

ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

---

**SCUOLA DI SCIENZE**  
Corso di Laurea in Informatica per il Management

**TEORIA E PRATICA DEGLI OPEN DATA:  
RASSEGNA ED IMPLEMENTAZIONE  
IN UN PORTALE WEB TURISTICO**

**Relatore:**  
Chiar.mo Prof.  
MARCO DI FELICE

**Presentata da:**  
LORENZO BIAGIO  
LANZARONE

**Correlatore:**  
MASSIMO SELVA

**Sessione II**  
**Anno Accademico 2017/2018**



*Alla mia famiglia ...*



# Indice

<b>Abstract</b>	<b>5</b>
<b>Introduzione</b>	<b>7</b>
<b>Parte I</b>	
<b>Rassegna sugli open data</b>	<b>11</b>
<b>1 L'importanza dei dati</b>	<b>13</b>
1.1 Dall'ipertesto al web . . . . .	14
1.2 L'evoluzione del web . . . . .	17
1.2.1 Il web 1.0 statico . . . . .	19
1.2.2 Il web 2.0 dinamico . . . . .	20
1.2.3 Il web 3.0 semantico . . . . .	21
1.3 L'invasione dei dati . . . . .	22
<b>2 Panoramica sugli open data</b>	<b>27</b>
2.1 Cosa sono gli open data . . . . .	27
2.1.1 Principi e caratteristiche degli open data . . . . .	28
2.1.2 Classificazione degli open data . . . . .	31
2.2 Le origini degli open data . . . . .	32
2.2.1 Il concetto di open government . . . . .	34
2.3 Limitazioni degli open data . . . . .	35
2.4 Le licenze aperte . . . . .	37
2.5 Tipologie e formati degli open data . . . . .	39

2.6	Come si utilizzano gli open data . . . . .	42
2.7	Principali fonti di open data . . . . .	44
2.7.1	OpenStreetMap . . . . .	45
2.7.2	Istat . . . . .	48
2.7.3	Dati.gov.it . . . . .	50
2.7.4	DatiOpen.it . . . . .	52
2.7.5	Comune di Bologna . . . . .	54
2.7.6	Arpa . . . . .	55
2.7.7	DBpedia . . . . .	57
2.7.8	Altre fonti . . . . .	59
2.8	Statistiche e valore degli open data . . . . .	60
2.9	Esempi di applicazione degli open data . . . . .	64

## Parte II

### Progettazione ed implementazione di open data in un portale web turistico **69**

<b>3</b>	<b>Progettazione</b>	<b>71</b>
3.1	Panoramica su OkBed&Breakfast . . . . .	71
3.2	Architettura e specifiche del progetto . . . . .	74
3.2.1	Pagine dei bed and breakfast . . . . .	76
3.2.2	Pagine di comuni, province e luoghi . . . . .	78
3.3	Open data utilizzati . . . . .	79
3.4	Scelte progettuali . . . . .	81
3.4.1	Download o API: vantaggi e svantaggi . . . . .	85
3.5	Screenshot . . . . .	86
<b>4</b>	<b>Implementazione</b>	<b>93</b>
4.1	Tecnologie utilizzate . . . . .	93
4.1.1	Drupal . . . . .	93
4.1.2	HTML . . . . .	94
4.1.3	CSS . . . . .	95

---

4.1.4	JavaScript . . . . .	96
4.1.5	jQuery . . . . .	96
4.1.6	AJAX . . . . .	97
4.1.7	PHP . . . . .	98
4.1.8	MySQL . . . . .	98
4.1.9	SPARQL . . . . .	99
4.1.10	Bootstrap . . . . .	100
4.1.11	OpenLayers . . . . .	101
4.1.12	AnimatedCluster . . . . .	102
4.1.13	Overpass API . . . . .	102
4.1.14	Nominatim API . . . . .	105
4.2	Codice implementato e file . . . . .	106
4.2.1	node-article.tpl.php . . . . .	106
4.2.2	taxonomy-term-location_istat.tpl.php . . . . .	110
4.2.3	taxonomy-term-luoghi.tpl.php . . . . .	112
4.2.4	markerOpenLayers.js . . . . .	113
4.2.5	page-ajax.tpl.php . . . . .	119
4.2.6	okbedbreakfast.module . . . . .	123
4.3	Errori e soluzioni . . . . .	125
<b>Conclusioni ed estensioni future</b>		<b>129</b>
<b>Elenco delle figure</b>		<b>133</b>
<b>Elenco delle tabelle</b>		<b>137</b>
<b>Bibliografia e sitografia</b>		<b>139</b>
<b>Ringraziamenti</b>		<b>147</b>



# Abstract

I dati sono sempre più importanti nella società e rappresentano una fondamentale risorsa. Da sempre l'essere umano ha la necessità di organizzare le informazioni e utilizzarle per creare nuova conoscenza. Nella presente tesi si approfondisce la tematica degli *open data*, cioè dati messi a disposizione da enti di vario genere, per essere utilizzati liberamente da chiunque per qualsiasi scopo.

Nella prima parte dell'elaborato si approfondiscono origini e aspetti teorici degli open data, mentre nella seconda gli aspetti pratici. Quanto descritto in quest'ultima parte è stato svolto durante un tirocinio presso l'agenzia web *Magic* di Funo di Argelato (Bologna). Il progetto riguarda la realizzazione di un esempio pratico di applicazione degli open data, in particolare la progettazione e implementazione degli stessi all'interno del portale web turistico *OkBed&Breakfast*.



# Introduzione

Al giorno d'oggi i dati sono sempre più importanti nella società e rappresentano una fondamentale risorsa. Da sempre l'essere umano ha la necessità naturale di organizzare le informazioni, accedervi, consultarle e utilizzarle per creare nuova conoscenza condivisa.

Alla base di questo studio vi è la volontà di approfondire la tematica degli *open data*. Si tratta di dati di vario tipo messi a disposizione da pubblica amministrazione, aziende e organizzazioni, per essere utilizzati nella maniera più libera possibile da chiunque e per qualsiasi scopo. L'importanza degli open data spesso era sottovalutata, solo negli ultimi anni è diventato un argomento in forte espansione e si è sviluppata una crescente attenzione verso questa importante risorsa, anche a livello legislativo.

La presente tesi è divisa in due parti distinte ma complementari, nella prima parte si approfondiscono origini e aspetti teorici degli open data, mentre nella seconda gli aspetti pratici. Parte del lavoro descritto nell'elaborato, specialmente quanto riportato nella seconda parte, è stato svolto durante un tirocinio della durata di due mesi presso l'agenzia web *Magic* di Funo di Argelato (Bologna), tra aprile e giugno 2018. Il progetto ha riguardato la realizzazione di un esempio pratico e reale di applicazione degli open data, in particolare la progettazione e implementazione degli stessi all'interno del sito web *OkBed&Breakfast*, un portale che mette in contatto viaggiatori e proprietari di bed and breakfast.

La tesi è strutturata in quattro capitoli, i primi due sono inclusi nella prima parte e gli ultimi due nella seconda parte.

