integer | codice postale della città in cui si trova l'attrazione |
| latitude | double | latitudine del punto in cui si trova l'attrazione |
| longitude | double | longitudine del punto in cui si trova l'attrazione |
| telephone | string | numero di telefono |
| fax | string | numero di fax |
| email | string | indirizzo email |
| url | string | Url associato all'attrazione |
| accessibility | string | condizioni di accessibilità al pubblico dell'attrazione |
| opening hours | string | orari di apertura relativi all'attrazione |
| entrance | string | tipo di ingresso (gratuito |

| | | |
|-------|--------|---|
| | | o no) e prezzo di ingresso |
| image | string | Url associata a un'immagine relativa all'attrazione |

(Tabella 5. *TDM*: campi, tipi di dato e descrizioni)

In totale, le risorse da recuperare e aggiornare sono 70.

4.4.2 Estensione del *Tourpedia Mapping Language*.

L'intera mappatura è specificata all'interno di un file di configurazione, attraverso un linguaggio specifico, il *Tourpedia Mapping Language (TML)*. All'interno del file di configurazione, per ogni sorgente, occorre specificare diverse informazioni, tra cui l'url da cui scaricare il dataset e il formato del dataset:

```
[MySource]
url = http://path/to/dataset.csv
format = CSV
```

Nella sua versione iniziale, il *TML* permette di estrarre informazione solo dai dati in formato *CSV*. Durante il lavoro di tesi, tuttavia, è stato esteso, per permettere l'estrazione di informazione anche dai dataset in formato *XLS*.

Riguardo agli altri formati (*XML*, *KML*) sono state create ad hoc funzioni specifiche per l'estrazione dei dati.

Di seguito un esempio relativo alla regione Toscana tratto dal file:

```
[Toscana_2]
url_attraction =
"http://mappe.regione.toscana.it/db-webgis/luoghi-spettacolo/example
_postgis.jsp?format=csv&_ga=2.262320655.261920716.1495633157-701941
840.1495028728"
dataset_attraction[format] = CSV
dataset_attraction[separator] = ","
dataset_attraction[first_data_row] = 3

attraction[name] = 3
attraction[lat] = 2
attraction[lon] = 1
attraction[address] = 4
attraction[city] = 5
attraction[province] = 6
attraction[category] = 8
attraction[description] = "Luoghi dello spettacolo"
attraction[region] = "Toscana"
```

Come si può notare, si tratta di un dataset in formato *CSV* che usa virgole come

delimitatori e relativo ai luoghi dello spettacolo. Le principali informazioni da esso estratte sono il nome, la latitudine e la longitudine, l'indirizzo, la città, la provincia e la categoria di luogo che riguardano i dati contenuti. I numeri assegnati a ogni campo indicano il numero del campo del file *CSV* in cui si trovano queste informazioni.

Il *TML* è stato esteso, inoltre, per permettere l'estrazione di informazioni trattate in modo più complesso nel dataset originario. È stata aggiunta la possibilità di segnalare che un certo tipo di informazione si trova in due o più campi del dataset al fine di raccoglierla in un unico campo nel database. Nel caso in cui latitudine e longitudine siano presenti in un unico campo nel dataset originale, ciò è stato indicato dal *TML* per permettere di separarli durante il processo di estrazione. Anche il caso di Url non ben formate, in cui manca ad esempio, la parte iniziale, è stato segnalato dal *TML* per permettere il loro corretto inserimento nel database. Riguardo ad alcuni dataset, inoltre, è stato necessario segnalare nel *TML* il numero di riga a partire dalla quale estrarre le informazioni necessarie. Ciò è stato necessario per i file *XLS* e per alcuni *CSV* in cui vi erano alcune righe vuote prima delle informazioni da estrarre.

Il seguente esempio relativo alla Sardegna mostra l'estensione del *TML* per segnalare la presenza di latitudine e longitudine nello stesso campo del dataset originario (*<lon*) e di un Url non ben formata (*<url*).

```
[Sardegna_3]
url_attraction =
"http://nurnet.crs4.it/geoserver/nurnet/ows?service=WFS&srsName=EPS
G:4326&version=1.1.0&request=GetFeature&typeName=nurnet:menhir&maxF
eatures=7000&outputFormat=csv"
dataset_attraction[format]      = CSV
dataset_attraction[separator]   = ", "
dataset_attraction[encoding]    = utf8

attraction[name] = 7
attraction[city] = 4
attraction[category] = 2
attraction[description] = "Menhir"
attraction[latitude] = 13<lon
attraction[url] = 12<url
attraction[region] = "Sardegna"
```

L'esempio seguente, invece, mostra come il *TML* indica la presenza dello stesso tipo di informazione, in questo caso numero di telefono e email, su più campi del dataset originario.

```
[Umbria_9]
```

```
url_attraction =
"http://dati.umbria.it/dataset/ec44f060-3225-4e14-a565-3f744cd2bf42
/resource/35883822-d131-44df-a627-e19086b42b23/download/2016opendat
aecomusei.csv"
dataset_attraction[format]      = CSV
dataset_attraction[separator]   = ";"
dataset_attraction[lastmodified] = 11

attraction[name] = 0
attraction[city] = 1
attraction[telephone] = 3or4
attraction[email] = 5or6
attraction[url] = 7
attraction[description] = "Ecomusei"
attraction[region] = "Umbria"
```

4.4.3 Arricchimento dei dati.

Tourpedia permette anche l'arricchimento delle informazioni presenti. Al momento, l'unico tipo di arricchimento fornito consiste nel recupero delle coordinate geografiche dell'attrazione, nel caso in cui queste non dovessero essere rese disponibili nel dataset originario. In questa evenienza, infatti, l'indirizzo e la città di un elemento vengono inviate all'*API* di *Google Maps* che ne fornisce le corrispondenti latitudine e longitudine.

È possibile, talvolta, che per qualunque ragione, una determinata risorsa relativa a una regione non sia più disponibile o recuperabile. In questo caso, il sistema recupera i dati precedentemente memorizzati.

CAPITOLO 5

COSTRUZIONE DI UN'ONTOLOGIA SULLE ATTRAZIONI TURISTICHE A PARTIRE DAGLI OPEN DATA

Dopo aver raccolto i diversi dataset relativi alle attrazioni turistiche, il passo successivo è consistito nel cercare di rappresentarli in modo formale attraverso la definizione di una ontologia. A tale scopo, sono stati presi in considerazione i campi *description* e *category*, relativi ad ogni risorsa, i quali potenzialmente possono contenere informazioni relative alla categoria o classe di appartenenza di ciascuna attrazione turistica. Il campo *description* fornisce l'indicazione del tipo principale di attrazione descritta dal dataset originario. Il valore del campo *description* è estratto direttamente dal nome del dataset, oppure da una macrocategoria inserita all'interno del dataset stesso. Il campo *category*, invece, indica quali tipologie o sottoclassi specifiche sono presenti nel dataset.

Sulla base dei due campi *description* e *category*, dunque, è stata realizzata una mappatura che mostra tutte le diverse descrizioni e categorie relative all'insieme dei dati raccolti e le mette in relazione con le singole regioni, indicando cioè, da quale regione provengono.

La seguente tabella mostra un esempio relativo ai dataset i cui elementi hanno *Biblioteche* come *description*.

| | Lazio | Lombardia | Toscana | Trentino | Marche |
|------------------------------------|-------|-----------|---------|----------|--------|
| Importante non specializzata | x | | | | x |
| Istituto di insegnamento superiore | x | | | | x |
| Nazionale | x | | | | |
| Non specificata | x | | | | |

| | | | | | |
|---------------|---|--|--|--|---|
| Pubblica | x | | | | x |
| Scolastica | x | | | | |
| Specializzata | x | | | | x |

(Tabella 6. Dataset che hanno *Biblioteche* come *description*)

A contenere la descrizione *Biblioteche* sono i dataset delle regioni Lazio, Lombardia, Toscana, Trentino e Marche. Le regioni Lazio e Marche presentano, inoltre, le categorie *Importante non specializzata*, *Istituto di insegnamento superiore*, *Pubblica*, *Specializzata*. Il Lazio presenta in più anche le categorie *Nazionale*, *Non specificata*, *Scolastica*.

In tutto, sono presenti settantacinque distinti valori di *description*. Quasi tutti, tranne 19, presentano una categorizzazione interna, ovvero hanno dei valori assegnati al campo *category*.

5.1 Criteri per la definizione dell'ontologia.

Per creare un'ontologia che descriva formalmente questo vasto e eterogeneo insieme di dati, o anche solo parte di esso, è necessario stabilire dei criteri da poter applicare in maniera standard all'intero insieme di dati.

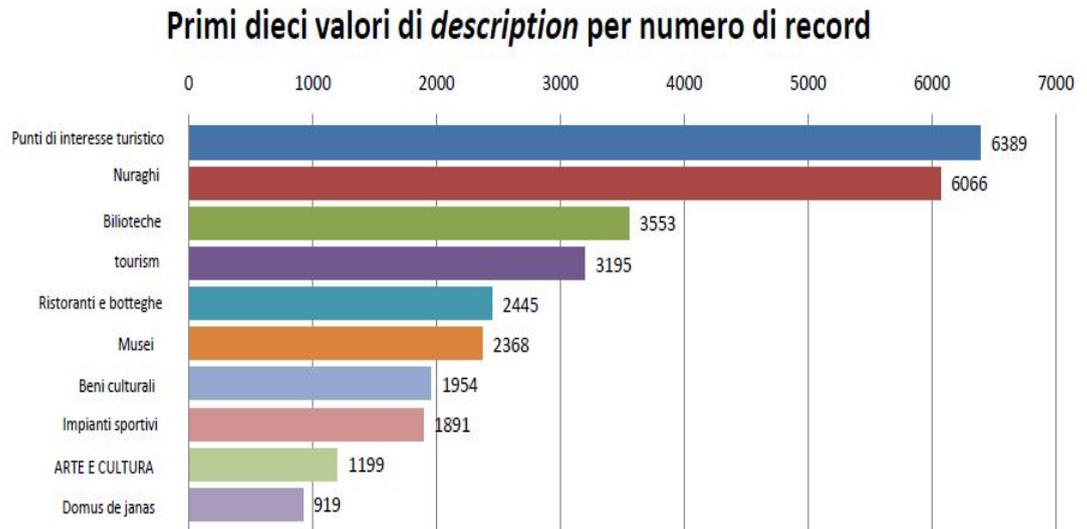
In primo luogo, si è scelto di definire un tipo base, comune a tutti i sottotipi e nominarlo *TouristAttraction*. Poi si è pensato di considerare tutti i distinti valori di *description* e di *category*, come tipi che sono sottoclassi del tipo *TouristAttraction*.

I tipi derivati da *description* vengono considerati sempre come sottoclassi dirette di *TouristAttraction*. I tipi derivati dai valori di *category* sono ovviamente considerati sottoclassi dei primi. Questo permette di attenersi al modo in cui i dati sono organizzati nel dataset originario. All'interno di ogni sottoclasse diretta di *TouristAttraction* le sue eventuali sottoclassi possono essere riorganizzate gerarchicamente, ad esempio, dando loro delle sopraclassi o compattandole in un'unica sottoclasse. Per quanto riguarda le proprietà dei vari tipi, esse sono scelte a partire dagli altri campi del dataset settati per ogni risorsa. Altre proprietà possono essere aggiunte per descrivere caratteristiche di tipi e sottoclassi che non sono

direttamente legate ad un preciso campo all'interno del dataset originario.

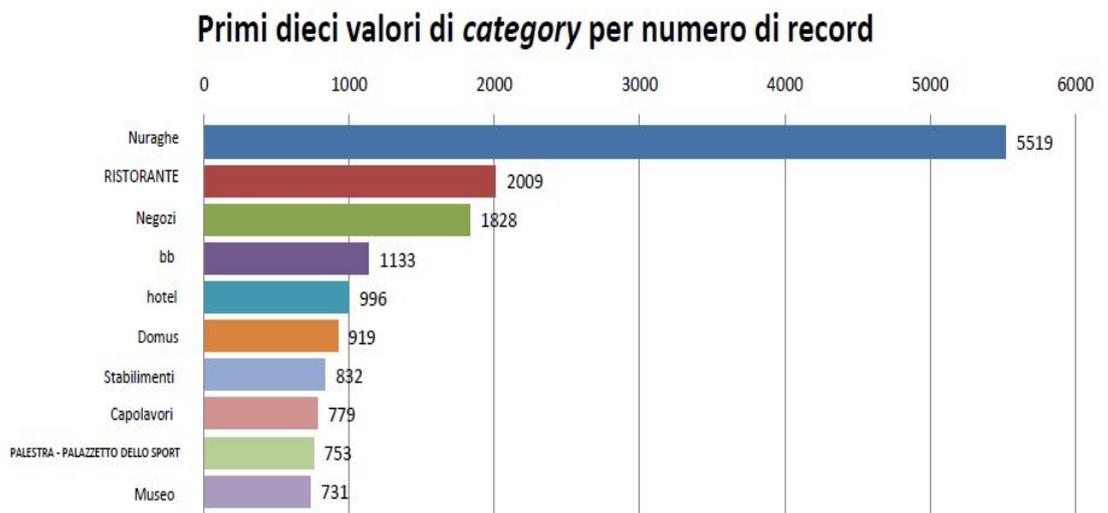
I nomi dei tipi vengono tradotti nel corrispondente in lingua inglese. Per la realizzazione dell'ontologia è stato utilizzato *Protege* (v. 3.3.1).

Il seguente grafico mostra i primi dieci valori di *description* per numero di record del database *Attrazioni* da essi descritto.



(Figura 6. Primi dieci valori di *description* per numero di record)

Il grafico che segue, invece, mostra i primi dieci valori di *category* per numero di record del database che hanno tali valori.



(Figura 7. Primi dieci valori di *category* per numero di record)

Data la grosse mole di categorie e sottocategorie, la definizione dell'ontologia è stata applicata ai soli casi delle Marche e della Sardegna.

5.2 Primo Caso d'uso: un'ontologia per la regione

Marche.