- Nel **primo capitolo** si sottolinea l'importanza che i dati hanno avuto sia in passato che nella società attuale, ripercorrendo la nascita e l'evoluzione dell'ipertesto e del web, fino ad arrivare al progetto di Tim Berners-Lee relativo ai linked open data, la massima espressione del concetto stesso di open data, nonché pilastro fondamentale del web 3.0.
- Nel **secondo capitolo** si introduce e si approfondisce la tematica degli open data, definendone principi, caratteristiche e classificazione. Si ripercorre la nascita e l'evoluzione degli open data, affrontando anche il concetto di *open government*. Successivamente è affrontato il tema delle licenze aperte, con le quali sono pubblicati gli open data, oltre a descrivere limiti, tipologie e formati degli stessi. L'ultima parte del secondo capitolo descrive come si utilizzano gli open data e si offre una panoramica delle principali fonti di vario genere. Inoltre sono riportate alcune statistiche per comprendere meglio il fenomeno degli open data e il loro valore economico e sociale. Infine sono raccolti degli esempi di applicazione di questo tipo di dati in progetti software come siti web e app.
- Nel **terzo capitolo** si descrive la fase progettuale del lavoro svolto durante il tirocinio, nel quale sono stati implementati gli open data di *OpenStreetMap* e di *Istat* nel portale web turistico OkBed&Breakfast. Innanzitutto si offre una panoramica di quest'ultimo e del suo funzionamento, in seguito si descrivono architettura e specifiche del progetto realizzato. Successivamente si approfondisce quali open data sono stati utilizzati e in che modo, oltre a raccogliere le principali scelte progettuali compiute in questa fase. Infine sono presentati alcuni *screenshot* che mostrano le porzioni di pagine web realizzate e il funzionamento del progetto dal lato dell'utente.
- Nel **quarto capitolo** si descrive la fase implementativa del lavoro svolto durante il tirocinio. Nella prima parte del capitolo è riportata una panoramica delle tecnologie utilizzate durante questa fase, descrivendone aspetti tecnici, caratteristiche principali e in che modo si sono usate nel progetto. Inoltre si approfondiscono aspetti pratici relativi agli open data geospaziali di OpenStreetMap e alle API utilizzate. Successivamente sono riportati gli *snippet* più rilevanti del codice implementato.

Tali porzioni di codice sono suddivise in base al file a cui appartengono e si riportano anche esempi di open data nei formati JSON e GeoJSON ricevuti come risposta alle interrogazioni. Infine si descrivono le problematiche riscontrate durante la fase implementativa e le loro eventuali soluzioni.



# Parte I

## Rassegna sugli open data

*"Il futuro è ciò che costruiamo."*

Tim Berners-Lee



# Capitolo 1

## L'importanza dei dati

L'obiettivo di questo capitolo è di sottolineare l'importanza che i dati hanno avuto sia in passato che in misura sempre maggiore nella società attuale, nella quale sono utilizzati tramite le moderne tecnologie informatiche e di comunicazione. I dati permettono di memorizzare informazioni e generare nuovo sapere, con il quale è possibile compiere progressi e sviluppare innovazioni. Per comprendere appieno il concetto di *open data* è importante capire prima quello di *dato* e analizzare l'evoluzione tecnologica che ha portato alla nascita del *web*, entrambi argomenti strettamente correlati agli open data.

Il valore dei dati e delle informazioni oggi è nettamente superiore a qualsiasi altra forma di contributo. I dati sono una risorsa preziosa, che si presta a infinite possibili applicazioni, tramite l'integrazione e la correlazione tra fonti informative diverse. Dalla ricerca scientifica alla realizzazione di progetti in ambito aziendale, si tratta di una risorsa ancora in gran parte inutilizzata e dalle enormi potenzialità. A partire dai dati si giunge alla consapevolezza, attraverso un percorso preciso:

**dati → informazioni → conoscenza → consapevolezza**

I dati, di per sé neutri, diventano informazioni quando vengono creati, estratti, elaborati e utilizzati per determinati obiettivi. Le informazioni, spesso costituite da più dati dello stesso tipo o di tipo diverso, diventano conoscenza quando vengono interpretate attraverso determinati strumenti, applicazioni e metodologie. Infine, la conoscenza acquisisce il valore di consapevolezza quando ad essa viene attribuita la funzione di cambiamento e

miglioramento della realtà interpretata grazie ai dati [1]. Fin dall'antichità l'essere umano ha la necessità di collezionare informazioni, categorizzarle, organizzarle in complesse strutture fisiche e mentali, ed infine indicizzarle per accedervi più facilmente. L'informatizzazione degli ultimi anni, con lo sviluppo dei nuovi mezzi tecnologici e di comunicazione, ha permesso di gestire, cercare e diffondere in un modo prima inimmaginabile i dati e di conseguenza il sapere umano.

## 1.1 Dall'ipertesto al web

Il bisogno dell'uomo di organizzare dati e informazioni ha portato alla nascita dell'ipertesto, cioè un'insieme di documenti che sono in relazione tra loro, tramite l'utilizzo di parole chiave. L'ipertesto può quindi essere immaginato come una rete, nella quale ogni nodo rappresenta un documento. La particolarità di questa organizzazione è che la sua lettura non è necessariamente lineare, ogni nodo può essere il punto di partenza o il punto di arrivo in base a ciò che si sta cercando.

Una forma primitiva di ipertesto si può trovare nell'antichità, ad esempio commenti alla Bibbia e ad altri importanti libri. Nel 1945 l'ingegnere statunitense Vannevar Bush, scrisse un articolo intitolato *As We May Think*, nel quale descriveva un sistema informativo analogico chiamato *Memex*, proponendo le basi concettuali dell'ipertesto. Tale idea influenzò e ispirò i successivi studi sulla materia. Il sociologo e filosofo statunitense Ted Nelson coniò nel 1963 il termine *hypertext*, in italiano *ipertesto*. Lo descriveva come un sistema elastico per organizzare informazioni, i dati non potevano essere stampati su carta, ma era necessario un computer per interagire con essi.

Il concetto di ipertesto raggiunse la sua massima espressione con la nascita del web. La prima rete di computer fu *Arpanet*, nata nel 1969, che era usata inizialmente per scopi militari e successivamente in ambito universitario. Nei primi anni ottanta dall'evoluzione di Arpanet nacque la rete *Internet* sulla quale si svilupperà il web, uno dei suoi principali servizi nonché interfaccia attraverso la quale è possibile consultare l'ipertesto globale. Prima della nascita e della diffusione del web, il concetto di ipertesto informatico

era confinato a centri di ricerca e laboratori universitari, dove vennero proposte diverse applicazioni che però portarono raramente a sistemi utilizzabili. Solo a partire dagli anni ottanta nacque un interesse commerciale verso l'ipertesto, anche grazie alle evoluzioni tecniche dei computer, così diverse aziende svilupparono appositi software in grado di costruire ipertesti.

Il software di questo tipo che ebbe il maggior successo fu *HyperCard*, sviluppato nel 1987 dall'informatico statunitense Bill Atkinson per Apple. Si trattava di un'applicazione per Mac che permetteva agli utenti di realizzare in modo intuitivo ipertesti con contenuti anche multimediali. Inoltre conteneva il linguaggio di programmazione *HyperTalk*, anch'esso semplice da utilizzare, ma in grado di gestire dati e interfaccia grafica. Nonostante Apple con HyperCard abbia sbagliato strategia di marketing, oltre a non aggiornarlo per molto tempo, questo software è rimasto nella storia per aver contribuito in modo significativo alla diffusione del modello di ipertesto, ispirando anche altre applicazioni dello stesso tipo. Rimase in commercio fino al 2004, quando Apple decise di cessarne lo sviluppo, non essendo più compatibile con la nuova versione del sistema operativo macOS e ormai superato dall'avvento del web.

Come accennato precedentemente, l'ipertesto informatico raggiunse la sua massima espressione con l'invenzione del web. L'informatico britannico Tim Berners-Lee, che lavorava al CERN<sup>1</sup> di Ginevra in Svizzera, scrisse nel 1989 un documento in cui si proponeva di creare un *ampio database ipertestuale con link* per risolvere un grande problema. Gli scienziati del CERN infatti faticavano molto per trovare dati e informazioni per i loro studi, poiché archiviati in strutture e calcolatori diversi che non potevano comunicare facilmente tra loro. Vi era anche un problema di incompatibilità tra computer con sistemi operativi e software differenti, oltre alla mancanza di uno standard comune che permettesse di accedere ai dati. Era però difficile per gli addetti ai lavori immaginare un sistema ipertestuale globale accessibile facilmente da chiunque come proposto da Berners-Lee. Si trattava di un'idea astratta ma allo stesso tempo innovativa, che dopo pochi anni avreb-

---

<sup>1</sup>CERN è l'organizzazione europea per la ricerca nucleare, nonché il più grande laboratorio al mondo di fisica delle particelle.

be cambiato il mondo, tuttavia inizialmente Berners-Lee non trovò il supporto sperato. Utilizzando un computer NeXT, egli realizzò gli elementi fondamentali del web: il protocollo HTTP, il primo server web, il linguaggio HTML<sup>2</sup> e il meccanismo degli URL<sup>3</sup>, con il quale ogni risorsa poteva essere raggiunta. Per funzionare, il web sfrutta il protocollo HTTP, acronimo di *Hypertext Transfer Protocol*. Si tratta di un protocollo di rete generico e *stateless*, appartenente al livello di applicazione del modello ISO/OSI, usato come principale sistema per la trasmissione dei dati tramite un'architettura client-server. Il primo esegue una richiesta e il secondo fornisce una risposta, scambiandosi specifici messaggi. I messaggi di richiesta sono composti da: riga di richiesta, sezione header, riga vuota e body. La riga di richiesta è composta da: metodo (come GET, POST e HEAD), URI e versione del protocollo. Invece i messaggi di risposta sono composti da: riga di stato, sezione header, riga vuota e body [2]. La prima pagina web venne pubblicata alla fine del 1990, Berners-Lee aveva creato quello che oggi è il World Wide Web, il più ampio ipertesto informatico del mondo [3].



Figura 1.1: Tim Berners-Lee

---

<sup>2</sup>HTML (Hypertext Markup Language) è un linguaggio di markup nato per realizzare documenti ipertestuali disponibili nel web.

<sup>3</sup>URL (Uniform Resource Locator) è una sequenza di caratteri che identifica univocamente l'indirizzo di una risorsa presente su un server web.

## 1.2 L'evoluzione del web

Inizialmente il web era utilizzato da centri di ricerca e in ambito accademico. Solo nel 1993 il CERN permise a chiunque di accedere gratuitamente e liberamente a questo innovativo mezzo di comunicazione. Si trattava di uno strumento dalle infinite potenzialità, non utile soltanto ai ricercatori, ma all'intera popolazione, che così poteva accedere facilmente a dati e informazioni. Il numero dei server connessi e delle pagine web disponibili iniziò a crescere in modo esponenziale, creando un ipertesto informatico accessibile da ogni parte del mondo.



Figura 1.2: Icone dei principali browser web

All'inizio degli anni novanta furono sviluppati da aziende e centri di ricerca i primi browser web, cioè i software che permettevano all'utente di collegarsi alla rete internet e navigare tra le pagine disponibili online. Dopo *Mosaic* (1993) e *Netscape Navigator* (1994), nel 1995 Microsoft decise di entrare prepotentemente nel mercato dei browser, lanciando *Internet Explorer* e installandolo di serie nel sistema operativo Windows 95. Questo nuovo ingresso nel mercato portò ad una serie di cause legali e alla cosiddetta *prima guerra dei browser*, finché nel 1999 Internet Explorer divenne il browser web più utilizzato. Negli anni seguenti nacquero altri importanti browser, *Firefox* di Mozilla (2002), *Safari* di Apple (2003), *Opera* (gratuito dal 2005) e *Chrome* di Google (2008). L'arrivo di questi nuovi software portò alla *seconda guerra dei browser*, Chrome divenne il più utilizzato (occupando attualmente circa il 60% del mercato [4]) e Microsoft fu costretta a sostituire Internet Explorer con *Edge*, arrivato nel 2015. Negli ultimi anni i browser si sono uniformati con funzioni e idee comuni. Questo grazie al W3C, acronimo di *World Wide Web Consortium*, nato nel 1994 su iniziativa di Tim Berners-Lee per promuovere e uniformare lo sviluppo delle tecnologie del WWW.

A causa dell'espansione del web, le pagine erano sempre più numerose e diventò un problema cercarle e raggiungerle tramite i browser. L'utente infatti doveva conoscere l'URL del contenuto che voleva consultare, per risolvere questo problema arrivarono i primi motori di ricerca, che indicizzavano tutte le pagine presenti online nei loro database. Questi potenti strumenti consentirono per la prima volta agli utenti di cercare dati e informazioni in modo semplice digitando specifiche parole chiave e restituendo i risultati più attinenti. Il primo software definibile come motore di ricerca fu *Archie*, sviluppato nel 1990 da Alan Emtage, studente dell'Università McGill di Montréal. Si trattava di un motore di ricerca per archivi di documenti distribuiti con il protocollo FTP<sup>4</sup>, ebbe il merito di attirare l'interesse di molti studiosi e dare il via allo sviluppo di questo tipo di software.

Negli anni novanta nacquero molti motori di ricerca e si assistette ad un processo di costante perfezionamento del loro funzionamento e dei loro algoritmi, rendendoli sempre più precisi ed intelligenti. Per esempio nel 1994 nacque *Yahoo!*, nel quale inizialmente i risultati erano indicizzati da esseri umani e non da *crawler*<sup>5</sup>. Nel 1995 nacque *AltaVista*, il primo in grado di comprendere *query* scritte in linguaggio naturale. Nel 1996, due studenti dell'Università di Stanford, Larry Page e Sergey Brin, iniziarono a lavorare su un nuovo motore di ricerca che chiamarono *BackRub*. Questo software introduceva il *page rank*, un innovativo algoritmo per indicizzare i siti web e in grado di mostrare i contenuti più pertinenti in cima all'elenco dei risultati<sup>6</sup>. Durante l'indicizzazione venivano premiate le pagine che erano citate maggiormente da altre pagine, in questo modo ogni contenuto possedeva una reputazione che ne influenzava il posizionamento. Nel 1998, sulle basi di *BackRub*, nasce *Google*, che diventerà il principale motore di ricerca del mondo. Page e Brin scelsero il nome *Google* ispirandosi al termine *Googol*, che rappresenta il numero uno seguito da cento zeri, per indicare l'enorme quantità di dati e informazioni che è in grado di gestire.