La regione Marche presenta quattro tipi principali di attrazioni diverse indicate dal campo *description*, ovvero *Punti di interesse turistico*, *Locali storici*, *Musei* e *Biblioteche*.

Nella costruzione dell'ontologia essi sono stati considerati come sottoclassi dirette del tipo *TouristAttraction* con i nomi rispettivamente di *TouristPOI*, *Club*, *Museum* e *Library*. Per ognuno di questi tipi sono definite nell'ontologia specifiche proprietà che riprendono quelle definite nel database *Attrazioni* per ogni risorsa. Quelle comuni a tutti i tipi, ovvero *latitude*, *longitude*, *city*, *codIstat*, *name*, *postalCode*, *province*, *region* e *street*, sono state assegnate al tipo *TouristAttraction* che li ingloba. Le proprietà *telephone*, *url* e *email* sono state assegnate a *Museum*, *Library* e *TouristPOI*. I tipi *Museum* e *TouristPOI* presentano in più le proprietà *image* e *openingHours*. Il tipo *TouristPOI* ha in più la proprietà *publicAccess*, mentre *Museum* ha in più la proprietà *entrance*. È stata, inoltre, definita la proprietà *historic* ed è stata assegnata a *Club* e *TouristPOI*, mentre la proprietà *openAir* è stata aggiunta per il tipo *Market*.

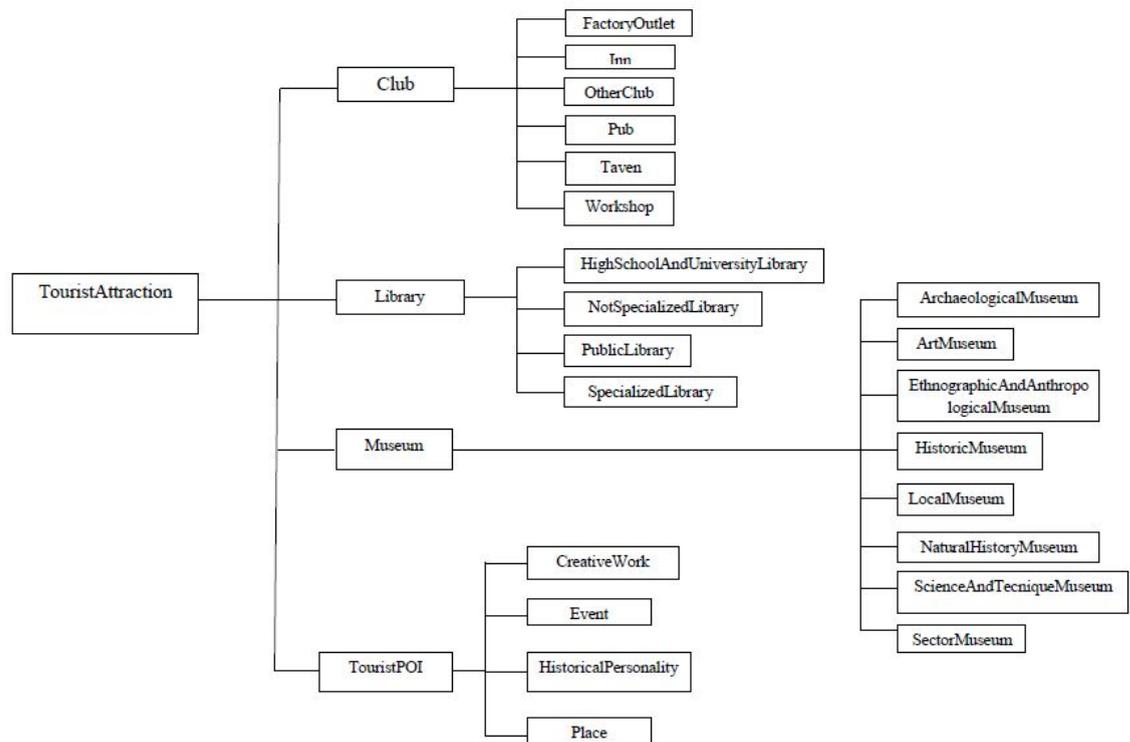
Ognuno di questi tipi di attrazione presenta una categorizzazione interna, specificata all'interno del database *Attrazioni* dai valori del campo *category*.

A possedere la categorizzazione più eterogenea è la classe *TouristPOI*, che comprende 45 categorie diverse. Si passa dai *Centri Benessere*, ai *Mercati pubblici all'aperto*, ai *Parchi divertimento*, alle *Spiagge*, ai *Porti*, ai *Fari*.

Alcune categorie, inoltre, hanno nomi poco rappresentativi delle entità descritte. Ad esempio, la categoria *Sports* potrebbe indicare indifferentemente impianti sportivi, oppure associazioni sportive. Per questo motivo, mentre per i tipi *Club*, *Museum* e *Library* è stata mantenuta nell'ontologia la categorizzazione originaria, per *TouristPOI* è stata riorganizzata la categorizzazione interna individuando tre sottoclassi principali, ovvero *CreativeWork*, *Event*, *HistoricalPersonality* e *Place*. Inoltre, i nomi di alcuni tipi sono stati modificati per descrivere meglio il tipo di entità da essi rappresentate.

Per ogni tipo e proprietà, nell'ontologia è fornita anche una descrizione.

La figura 8 mostra i diversi tipi di *TouristAttraction* e le loro sottoclassi dirette definiti nell'ontologia della regione Marche. Per motivi di leggibilità della figura sono state omesse le sottoclassi di *CreativeWork*, *Event*, *HistoricalPersonality* e *Place*.



(Figura 8. Tipi di attrazioni e sottoclassi nell'ontologia delle Marche)

5.3 Secondo Caso d'uso: un'ontologia per la regione Sardegna.

La regione Sardegna presenta diciannove tipi principali di attrazioni diverse indicate dal campo *description*, ovvero *aeroway*, *tourism*, *historic*, *natural*, *amenity*, *building*, *place*, *landuse*, *highway*, *public_transport*, *user_defined*, *Nuraghi*, *Menhir*, *Domus da janas*, *Tombe dei giganti*, *Dolmen*, *Villaggi*, *Culto dell'acqua*, *Unicum* e *tafoni*. Si può notare che, a differenza della regione Marche (5.1), la maggiorparte di essi ha nomi in inglese e questo ha reso più facile il loro inserimento nell'ontologia che, per ragioni di maggiore standardizzazione, è definita appunto, in questa lingua. Nel caso in cui il nome sia in italiano e si riferisca a categorie specifiche, invece, esso è stato

mantenuto nell'ontologia poiché non esiste il corrispettivo inglese e si tratta, nel caso di *Nuraghi*, *Domus da janas* e *Tombe dei giganti*, di attrazioni tipiche della Sardegna.