---

<sup>4</sup>FTP (File Transfer Protocol) è un protocollo per la trasmissione di dati tra host basato su TCP e con architettura client-server.

<sup>5</sup>Un crawler è un software che analizza i contenuti di una rete o di un database in un modo metodico e automatizzato, in genere per conto di un motore di ricerca.

<sup>6</sup>Dal quale in seguito nascerà il SEO (Search Engine Optimization), che definisce tutte le attività di ottimizzazione di un sito web volte a migliorarne il posizionamento nei risultati dei motori di ricerca.

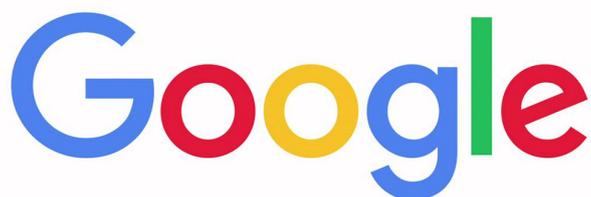


Figura 1.3: Logo di Google

Con il passare degli anni l'algoritmo di Google è stato costantemente ottimizzato, diventando il più preciso e di conseguenza il più utilizzato, seguito da Yahoo! e da *Bing*, il motore di ricerca di Microsoft lanciato nel 2009. Aspetti determinanti nel successo di Google sono stati la sua velocità di risposta alle ricerche degli utenti e l'interfaccia grafica semplice e dall'ottima usabilità. Oggi Google è leader del mercato, ha allargato i suoi interessi anche ad altri settori, ed è una delle aziende informatiche più importanti del mondo, controllata dal 2015 dalla holding *Alphabet* [5].

Da quando è nato ad oggi, il web ha subito un'evoluzione, nel corso degli anni le tecnologie legate ad esso sono migliorate costantemente e lo hanno reso sempre più coinvolgente e utile per gestire dati e informazioni. È possibile individuare tre fasi principali che hanno caratterizzato l'evoluzione del web: *web 1.0*, *web 2.0* e *web 3.0*.

### 1.2.1 Il web 1.0 statico

Il web 1.0 indica la prima fase di vita del World Wide Web, la parola d'ordine di questa versione è *staticità*. Negli anni novanta l'utente visitava i siti web senza poter interagire con essi, quindi senza poterne modificare il contenuto. Le pagine erano semplici file HTML e CSS<sup>7</sup> statici, che non venivano elaborati dal server prima della trasmissione o dal client dopo la ricezione. All'epoca quindi, la creazione di un sito web non includeva una vera e propria fase di programmazione, ma si limitava alla progettazione e visualizzazione di contenuti statici. La manutenzione di un sito web statico risulta più scomoda, poiché

---

<sup>7</sup>CSS (Cascading Style Sheets) è un linguaggio usato per definire lo stile di un documento HTML e quindi l'interfaccia grafica del sito web.

anche per una piccola modifica ad una pagina è necessario agire direttamente sul codice HTML, ma contemporaneamente è anche il paradigma più semplice per implementare un sito web. Ancora oggi i siti web statici sono utilizzati, specialmente per pubblicazioni di dimensioni contenute e soprattutto a scopo informativo, come siti web personali e aziendali.

### 1.2.2 Il web 2.0 dinamico

Il web 2.0 è l'evoluzione della versione 1.0. In questa seconda fase di vita del web la parola d'ordine è *dinamicità*. Tra la fine degli anni novanta e l'inizio degli anni duemila il web subisce un'importante evoluzione: grazie alle nuove tecnologie introdotte l'utente può interagire e modificare i contenuti, diventando un soggetto attivo nella rete, oltre ad essere gli stessi contenuti a modificarsi in base all'utente che li visualizza.

Il termine *web 2.0* è stato utilizzato per la prima volta nel 2004 dall'editore irlandese Tim O'Reilly. Questa seconda versione è caratterizzata dall'arrivo di nuovi linguaggi di programmazione lato server, come PHP, e lato client, come JavaScript, che hanno permesso di rendere dinamiche e interattive le pagine create. Tali tecnologie sono state essenziali per la nascita e la diffusione di numerose applicazioni e servizi web, realizzate anche tramite l'uso di CMS<sup>8</sup> e database, che consentirono per la prima volta una migliore gestione di dati e contenuti complessi.

Si assistette alla nascita di blog, forum, wiki, siti web di e-commerce, piattaforme multimediali e di streaming, chat e social network, nei quali l'utente è il protagonista e può contribuire in prima persona, partecipando anche a *community* di utenti che hanno gli stessi interessi per scambiarsi informazioni e opinioni. Proprio in questa fase nacque *Wikipedia*, la più vasta enciclopedia online del mondo, gestita e aggiornata dagli utenti stessi. Il web si è arricchito di contenuti di ogni genere ed è migliorata notevolmente l'interfaccia grafica e l'usabilità delle pagine, anche per i dispositivi mobili tramite il *responsive design*. Tutto questo ha reso la rete internet il mezzo di comunicazione più

---

<sup>8</sup>CMS (Content Management System) è uno strumento software, installato su un server web, il cui compito è facilitare la gestione dei contenuti del sito web.

coinvolgente, completo ed intrinseco nella vita degli utenti e nella società. Inoltre si è iniziato ad assistere ad una progressiva fusione di tutti gli altri mezzi di comunicazione con il web, poiché utilizzabili direttamente tramite la rete internet.

### 1.2.3 Il web 3.0 semantico

Il web 3.0 indica la fase attuale nella quale si colloca lo sviluppo del web. Anche in questo caso si tratta di un'evoluzione delle tecnologie e dei concetti nati durante le fasi precedenti. La parola che caratterizza il web 3.0 è *semanticità*, cioè la capacità di dare un significato e un contesto ai dati online, per arrivare al cosiddetto *web of data*.



Figura 1.4: Rappresentazione astratta della rete internet mondiale

L'idea alla base del web 3.0, ancora in fase di affermazione, è di creare un web che sia semantico, con la tendenza ad agglomerare le informazioni in un unico database. Queste operazioni permetterebbero lo sviluppo dell'intelligenza artificiale, che potrebbe così utilizzare la sterminata quantità di dati disponibili online. Per attribuire significato alle informazioni presenti sul web sono stati progettati appositi linguaggi di markup e di interrogazione, come RDF, OWL e SPARQL, proposti direttamente dal W3C e pensati per formalizzare i contenuti in modo che possano essere interpretabili anche dalle macchine. Si iniziò a parlare di web 3.0 nel 2006, anno in cui Tim Berners-Lee dichiarava [6]:

*“Le persone si continuano a chiedere cosa sia il web 3.0. Penso che forse, quando si sarà ottenuta una sovrapposizione della grafica vettoriale scalabile sul web 2.0, e l’accesso ad un web semantico integrato in uno spazio enorme di dati, si potrà ottenere l’accesso ad un’incredibile risorsa di informazioni.”*

Il web 3.0 può essere immaginato come la possibilità di rendere internet intelligente, in modo che dati e informazioni connesse tra loro siano interpretabili semanticamente da dispositivi elettronici. Le colonne portanti del web semantico sono i *linked data*, presentati da Tim Berners-Lee durante la conferenza *TED* nel 2009. Si tratta di dati pubblicati online che rappresentano risorse appartenenti allo stesso dominio o a domini differenti, collegate tra loro e utilizzabili tramite interrogazioni semantiche. Inoltre il W3C, tramite il progetto *Linking Open Data* promuove lo sviluppo dei *linked open data*, cioè la forma aperta e libera dei linked data [7]. I linked open data sono in continua espansione e rappresentano la maggiore espressione di apertura, accessibilità e condivisione del concetto stesso di open data, che verrà analizzato nel secondo capitolo. I dati, se considerati singolarmente, possono avere poco valore, ma se più dataset provenienti da fonti diverse vengono analizzati attraverso tecniche di *data mining* e messi in correlazione tra loro, portano alla creazione di nuova conoscenza condivisa. È proprio questo lo scopo dei linked data e di conseguenza del web 3.0: connettere le informazioni per rendere il web intelligente e utilizzarlo come generatore di nuova conoscenza [8] [9].

### 1.3 L’invasione dei dati

Oggi, il mondo in cui viviamo è invaso da dati e informazioni che vanno gestite, organizzate e riutilizzate, essendo una fonte di inestimabile valore. L’evoluzione del web, ha avuto un ruolo di primo piano in questa sempre crescente necessità naturale dell’essere umano di gestire i dati. Ovviamente il web è in continua evoluzione e non si fermerà nella terza fase, anzi, si sta già iniziando a sviluppare il futuro web 4.0. Sarà il web *simbiotico*, in grado di dialogare direttamente con il cervello umano e funzionare come un sistema operativo, di conseguenza *intelligenza artificiale* e *realtà aumentata* saranno sempre più protagoniste [10].

Fattori che hanno accompagnato lo sviluppo del web da quando è nato ad oggi, e di conseguenza hanno portato all'uso sempre più massiccio di dati, sono stati il miglioramento della velocità di connessione alla rete internet e lo sviluppo di nuovi dispositivi tecnologici. Il dispositivo che ha dato veramente inizio a questo cambiamento è stato l'*iPhone* di Apple. Presentato nel 2007 da Steve Jobs, è uno strumento mobile dalle infinite potenzialità, che inaugurò il settore degli *smartphone*, e portò allo sviluppo di tutti i prodotti derivati, come i *tablet*. Questi nuovi dispositivi, essendo costantemente connessi alla rete internet, hanno permesso una diffusione ancora più capillare del web, utilizzandolo ed interagendo con esso anche tramite le *app*. Dati di ogni genere rappresentano il centro del loro utilizzo ed è necessario proteggerli poiché si tratta di una preziosa risorsa.

Dall'utilizzo dei social network, come *Facebook*, *Twitter*, *Google+*, *Instagram* e *YouTube*, ai sensori elettronici integrati negli smartphone che generano costantemente dati, passando per la diffusione di dispositivi indossabili come gli *smartwatch*, fino alle più recenti tecnologie di domotica e del cosiddetto *internet of things*, siamo costantemente connessi alla rete e di conseguenza completamente circondati da dati di ogni genere [11]. L'intera società, negli ultimi anni, ha la naturale tendenza a trasferire online ogni aspetto della sua gestione, dove dati e informazioni sono connesse, accessibili da ogni luogo e in qualsiasi momento. Anche l'archiviazione dei dati e la gestione dei database sono cambiate con l'avvento di internet, dalla carta ai supporti meccanici e analogici fino ad arrivare all'informatizzazione e ai supporti digitali. Oggi con il *cloud* si è compiuto un ulteriore passo in avanti, archiviando tutti i dati online e rendendoli sempre accessibili da ogni dispositivo [12]. Di seguito alcune statistiche per comprendere meglio il fenomeno appena descritto [13]:

- **Internet nel mondo:** 4.021 miliardi di utenti su una popolazione di 7.593 miliardi di persone (penetrazione del 53%). Usano internet da mobile 3.722 miliardi di persone (penetrazione del 49%). Usano i social network 3.196 miliardi di persone (penetrazione del 42%).

- **Internet in Italia:** 43.31 milioni di utenti su una popolazione di 59.33 milioni di persone (penetrazione del 73%). Usano internet da mobile 38.81 milioni di persone (penetrazione del 65%). Usano i social network 34.00 milioni di persone (penetrazione del 57%).

Da questa analisi si può comprendere l'importanza fondamentale che hanno i dati e le informazioni. Grazie ai nuovi mezzi tecnologici, informatici e di comunicazione essi possono diffondersi ed essere utilizzati in ogni circostanza, ma è importante ricordare che essendo un'enorme risorsa, i dati vanno anche mantenuti e protetti. Oggi, e in misura sempre maggiore in futuro, avremo a disposizione una quantità e qualità di conoscenza talmente vasta che è difficilmente quantificabile. Mai in passato l'essere umano ha avuto un accesso così immediato e libero al sapere, di conseguenza è necessario imparare a utilizzarlo nel migliore dei modi.





# Capitolo 2

## Panoramica sugli open data

Come spiegato nel primo capitolo, i dati sono estremamente importanti per l'evoluzione della società e per la creazione di nuova conoscenza. Di conseguenza, più i dati possono essere utilizzati liberamente, maggiori saranno i benefici ottenibili dall'utilizzo degli stessi. Proprio per soddisfare questi bisogni di accessibilità alle informazioni nascono gli *open data*, in italiano *dati aperti*. L'obiettivo di questo capitolo è di offrire una panoramica sul concetto di open data, analizzando gli aspetti caratterizzanti e i principi alla base di questa particolare forma di dati.

### 2.1 Cosa sono gli open data

Gli open data sono dati pubblici resi disponibili in varie forme dalla pubblica amministrazione o da altri enti in modo libero e gratuito per tutti i cittadini. Gli open data possono riguardare qualsiasi ambito della società che ci circonda, ad esempio dati relativi a popolazione, ambiente, economia, trasporti, politica e molto altro. Questi dati possono essere utilizzati da chiunque, ad esempio cittadini e aziende, per consultarli e diffondere nuova conoscenza o per sviluppare software che combinano tra loro queste informazioni.