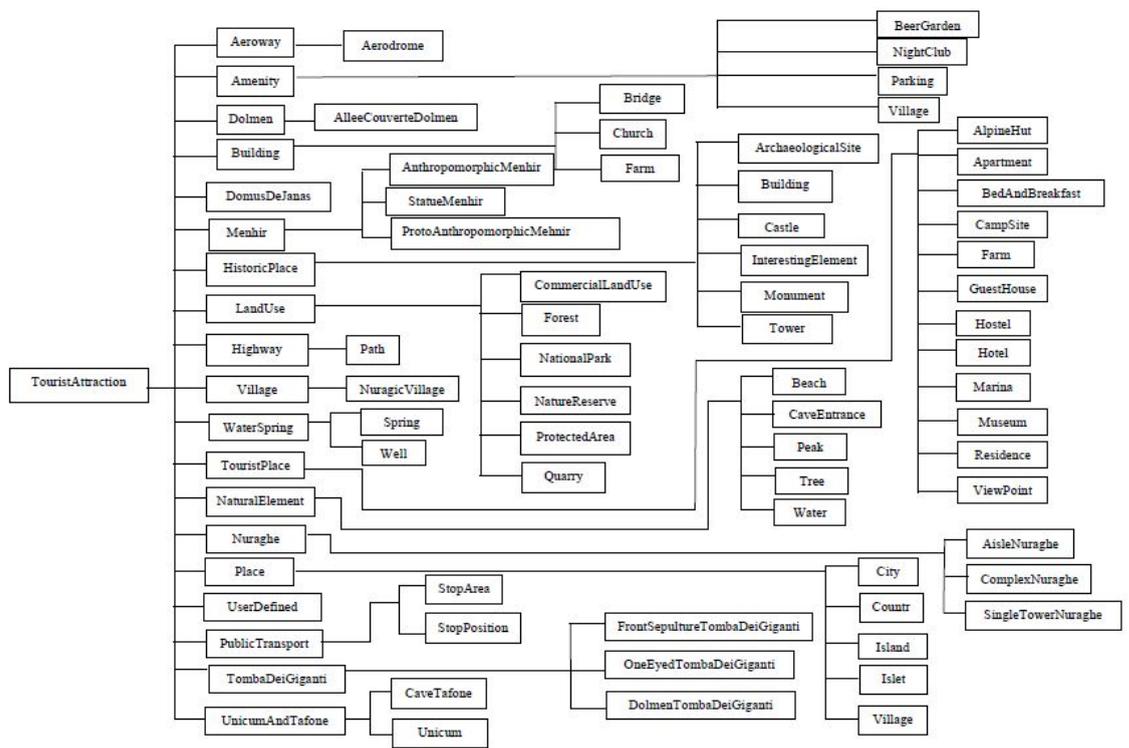
Nella costruzione dell'ontologia i tipi principali individuati sopra sono stati considerati come sottoclassi dirette del tipo *TouristAttraction*. Per ognuno di questi tipi sono definite nell'ontologia specifiche proprietà che riprendono quelle definite nel database *Attrazioni* per ogni risorsa. Rispetto al caso delle Marche (5.1) il database fornisce un numero minore di proprietà per questi tipi di attrazione. Non sono presenti, infatti, le proprietà *codIstat*, *postalCode*, *province*, *telephone* e *email*. Le proprietà fornite sono, invece, *city*, *latitude*, *longitude*, *name*, *region*, *street* e *url*. Esse sono applicabili a ognuno dei tipi di attrazione definito e, per questo, viene loro assegnato come dominio il tipo più generico *TouristAttraction*.

Alcuni tra i tipi principali di attrazione, inoltre, hanno nomi poco adeguati a un'ontologia. Ad esempio, *natural*, *historic* e *tourism* sembrano nomi più consoni per proprietà piuttosto che per tipi. Per questo, nell'ontologia sono stati modificati in *NaturalElement*, *HistoricPlace*, *TouristPlace*. Anche il nome *culto dell'acqua* è stato reso nell'ontologia come *WaterSpring*.

Ognuno dei tipi principali di attrazione turistica presenta una categorizzazione interna, specificata nel database *Attrazioni* dai valori del campo *category*. Rispetto alla regione Marche (5.1), si tratta di una categorizzazione più precisa e meno eterogenea che non ha richiesto una ulteriore riorganizzazione e gerarchizzazione delle varie sottoclassi nell'ontologia. Al massimo, sono stati eliminati dall'ontologia quei valori di *category* uguali a quelli di *description* e che descrivevano lo stesso tipo di entità. Per i valori di *description* *Nuraghi*, *Menhir*, *Domus da janas*, *Tombe dei giganti*, *Dolmen*, *Villaggi*, infatti, erano presenti valori di *category* con lo stesso nome. Queste categorie sono state eliminate dall'ontologia poiché inglobate dai tipi designati da *description*.

Per ogni tipo e proprietà definiti, nell'ontologia è fornita anche una descrizione.

La figura 9 mostra tutti i tipi e le sottoclassi definiti nell'ontologia della regione Sardegna.



(Figura 9. Tipi e sottoclassi dell'ontologia della regione Sardegna)

CONCLUSIONI

In questo lavoro di tesi è stato effettuato uno studio sugli Open Data a livello italiano relativo alle attrazioni turistiche. Il presente studio ha dimostrato come vi sia una grossa eterogeneità nella rappresentazione dei dati: non solo ogni regione o provincia fornisce una rappresentazione personalizzata delle attrazioni turistiche, ma le diverse organizzazioni governative definiscono in modo differente il concetto di attrazione turistica. Questa eterogeneità dei dati è stata argomentata nella tesi attraverso un'analisi specifica dei diversi dataset e si è giunti alla conclusione che anche la qualità dei dati varia a seconda della sorgente. In alcuni casi si parla di veri e propri problemi redazionali, che richiederebbero tecniche sofisticate di correzione.

Per quanto riguarda la raccolta dei dati, il presente lavoro di tesi ha contribuito ad arricchire il portale *Tourpedia* di nuovi dati relativi alle attrazioni turistiche. Ciò aprirà sicuramente la strada verso la costruzione di applicazioni relative al turismo in generale e alle attrazioni turistiche in particolare. Ad esempio, sulla base dei dati raccolti in *Tourpedia*, potranno essere costruite applicazioni per un turismo sostenibile, cioè un tipo di turismo che cerca di valorizzare i luoghi meno famosi, le loro tradizioni, usi, costumi, eccetera.

Con l'ultima parte di questo lavoro di tesi - lo sviluppo delle ontologie - si è cercato di dimostrare praticamente l'eterogeneità dei dati e si è giunti alla conclusione che l'unica strategia adottabile per l'identificazione di ciò che può essere definito attrazione turistica consiste nel *declassare* il tipo *TouristAttraction* a proprietà. Questa strategia, adottata da *Schema.org*, permetterà di definire di volta in volta ciò che si intende per attrazione turistica.

Appendice

Tabella 1A: Dataset raccolti per ogni regione italiana.