L'idea alla base degli open data richiama il concetto di *open government*, cioè un esercizio del potere che consenta alle amministrazioni di essere aperte e trasparenti nei confronti dei cittadini. Utilizzando quindi i nuovi mezzi tecnologici, informatici e di comunica-

zione per divulgare informazioni che però non compromettano la *privacy* dei cittadini stessi, ma anzi creino nuovo valore proprio per questi ultimi. Per comprendere meglio la differenza tra varie tipologie di dati, aperti e non, è importante soffermarsi sulle seguenti definizioni:

- **Dato:** i dati sono strettamente correlati a fatti, sono potenzialmente riproducibili e possono fare parte di strutture informative più vaste. In informatica, un dato è un valore, rappresentato in bit, che può essere elaborato da un calcolatore elettronico. Le memorie di massa consentono di archiviare i dati in modo permanente. Singoli dati o insiemi di dati possono essere conservati in file o database. Un dato non aperto può essere soggetto a limiti di utilizzo o essere a pagamento [1].
- **Big data:** il termine big data indica un volume enorme di dati, strutturati e non strutturati, impossibili da analizzare senza l'utilizzo di software e calcolatori elettronici. Si tratta di dati che provengono da fonti diverse, come social network, transazioni economiche, sistemi di analisi e sensori elettronici [14].
- **Open data:** come specificato dall'*Open Definition*, un dato si definisce aperto quando chiunque ha libertà di accesso, uso, modifica e condivisione dello stesso. Un dato aperto può essere soggetto al massimo a misure che ne preservino la provenienza e l'apertura. Di conseguenza l'utilizzatore può dover citare la fonte o pubblicare il dato prodotto con lo stesso tipo di licenza con cui è stato rilasciato originariamente [15].
- **Linked open data:** come descritto nel primo capitolo, i linked open data rappresentano la massima espressione del concetto stesso di open data, applicato ai linked data. Si tratta quindi di dati aperti e strutturati, collegati fra di loro e utilizzabili attraverso interrogazioni semantiche [16].

### 2.1.1 Principi e caratteristiche degli open data

L'*Open Knowledge International*, fondazione no profit nata nel 2004 che promuove l'apertura dei dati, definisce la conoscenza aperta tramite i seguenti principi [17]:



Figura 2.1: Logo di Open Knowledge International

- **Disponibilità e accesso:** i dati devono essere disponibili nel loro complesso, per un prezzo non superiore ad un ragionevole costo di riproduzione, preferibilmente tramite il download gratuito via internet. Qualsiasi informazione aggiuntiva vincolata dalla licenza (come i nomi dei contributori richiesti nel caso del requisito di attribuzione) deve accompagnare l'opera. I dati devono essere disponibili in un formato utile e modificabile, che non ponga inutili ostacoli tecnologici allo svolgimento dei permessi forniti dalla licenza. Nello specifico, i dati devono essere processabili da un calcolatore, disponibili in massa e messi a disposizione in un formato aperto, vale a dire un formato le cui specifiche siano pubblicamente e liberamente disponibili e che non imponga nessuna restrizione economica o di altro tipo al suo utilizzo o che, come caso estremo, possa essere elaborato da almeno un software libero e gratuito.
- **Riutilizzo e redistribuzione:** i dati devono essere forniti a condizioni tali da permetterne il riutilizzo e la redistribuzione. Ciò comprende la possibilità di combinarli con altre basi di dati. L'opera deve essere disponibile con una licenza aperta, qualsiasi condizione aggiuntiva che accompagna l'opera (come condizioni d'uso o brevetti del licenziatario) non deve contraddire le condizioni della licenza.
- **Partecipazione universale:** tutti devono essere in grado di usare, riutilizzare e redistribuire i dati. Non ci devono essere discriminazioni né di ambito di iniziativa né contro soggetti o gruppi. Ad esempio, la clausola *non commerciale*, che vieta l'uso a fini commerciali o restringe l'utilizzo solo per determinati scopi (come quello educativo) non è ammessa.

Gli open data, come riportato sul documento *Ten Open Data Guidelines* pubblicato da *Transparency International Georgia*, devono avere le seguenti caratteristiche [18]:

- **Completi:** tutti i dati compresi in dataset pubblici dovrebbero essere pubblicati, inclusi i dati di archiviazione non digitali (che andrebbero convertiti) e quelli utilizzati per generare dati aggregati o derivati.
- **Primari:** i dati dovrebbero essere raccolti all'origine e dovrebbero essere pubblicati con il livello di granularità con cui sono stati raccolti.
- **Tempestivi:** i dati devono essere resi disponibili il più rapidamente possibile al fine di massimizzarne il valore per il pubblico, che deve essere in grado di trovarli velocemente tramite apposite funzioni di ricerca.
- **Accessibili:** i dati devono essere resi disponibili al maggior numero possibile di utenti e per la più ampia gamma possibile di scopi. I dati devono essere facili da condividere digitalmente e per il loro utilizzo non deve mai essere richiesta alcuna forma di registrazione o pagamento. I download collettivi devono essere resi disponibili tramite appositi protocolli come FTP, oppure tramite una API ben documentata per l'accesso automatizzato ai dati.
- **Leggibili da computer:** i dati devono essere archiviati utilizzando un formato e una struttura che consentano facilmente l'elaborazione automatizzata. La leggibilità da parte del computer non dovrebbe sostituire la leggibilità umana, quindi dovrebbero essere offerti entrambi i formati.
- **In formati non proprietari:** i dati devono essere codificati in formati aperti, sui quali nessuna entità abbia il controllo esclusivo. Sono preferibili i formati con le codifiche più semplici e maggiormente supportati.
- **Liberamente utilizzabili:** I dati devono essere resi gratuiti per tutti i tipi di utilizzo, incluso l'uso commerciale, senza restrizioni. Devono essere caratterizzati da licenze che non ne limitino l'uso, di conseguenza protezioni della proprietà intellettuale, come *copyright* o *trademark*, non possono essere utilizzate.

- **Rivedibili:** ogni ente pubblico o privato che rilascia dati dovrebbe designare una persona per rispondere a domande e reclami sui dati. Inoltre le informazioni di contatto di questa persona dovrebbero essere incluse con i dati.
- **Ricercabili:** i dati devono poter essere trovati da chi li sta cercando. Devono essere inclusi in appropriati cataloghi aggiornati, e i relativi siti web devono essere accessibili ai motori di ricerca.
- **Permanenti:** Poiché i dati invecchiano, dovrebbero essere archiviati in modi che soddisfino le caratteristiche sopra indicate durante il loro intero ciclo di vita.

### 2.1.2 Classificazione degli open data

Tim Berners-Lee, inventore del World Wide Web, ha proposto un modello per classificare il formato dei dati su una scala che va da una a cinque stelle. Una stella corrisponde al minimo di apertura, cinque stelle al massimo di apertura. Nella scala, un dato può considerarsi aperto se possiede almeno tre stelle [19].

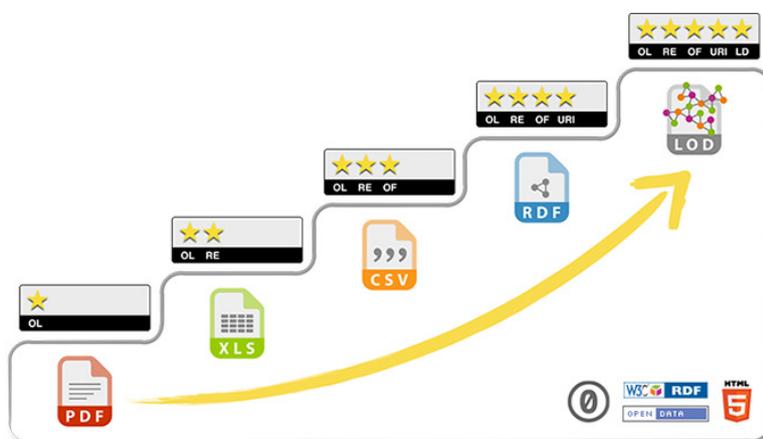


Figura 2.2: Le 5 stelle degli open data

- **1 stella:** è il livello più basso di apertura, indica dati non strutturati (cioè conservati senza uno schema preciso) e pubblicati sul web con una licenza aperta, ad esempio un file PDF o un'immagine JPG.