Per ogni regione italiana è rappresentato un elenco con le *URL* dei dataset raccolti, per ognuno dei quali è indicato il formato, una breve descrizione e se è stato o meno selezionato come insieme di attrazioni turistiche.

| Regione | Dataset | Forma to | Descrizione | Selezio nato |
|--------------------|---|-------------|--|-----------------|
| Abruzzo | nessuno | | | |
| Basilicata | http://dati.comune.matera.it/storage/f/2014-02-22T17%3A48%3A27.188Z/siti-matera.csv | CSV | siti archeologici del comune di Matera | SI |
| | http://dati.comune.matera.it/storage/f/2014-02-22T18%3A55%3A03.355Z/beniculturalicomunali.csv | CSV | Beni culturali del comune di Matera | SI |
| | http://dati.comune.matera.it/storage/f/2015-03-22T12%3A23%3A58.931Z/nuove-immagini-chiese-rupestri-rev0.csv | CSV | Chiese rupestri del comune di Matera | SI |
| Calabria | nessuno | | | |
| Campania | nessuno | | | |
| Emilia-Ro magna | http://dati.comune.bologna.it/download/file/fid/3190 | CSV | Porte storiche e torri del comune di Bologna | SI |
| | http://dati.comune.bologna.it/download/file/fid/2644 | CSV | Fontane, ponti, canali del comune di Bologna | SI |
| | http://dati.comune.bologna.it/download/file/fid/3383 | CSV | Ville e dimore storiche di Bologna | SI |
| | http://dati.comune.bologna.it/download/file/fid/3480 | CSV | Musei, gallerie, luoghi storici, | SI |

| | | | | |
|-----------------------|---|-----|--|----|
| | | | teatri storici del comune di Bologna | |
| | http://dati.comune.bologna.it/download/file/fid/3380 | CSV | Mercati cittadini storici e itineranti del comune di Bologna | SI |
| | http://dati.comune.bologna.it/download/file/fid/3197 | CSV | Luoghi e personaggi di interesse turistico del comune di Bologna | SI |
| | http://dati.comune.bologna.it/download/file/fid/2846 | KML | Biblioteche comunali di Bologna | SI |
| | http://dati.comune.bologna.it/download/file/fid/2615 | CSV | Servizi igienici pubblici del comune di Bologna | NO |
| | http://dati.comune.bologna.it/download/file/fid/3248 | CSV | Parcheggi del comune di Bologna | NO |
| Friuli-Venezia Giulia | nessuno | | | |
| Lazio | http://dati.lazio.it/catalog/it/dataset/4ff54e24-d768-4e32-bdab-75752f224305/resource/ed9470a0-6d53-4660-8bda-3651705b3a6f/download/arene2013.csv | CSV | Arene | SI |
| | http://dati.lazio.it/catalog/it/dataset/4ff54e24-d768-4e32-bdab-75752f224305/resource/3345fa82-b5bd-4f0b-9aae-9780b794d2d8/download/salecinematografiche2013.csv | CSV | Sale cinematografiche | SI |
| | http://dati.lazio.it/catalog/it/dataset/4ff54e24-d768-4e32-bdab-75752f224305/resource/df17214 | CSV | Cinema-teatro | SI |

| | | | | |
|---------|---|-----|------------------------------|----|
| | 6-7580-4bde-8dce-d92482b107cf/download/cinemateatro2013.csv | | | |
| | http://dati.lazio.it/catalog/it/dataset/4ff54e24-d768-4e32-bdab-75752f224305/resource/07e1c4d0-ec8a-4ee5-b6c1-e562dd007249/download/multisala2013.csv | CSV | Multisala | SI |
| | http://dati.lazio.it/catalog/it/dataset/64281dd7-2222-4513-bdb7-0061c5d8cd25/resource/eb98bb58-3d7e-46b4-a6d2-305054ed8cc5/download/elencobibliotechelazio20142b.csv | CSV | Biblioteche | SI |
| | http://dati.lazio.it/catalog/it/dataset/d2c9dfdb-51af-4faf-bdd2-52ce662e63d8/resource/b4950fa1-75a7-4324-aa4c-f04e7cde04e9/download/censimentodeglispazidipubblicospettacolo2.csv | CSV | Spazi di pubblico spettacolo | SI |
| | http://dati.lazio.it/catalog/it/dataset/64281dd7-2222-4513-bdb7-0061c5d8cd25/resource/25f320a5-ee4f-486c-a8e7-408733c606c5/download/elencomusei20141.csv | CSV | Musei | SI |
| | http://dati.lazio.it/catalog/it/dataset/886075d1-b5a6-4ee8-ae74-38feba74108/resource/55f2c0d3-7860-42bb-8f04-c99d2f4ddd1e/download/parafarmaciereglazio.csv | CSV | Parafarmacie | NO |
| | http://dati.lazio.it/catalog/it/dataset/551579e1-a65d-4e7e-80d7-9bfe07fb2bdc/resource/e0497f9d-ddec-4983-b8d1-9701266d0631/download/farmaciereglazio.csv | CSV | Farmacie | NO |
| Liguria | http://www.regione.liguria.it/sep-servizi-online/catalogo-servizi-online/opendata/download/84/6839/48.html | CSV | Strutture sanitarie private | NO |

| | | | | |
|-----------|---|-----|--|----|
| | http://cartodownloadpubb.regione.liguria.it/webpubb/geoservices/download/repository_pubb//79866/Teatri.zip | KML | Teatri | SI |
| | http://cartodownloadpubb.regione.liguria.it/webpubb/geoservices/download/repository_pubb//79867/Siti_Archeologici.zip | KML | Siti archeologici | SI |
| | http://cartodownloadpubb.regione.liguria.it/webpubb/geoservices/download/repository_pubb//79868/Musei.zip | KML | Musei | SI |
| | http://cartodownloadpubb.regione.liguria.it/webpubb/geoservices/download/repository_pubb//79869/Gallerie_d_arte.zip | KML | Gallerie d'arte | SI |
| | http://cartodownloadpubb.regione.liguria.it/webpubb/geoservices/download/repository_pubb//79870/Borghi_Storici.zip | KML | Borghi storici | SI |
| | http://cartodownloadpubb.regione.liguria.it/webpubb/geoservices/download/repository_pubb//79871/Giardini.zip | KML | Giardini | SI |
| Lombardia | https://www.dati.lombardia.it/api/views/3syc-54zf/rows.csv?accessType=DOWNLOAD | CSV | Musei | SI |
| | https://www.dati.lombardia.it/api/views/k2xm-fcgg/rows.csv?accessType=DOWNLOAD | CSV | Biblioteche | SI |
| | https://www.dati.lombardia.it/api/views/kf9b-rj2t/rows.csv?accessType=DOWNLOAD | CSV | Architetture | NO |
| | https://www.dati.lombardia.it/api/views/q9ef-6jgc/rows.csv?accessType=DOWNLOAD | CSV | Luoghi della cultura in provincia di Monza-Brianza | SI |
| | https://www.dati.lombardia.it/api/views/4mr7-hfsh/rows.csv?accessType=DOWNLOAD | CSV | Beni culturali | SI |

| | | | | |
|--------|---|-----|--|----|
| | https://www.dati.lombardia.it/api/views/5dhz-9ubp/rows.csv?accessType=DOWNLOAD | CSV | Biblioteche della provincia di Monza-Brianza | SI |
| | https://www.dati.lombardia.it/api/views/4kkk-xzn5/rows.csv?accessType=DOWNLOAD | CSV | Musei della provincia di Monza-Brianza | SI |
| | https://www.dati.lombardia.it/api/views/cq9n-g2b3/rows.csv?accessType=DOWNLOAD | CSV | Cinema e teatri della provincia di Monza-Brianza | SI |
| | https://www.dati.lombardia.it/api/views/sd8x-w4h3/rows.csv?accessType=DOWNLOAD | CSV | Impianti sportivi | SI |
| | https://www.dati.lombardia.it/api/views/haxp-4dkd/rows.csv?accessType=DOWNLOAD | CSV | Grandi strutture di vendita | SI |
| Marche | http://goodpa.regione.marche.it/dataset/d8444ecf-586c-4410-9a78-f24fcf938822/resource/4b48dfbf-59a6-40a6-8be1-3c3aaa6e30e7/download/8poisettembre.xls | XLS | Punti di interesse turistico | SI |
| | http://goodpa.regione.marche.it/dataset/15e2a15a-6714-434d-9257-6641d97859f0/resource/27567e3e-fedf-4011-8fd7-60b41bce9ed3/download/localistorici.xls | XLS | Locali storici | SI |
| | http://goodpa.regione.marche.it/dataset/22e82da2-a4d1-4e83-928d-e384b85da978/resource/0155710f-f2b6-412d-a138-1db184c7c3ab/download/biblioteche.xls | XLS | Biblioteche | SI |
| | http://goodpa.regione.marche.it/dataset/b8c4dfa8-a0ab-46b6-b078-3def5bbe6d35/resource/cd75ebf3-809e-475e-b91e-0b3e6fe07ad6/download/musei.xls | XLS | Musei | SI |
| Molise | nessuno | | | |

| | | | | |
|----------|---|-----|--|----|
| Piemonte | http://www.dati.piemonte.it/catalogodati/scarica.html?idallegato=458 | CSV | Musei | SI |
| | http://www.dati.piemonte.it/catalogodati/scarica.html?idallegato=459 | CSV | Beni architettonici di interesse | SI |
| | http://www.dati.piemonte.it/catalogodati/scarica.html?idallegato=1141 | CSV | Centri commerciali | SI |
| | http://www.dati.piemonte.it/catalogodati/scarica.html?idallegato=1135 | CSV | Medie e grandi strutture di vendita | NO |
| | http://www.dati.piemonte.it/catalogodati/scarica.html?idallegato=800 | CSV | Farmacie pubbliche e private | NO |
| | http://www.dati.piemonte.it/catalogodati/scarica.html?idallegato=802 | CSV | Strutture sanitarie regionali pubbliche | NO |
| | http://www.dati.piemonte.it/catalogodati/scarica.html?idallegato=803 | CSV | Strutture sanitarie private e case di cura | NO |
| | http://www.dati.piemonte.it/catalogodati/scarica.html?idallegato=801 | CSV | Strutture sanitarie equiparate | NO |
| | http://www.dati.piemonte.it/catalogodati/scarica.html?idallegato=805 | CSV | Residenze Sanitarie assistenziali private | NO |
| | http://www.dati.piemonte.it/catalogodati/scarica.html?idallegato=808 | CSV | Unità Organizzative Funzionali delle strutture sanitarie pubbliche | NO |
| | http://www.dati.piemonte.it/catalogodati/scarica.html?idallegato=809 | CSV | Punti di erogazione pubblici regionali di | NO |

| | | | | |
|----------|---|-----|---|----|
| | | | prestazioni sanitarie | |
| Puglia | http://www.dataset.puglia.it/dataset/fb554ebc-5bd6-48ea-8b01-ae14ea555c81/resource/1d30aec5-b5b6-451e-ad82-2eb390809002/download/reportluoghiinteresse.csv | CSV | Luoghi di interesse turistico, culturale, naturalistico | SI |
| | http://www.dataset.puglia.it/dataset/ad9df93d-46fc-4486-83d0-f62cd3845e60/resource/bc957644-7994-47f9-942d-2ce6a600cc45/download/stabilimentibalneari.csv | CSV | Stabilimenti balneari | SI |
| | http://www.dataset.puglia.it/dataset/anagrafe-strutture-sanitarie/resource/c34b0881-3696-4c22-9501-d5fa5252d591 | CSV | Strutture sanitarie | NO |
| | http://www.dataset.puglia.it/dataset/anagrafe-asl/resource/bcba8e5e-e310-4b96-ba14-21f75ccc e1d6 | CSV | ASL | NO |
| | http://www.dataset.puglia.it/dataset/farmacie-territoriali-innovapuglia/resource/03710055-0531-45ca-be6d-79e1773d35ec | CSV | Farmacie convenzionate | NO |
| Sardegna | http://webgis2.regione.sardegna.it:80/geoserver/dbu/ows?service=WFS&version=1.0.0&request=GetFeature&typeName=dbu:IDT_OP08V_POI_EXPORT&outputFormat=CSV | CSV | Luoghi di interesse turistico, culturale, naturalistico | SI |
| | http://nurnet.crs4.it/geoserver/nurnet/ows?service=WFS&srsName=EPSG:4326&version=1.1.0&request=GetFeature&typeName=nurnet:dolmen&maxFeatures=7000&outputFormat=csv | CSV | Dolmen | SI |
| | http://nurnet.crs4.it/geoserver/nurnet/ows?service=WFS&srsName=EPSG:4326&version=1.1. | CSV | Unicum e tafoni | SI |

| | | | | |
|---------|---|-----|-----------------------|----|
| | 0&request=GetFeature&typeName=nurnet:acqua&maxFeatures=7000&outputFormat=csv | | | |
| | http://nurnet.crs4.it/geoserver/nurnet/ows?service=WFS&srsName=EPSG:4326&version=1.1.0&request=GetFeature&typeName=nurnet:acqua&maxFeatures=7000&outputFormat=csv | CSV | Fonti d'acqua | SI |
| | http://nurnet.crs4.it/geoserver/nurnet/ows?service=WFS&srsName=EPSG:4326&version=1.1.0&request=GetFeature&typeName=nurnet:villaggi&maxFeatures=7000&outputFormat=csv | CSV | Villaggi archeologici | SI |
| | http://nurnet.crs4.it/geoserver/nurnet/ows?service=WFS&srsName=EPSG:4326&version=1.1.0&request=GetFeature&typeName=nurnet:domus&maxFeatures=7000&outputFormat=csv | CSV | Domus de janas | SI |
| | http://nurnet.crs4.it/geoserver/nurnet/ows?service=WFS&srsName=EPSG:4326&version=1.1.0&request=GetFeature&typeName=nurnet:tombe&maxFeatures=7000&outputFormat=csv | CSV | Tombe dei giganti | SI |
| | http://nurnet.crs4.it/geoserver/nurnet/ows?service=WFS&srsName=EPSG:4326&version=1.1.0&request=GetFeature&typeName=nurnet:menhir&maxFeatures=7000&outputFormat=csv | CSV | Menhir | SI |
| | http://nurnet.crs4.it/geoserver/nurnet/ows?service=WFS&srsName=EPSG:4326&version=1.1.0&request=GetFeature&typeName=nurnet:nuraghi&maxFeatures=7719&outputFormat=csv | CSV | Nuraghi | SI |
| Sicilia | nessuno | | | |
| Toscana | http://dati.toscana.it/dataset/b68 | CSV | Ospedali | NO |