- **2 stelle:** indicano dati strutturati (cioè conservati con uno schema preciso) ma codificati in un formato proprietario, quindi possono essere elaborati efficacemente da un software in grado di leggerli. Questo tipo di dati, poiché strutturati, possono essere convertiti in open data. Ad esempio, un file XLS di Microsoft Excel.
- **3 stelle:** indicano un dato strutturato e codificato in un formato non proprietario, si tratta del primo livello utile a definire un dato come aperto. Ad esempio, un file CSV o XML che può essere letto da qualsiasi software non necessariamente proprietario.
- **4 stelle:** dati strutturati, codificati in formato non proprietario e dotati di un URI (Uniform Resource Identifier), cioè un identificativo univoco della risorsa. Un esempio è lo standard RDF (Resource Description Framework), che applica al dato un significato semantico condiviso e può essere elaborato da un software.
- **5 stelle:** indicano il massimo livello di apertura dei dati, rappresentato dai LOD (Linked Open Data). Si tratta di open data, collegati dinamicamente ad altri insiemi di dati aperti provenienti anche da fonti diverse.

## 2.2 Le origini degli open data

La filosofia alla base del concetto di dati aperti esiste da molto tempo e le sue origini risalgono all'illuminismo, sviluppatosi in Europa durante il XVIII secolo, ma il termine *open data* è molto più recente e ha assunto importanza grazie all'avvento di internet e del web. Nel 2003 l'Unione Europea emanò la direttiva sul *riutilizzo dell'informazione del settore pubblico*, per promuovere il più possibile la diffusione e il riutilizzo di dati e documenti in possesso delle pubbliche amministrazioni. Nel 2013 la direttiva è stata aggiornata rendendo obbligatoria la pubblicazione degli open data, ed estendendo tale disposizione anche alle istituzioni culturali. Purtroppo però, a seguito della direttiva emanata nel 2003 furono estremamente limitati i tentativi di attuazione da parte dei vari Stati europei. Il portale web dell'Unione Europea dedicato alla diffusione degli open data, *European Union Open Data Portal*, arrivò solo nel 2012.

Un'importante spinta all'utilizzo e alla creazione degli open data è stata data senza dubbio dagli Stati Uniti d'America. Nel 2009 l'ex presidente Barack Obama approvò la direttiva sull'*open government*, con il quale il suo governo si impegnavano ad attuare una politica di trasparenza senza precedenti verso i cittadini, per potenziare la democrazia. L'annuncio avvenne tramite un memorandum intitolato *Transparency and Open Government*, nel quale si affermava [20]:

*“Fino a dove è possibile e sottostando alle sole restrizioni valide, le agenzie devono pubblicare le informazioni online utilizzando un formato aperto che possa cioè essere recuperato, soggetto ad azioni di download, indicizzato e ricercato attraverso le applicazioni di ricerca web più comunemente utilizzate. Per formato open si intende un formato indipendente rispetto alla piattaforma, leggibile dall'elaboratore e reso disponibile al pubblico senza che sia impedito il riuso dell'informazione veicolata.”*

Sulle basi di questa direttiva nacque, sempre nel 2009, il portale web *Data.gov*, che raccoglie buona parte degli open data disponibili negli Stati Uniti d'America. Successivamente, proprio a *Data.gov* si ispireranno tutti gli altri portali dello stesso tipo nati negli altri Stati del mondo.

In Italia il concetto di open data è arrivato per la prima volta nel 2007, grazie alla diffusione del progetto *OpenStreetMap*, un sistema di creazione di mappe e cartografie a contenuto libero di tutto il mondo (in modo simile a *Google Maps*, che però non è open). Altro importante passo in avanti arrivò senza dubbio il 18 ottobre 2011, quando venne inaugurato il portale web *Dati.gov.it*, che raccoglie buona parte degli open data disponibili nel nostro paese. Sempre nello stesso anno venne pubblicato un vademecum dedicato agli open data, dal titolo *Come rendere aperti i dati delle pubbliche amministrazioni*, che spiegava quali erano i passi da compiere per pubblicare gli open data [21].

Tra il 2012 e il 2013, a seguito della direttiva europea descritta precedentemente, venne applicato anche in Italia l'obbligo per le pubbliche amministrazioni di pubblicare i dati in formato aperto, oltre al principio del cosiddetto *open by default*, secondo il quale tutti i dati diffusi online dalla pubblica amministrazione che fossero sprovvisti di un'esplicita

licenza di utilizzo sono da intendersi come open data, utilizzabili da chiunque anche per finalità commerciali. Successivamente, negli anni seguenti, alcune città e regioni italiane hanno realizzato i propri siti web dedicati alla raccolta e alla distribuzione degli open data locali. Oltre ai portali gestiti direttamente da città e regioni, dal 2012 esistono altre due importanti fonti di open data in Italia, che consentono di scaricare centinaia di dataset<sup>1</sup> liberamente utilizzabili: *DatiOpen.it* e *Open Data Hub Italia*. Anche l'Istat dispone di due portali per utilizzare gli open data realizzati dall'Istituto nazionale di statistica: *Linked Open Data* e *I.Stat*. La maggiore community italiana dedicata alla discussione e allo scambio di idee e informazioni relativamente agli open data è *Spaghetti Open Data*.

Nel 2013, durante il vertice del G8 in Irlanda del Nord, i principali leader mondiali firmarono la *Open Data Charter*, cioè la *Carta dei Dati Aperti*. Con questo accordo si definirono i principi per applicare il concetto di open by default al proprio patrimonio informativo, e per aumentare quantità, qualità e facilità di riutilizzo dei dati stessi. Nel 2015 venne fatto un ulteriore passo in avanti, quando i governi di 17 paesi firmarono la *International Open Data Charter*, cioè la *Carta Internazionale dei Dati Aperti*. Con questa convenzione si definirono i principi e le buone pratiche per la pubblicazione degli open data, nel tentativo di arrivare ad uno standard comune.

### 2.2.1 Il concetto di open government

Da quanto descritto precedentemente, è chiaro che per promuovere l'utilizzo e la diffusione degli open data sono stati necessari interventi a livello governativo, in grado di definire leggi specifiche sul tema. Inizialmente, la diffusione degli open data si è sviluppata intorno al settore pubblico, poiché le informazioni prodotte appartengono alla collettività. Ma in futuro, seguendo l'esempio di questo settore, gli open data potranno essere prodotti e diffusi sempre di più anche da enti privati, come aziende e organizzazioni, in modo da incrementare lo sviluppo economico e di nuova conoscenza, oltre alla creazione di nuovi servizi.

---

<sup>1</sup>Un dataset è una collezione di dati strutturati, rappresentata ad esempio dal contenuto di una singola tabella di un database.