| | | | | |
|--|---|-----|---|----|
| | 70628-2217-413a-bfd8-adf147a893a2/resource/c14d0f56-665c-46d7-8b5c-bd98bf499aca/download/ospedalierefeature.csv | | | |
| | http://dati.toscana.it/dataset/b6870628-2217-413a-bfd8-adf147a893a2/resource/39231648-3cc0-4a67-adfd-ee596f7e19c3/download/riaibitativefeature.csv | CSV | Strutture riabilitative | NO |
| | http://dati.toscana.it/dataset/b6870628-2217-413a-bfd8-adf147a893a2/resource/8ded0b73-24c7-451a-8972-2a92e6bbaf0d/download/territorialifeature.csv | CSV | Strutture sanitarie territoriali | NO |
| | http://www301.regione.toscana.it/bancadati/farmacie/index.csv?_ga=2.268530097.1155031691.1499010437-701941840.1495028728 | CSV | Farmacie | NO |
| | http://mappe.regione.toscana.it/db-webgis/biblio/example_postgis.jsp?format=csv&_ga=2.181972322.412019626.1495551945-701941840.1495028728 | CSV | Biblioteche | SI |
| | http://www.datiopen.it/it/opendata/Regione_Toscana_Universit_e_conservatori | CSV | Università pubbliche e conservatori | NO |
| | http://dati.provincia.fi.it/it/storage/f/2016-06-29T070307/SERVIZI-E-UTILIT%C3%80-punti-di-informazione-turistica.csv?_ga=2.194778605.1155031691.1499010437-701941840.1495028728 | CSV | Punti di informazione turistica a Firenze e nei comuni del territorio | NO |
| | http://mappe.regione.toscana.it/db-webgis/luoghispettacolo/example_postgis.jsp?format=csv&_ga=2.262320655.261920716.1495633157-701941840.1495028728 | CSV | Luoghi dello spettacolo | SI |
| | http://mappe.regione.toscana.it/ | CSV | Musei | SI |

| | | | | |
|----------------------------|---|------|---|----|
| | db-webgis/musei/example_post gis.jsp?format=csv&_ga=2.839 57593.1351149729.149563366 7-701941840.1495028728 | | | |
| | http://dati.toscana.it/dataset/75d6183f-8b8d-4150-a1f0-f69eb984c238/resource/80020491-6b17-4e1b-94be-bd8d1a41f62a/download/esportazioneadesioni10102016.csv | CSV | Ristoranti e botteghe | SI |
| Trentino- Alto Adige | http://dati.trentino.it/dataset/e491b841-cc4a-47e3-9d6e-0c54436bb98f/resource/e572c92e-05dc-41b7-ba22-e9898bddc85c/download/apttrentocensimentoappwudefinitivopart2.csv | CSV | Punti di interesse di Trento e della comunità della Valle dei Laghi | SI |
| | http://dati.trentino.it/dataset/be822323-a6fb-4ae3-a5b7-abbcf8934f2/resource/71d4e23c-04fa-4b43-90b1-2d58f0b49f20/download/wu2013valsugana.json | JSON | Punti di interesse della Valsugana | SI |
| | http://dati.trentino.it/dataset/be822323-a6fb-4ae3-a5b7-abbcf8934f2/resource/71d4e23c-04fa-4b43-90b1-2d58f0b49f20/download/wu2013valsugana.json | JSON | Punti di interesse dell'Altopiano di Pinè e Val di Cembra | SI |
| | http://www.commercio.provincia.tn.it/binary/pat_commercio/marchi_prodotto/Elenco_osterie_tipiche_31ott13.1387207601.xls | XLS | Osterie tipiche della provincia di Trento | SI |
| | http://dati.trentino.it/dataset/c9270e2a-041f-449b-a27b-100bb02de4b2/resource/0de24153-8dc2-4b98-84c3-2cda64a37658/download/albobotteghistorichefeb2017.csv | CSV | Botteghe storiche della Provincia di Trento | SI |
| | http://dati.trentino.it/dataset/e34151d9-6eea-4078-a17a-e563a4c9738e/resource/a1964337-79ee-4498-8d80-af481ccc491e/download/wu2013poi.json | JSON | Punti di interesse della Provincia di Trento | SI |

| | | | | |
|--------|---|-----|--|----|
| | http://www.cbt.biblioteche.provincia.tn.it/oseegenius/opendata/biblioteche.xml | XML | Biblioteche della provincia di Trento | SI |
| Umbria | http://dati.umbria.it/dataset/ab39e23d-540f-4aa1-b27f-3fb244837545/resource/1ae8a517-4083-44e2-b141-0c8e613b6052/download/attrattoricsvitatattrattoriiti.csv | CSV | Maggiori attrattori turistici | SI |
| | http://dati.umbria.it/dataset/32cca8e8-4e6e-43ed-b7df-2f577de8ac4d/resource/520895ad-05d8-4734-94f7-7a6b8acfea36/download/iatesvilit.zipiatitit.csv | CSV | Uffici IAT | SI |
| | http://dati.umbria.it/dataset/773e2bd0-b0ce-422e-80be-9da831f77f26/resource/c145f43d-b7e7-4033-b587-70c16bf8ac7b/download/bibliotechecomunalisezioneragazzicleaned.csv | CSV | Biblioteche comunali con sezione ragazzi | SI |
| | http://dati.umbria.it/dataset/88cb23f2-0e34-4e0c-8fc3-4c90ca872d5d/resource/f5e7f052-903e-4bad-8a62-98ec4f5ed418/download/reteregionalebibliotechenpl.csv | CSV | Biblioteche con sezione "Nati per leggere" | SI |
| | http://dati.umbria.it/dataset/c3d9e1c7-67aa-4ea7-bf95-b5ec299db957/resource/c2e0a071-ba78-4d07-a574-2d8907ce3a01/download/elencobibliotechesbn.csv | CSV | Biblioteche SBN | SI |
| | http://dati.umbria.it/dataset/2a580dcb-1319-4726-8fa1-131dd3eef77c/resource/124381ae-1e1b-485e-93ee-69c667c91635/download/opendataanagraficacomunalicleaned.csv | CSV | Biblioteche comunali | SI |
| | http://dati.umbria.it/dataset/a06d6b87-293d-4a18-ac6d-6c49924e832c/resource/ccdc8f3e-8b48-4072-adc4-2e686bfe90e3/download/luoghiculturasettembre20 | CSV | Musei e luoghi della cultura | SI |

| | | | | |
|---------------|---|-----|--|----|
| | 16cleaned.csv | | | |
| | http://dati.umbria.it/dataset/cbb56a52-5707-4397-9853-95c0ab8f026c/resource/ccd7d772-2973-4ae4-ac77-c8df0d35945e/download/museichehannostoffa.csv | CSV | Musei che hanno stoffa | SI |
| | http://dati.umbria.it/dataset/ec44f060-3225-4e14-a565-3f744cd2bf42/resource/35883822-d131-44df-a627-e19086b42b23/download/2016opendataecomusei.csv | CSV | Ecomusei | SI |
| | http://dati.umbria.it/dataset/1b0ef5fa-a3f5-465e-8aab-c6ea3531060c/resource/4e88408f-4dd6-45ce-a0a6-236081367c78/download/strutturetermaliregioneumbriastrutturetermaliregioneumbria.csv | CSV | Strutture termali | SI |
| | http://dati.umbria.it/dataset/097fa968-78a9-412f-ae60-5c246c9d4e43/resource/6c06bc75-488c-4d87-b71b-dedf07640643/download/datianagraficistrutture.csv | CSV | Strutture sanitarie pubbliche e private | NO |
| Valle D'Aosta | http://www.regione.vda.it/statistica/Opendata/Opendataregione/Openlink.asp?id=14 | CSV | Strutture sanitarie, socio-sanitarie, socio-educative, assistenziali | NO |
| Veneto | http://idt.regione.veneto.it/app/metacatalog/index?deflevel=165 | XML | Stazioni ferroviarie | NO |
| | http://dati.veneto.it/storage/f/2014-01-16T14%3A03%3A54.281Z/veneto-cinema-2012.xls | XLS | Sale cinematografiche | SI |
| | http://dati.veneto.it/dataset/58b96ef0-f088-4d5b-ac90-53183b489149/resource/82c990ce-bd26 | XLS | Spazi teatrali | SI |

| | | | | |
|--|--|--|--|--|
| | -4b21-9c79-78ecedcffbe7/download/teatriveneto42014.xls | | | |
|--|--|--|--|--|

Bibliografia

(AA. VV. 2010) Department of Economic and Social Affairs of the United Nations Secretariat. 2010. *International Recommendations for Tourism Statistics 2008*. New York, United Nations Publication. Disponibile su:
https://unstats.un.org/unsd/publication/Seriesm/SeriesM_83rev1e.pdf#page=12 [Data accesso: 15 giugno 2017]

(AA. VV. 2014) United Nations World Tourism Organization. 2014. *Glossary of tourism terms*. Disponibile su:
<http://cf.cdn.unwto.org/sites/all/files/Glossary+of+terms.pdf> [Data accesso: 15 giugno 2017]

(AA. VV. 2016) Committee on Tourism and Competitiveness of the United Nations World Tourism Organization. 2016. *Report of the Committee on Tourism and Competitiveness*. Madrid, World Tourism Organization. Disponibile su:
<http://cf.cdn.unwto.org/sites/all/files/docpdf/generalprogrammeofworkdmgt.pdf> [Data accesso: 17 giugno 2017]

(Bahramian e Abbaspour 2015) Bahramian, Z, e R. Ali Abbaspour. 2015. *An ontology-based tourism recommender system based on spreading activation model*. In: Arefi, H., e M. Motagh. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XL-1/W5*, pp. 83-90. Disponibile su:
<http://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XL-1-W5/83/2015/iSprsarchives-XL-1-W5-83-2015.pdf> [Data accesso: 1 luglio 2017]

(Harris e Howard 1996) Harris, Robert, e Joy Howard. 1996. *Dictionary of Travel, Tourism and Hospitality Terms*. Elsternwick, Hospitality Press Pty Ltd.

(Höpken e Ciaran 2008) Höpken Wolfram, e Clissmann Ciaran. 2008. *HarmoNET*

Tourism Ontology User Manual. Disponibile su:

http://euromuse.harmonet.org/c/document_library/get_file?p_1_id=15412&folderId=16041&name=DLFE-18.pdf [Data accesso: 1 luglio 2017]

(Mancinelli 2017) Mancinelli, M. L. 2017. *La catalogazione del patrimonio archeologico, architettonico, paesaggistico, storico-artistico e demoetnoantropologico: Principi generali di Catalogazione*. MiBACT ICCD.

Disponibile su: <http://www.iccd.beniculturali.it/getFile.php?id=5764> [Data accesso: 22 giugno 2017]

(Marchioro 2011) Marchioro, Stefano. 2011. *Destinazione turistica, destination management e le DMO*. Ferrara. Disponibile su:

http://www.unife.it/lettere/ filosofia/comunicazione/insegnamenti/marketing/materiale_didattico/DM_Ferrara_5_dicembre_2011.pdf [Data accesso: 15 giugno 2017]

(Medlik 2003) Medlik, S. 2003. *Dictionary of Travel, Tourism and Hospitality*. Burlington, Elsevier Science.

(Prantner 2004) Prantner Kathrin. 2004. *OnTour: The Ontology*. Disponibile su:

<http://e-tourism.deri.at/ont/docu2004/OnTour%20-%20The%20Ontology.pdf> [Data accesso: 1 luglio 2017]

(Tosqui-Lucks e Dias Da Silva 2012) Tosqui-Lucks Patrícia, e Dias Da Silva Bento Carlos. 2012. *Structuring an Ontology of the Basic Vocabulary of Tourism*.

“International Journal of Information and Education Technology” vol. 2, no. 4, pp. 331-334. Disponibile su: <http://www.ijiet.org/papers/144-T046.pdf> [Data accesso: 1 luglio 2017]

(Varra 2012) Varra, Lucia (a cura di). 2012. *Dal dato diffuso alla conoscenza condivisa*. Firenze, University Press. Disponibile su:

<https://books.google.it/books?id=5tKk5EILyhwC&printsec=frontcover&dq=isbn:8866551767&hl=it&sa=X&ved=0ahUKEwjnso-ZpvLUAhUBOxQKHV-wAlIQ6AEIJz>

AA#v=onepage&q&f=false [Data accesso: 15 giugno 2017]

Sitografia

5stardata.info, *5-star Open Data*. Disponibile su: <http://5stardata.info/en/> [Data accesso: 20 giugno 2017]

Developer-tripadvisor.com, *Subcategories, Groups, and Types*. Disponibile su: <https://developer-tripadvisor.com/content-api/business-content/categories-subcategories-and-types/> [Data accesso: 1 luglio 2017]

Opendefinition.org, *Open Definition 2.1 - Open Definition - Defining Open in Open Data, Open Content and Open Knowledge*. Disponibile su: <http://opendefinition.org/od/2.1/en/> [Data accesso: 20 giugno 2017]

Open Knowledge International Blog, *Why the Open Definition Matters for Open Data: Quality, Compatibility and Simplicity*. Disponibile su: <https://blog.okfn.org/2014/09/30/why-the-open-definition-matters-for-open-data-quality-compatibility-and-simplicity/> [Data accesso: 20 giugno 2017]

Qallme.fbk.eu, *Quall-me ontology*. Disponibile su: <http://qallme.fbk.eu/index.php?location=ontology> [Data accesso: 1 luglio 2017]

W3.org, *W3C Semantic Web FAQ*. Disponibile su: <https://www.w3.org/2001/sw/SW-FAQ#WhatIsTheSW> [Data accesso: 22 giugno 2017]

Wordnet.princeton.edu, *About WordNet - WordNet - About WordNet*. Disponibile su: <https://wordnet.princeton.edu/> [Data accesso: 1 luglio 2017]