Gli open data, nell'ambito della pubblica amministrazione, sono fortemente legati al concetto di *open government*, cioè una modalità di esercizio del potere, sia a livello centrale sia locale, basato su strumenti e tecnologie che consentono alle amministrazioni di essere aperte e trasparenti nei confronti dei cittadini, i quali possono così controllarne l'operato. L'open government può quindi essere inteso come un'evoluzione dell'*e-government*, cioè il processo iniziato negli anni novanta con il quale la pubblica amministrazione si è dotata delle nuove tecnologie informatiche e di comunicazione. I tre principi fondamentali su cui si basa l'open government sono i seguenti [22]:

- **Trasparenza:** i cittadini devono avere la possibilità di accedere a tutte le informazioni necessarie per conoscere il funzionamento e l'operato delle pubbliche amministrazioni.
- **Partecipazione:** i cittadini, senza alcuna discriminazione, devono essere coinvolti nei processi decisionali, collaborando nella definizione delle politiche per contribuire al bene comune e all'efficienza delle amministrazioni.
- **Responsabilità:** i governi hanno l'obbligo di informare i cittadini del proprio operato e delle proprie decisioni, garantendo la piena responsabilità dei risultati conseguiti.

## 2.3 Limitazioni degli open data

Come spiegato nei capitoli precedenti, i dati e le informazioni al giorno d'oggi sono sempre più preziose e rappresentano un'importante risorsa. Ne consegue che paradossalmente è proprio questo il principale limite alla diffusione degli open data, poiché non tutti sono disposti a rendere di pubblico utilizzo il proprio patrimonio informativo. I sostenitori del concetto di open data affermano però che qualsiasi tipo di limitazione alla diffusione degli stessi sia dannosa per il bene della comunità e i dati dovrebbero essere disponibili senza restrizioni o forme di pagamento, per aumentare la diffusione e la creazione di conoscenza. Il problema appena descritto è riassumibile con la seguente dichiarazione di John Wilbanks, chief commons officer di *Sage Bionetworks*:

*“Numerosi scienziati hanno sottolineato con ironia che proprio nel momento storico in cui disponiamo delle tecnologie per consentire la disponibilità dei dati scientifici a livello globale e dei sistemi di distribuzione che ci consentirebbero di ampliare la collaborazione e accelerare il ritmo e la profondità della scoperte... siamo occupati a bloccare i dati e a prevenire l’uso di tecnologie avanzate che avrebbero un forte impatto sulla diffusione della conoscenza.”*

Mentre i dati generati da aziende private, le quali hanno investito risorse per poterli produrre, è comprensibile che vengano distribuiti in modo aperto solo su decisione dell’azienda stessa, tale ragionamento non può essere fatto per i dati generati dalla pubblica amministrazione. Questi ultimi infatti, grazie all’open government, devono tornare alla collettività sotto forma di dati aperti e gratuiti, in quanto prodotti con denaro pubblico. Spesso accade che siano gli stessi creatori di dati a sottovalutarne l’importanza, e di conseguenza a non considerare nemmeno l’idea di pubblicarli tramite licenze aperte. Le sfide da affrontare in futuro per lo sviluppo degli open data possono riassumersi nei seguenti tre punti chiave [14]:

- Pubblica amministrazione, aziende, organizzazioni ed enti di altro genere devono pubblicare sempre più dati in formato aperto, migliorandone anche qualità, precisione, completezza e consistenza. Inoltre devono rendere la modalità di accesso a tali informazioni sempre più efficiente e semplice da utilizzare.
- Solo pochi soggetti sono in grado di utilizzare direttamente gli open data, molti altri potranno accedere a questi dati grazie al lavoro di intermediari e sviluppatori capaci di riorganizzare, trasformare e presentare nuovamente i dati tramite apposite soluzioni software.
- È necessario che aumenti la consapevolezza del valore degli open data e incrementare il loro contributo alla crescita economica e sociale. Di conseguenza l’aggregazione e l’interpretazione dei dati sono fondamentali affinché essi diventino di interesse per tutti i soggetti coinvolti (settore pubblico, settore privato, settore no profit e cittadini).

## 2.4 Le licenze aperte

Esistono molti tipi diversi di licenze, sia aperte che chiuse, ovviamente gli open data devono essere dotati di licenze aperte che ne consentano la consultazione e il riutilizzo. Solitamente, la distribuzione di contenuti creativi è tutelata da un sistema che non permette di utilizzare e modificare liberamente i materiali protetti. Le *licenze chiuse* rappresentano quindi la soluzione tradizionale, che consiste nel riservare *tutti i diritti* al titolare del *copyright* (diritto di copia) sull'opera. Di conseguenza l'utente, per utilizzare i contenuti, deve chiedere esplicitamente il consenso al titolare dei diritti.



Open Data Commons



Figura 2.3: Loghi di alcune licenze aperte

Per questi motivi, le licenze chiuse violano i principi fondamentali degli open data, che quindi devono essere accompagnati da un altro tipo di licenze: le *licenze aperte*, che operano tramite il principio del *copyleft* (permesso d'autore). Gli enti che pubblicano open data sono tenuti ad applicare agli stessi una licenza aperta, che permetta agli utilizzatori dei dataset di capire con chiarezza che si tratta di open data. Le licenze aperte, non si soffermano troppo sui limiti di utilizzo dei dati, ma tendono a garantire una serie di diritti a chi entra in possesso degli stessi. Questo tipo di licenze, tutelano comunque l'autore dei dati, attribuendogli la paternità dell'opera ed evitando che quanto pubblicato possa essere alterato senza controllo e senza rispettare regole precise. Ma concedono agli utenti maggiori diritti, dandogli la possibilità di ridistribuire e modificare liberamente i dati, al fine di creare opere derivate.

Le più importanti licenze aperte da applicare agli open data e ad altri tipi di contenuti sono le *Creative Commons* (CC). Nate nel 2002, questo tipo di licenze consentono la libera circolazione del materiale creativo protetto dal diritto d'autore, ispirandosi al

modello *alcuni diritti riservati*. Sono strutturate in due parti, la prima indica le libertà che l'autore vuole concedere (condividere e rielaborare), la seconda individua le condizioni per l'utilizzo dell'opera (attribuzione, non commerciale, non opere derivate e condividi allo stesso modo). È l'autore che ha la facoltà di decidere quali diritti riservarsi e quali concedere liberamente, combinando tra loro le quattro clausole di base possono essere create sei diverse licenze principali, oltre alla licenza più permissiva *CC Zero* (CC0) che rappresenta il pubblico dominio. Ovviamente le clausole *non commerciale* e *non opere derivate* non rispettano i principi degli open data, di conseguenza non possono essere utilizzate.

Un'altra importante licenza aperta è la *Open Database License* (ODbL), nata nel 2009 e creata dall'*Open Data Commons*, parte dell'*Open Knowledge International*. Questa licenza è pensata appositamente per i database e i dati, al contrario delle *Creative Commons* che possono essere utilizzate anche per altri tipi di contenuti. Le mappe di *OpenStreetMap*, ad esempio, utilizzano proprio la licenza ODbL, che definisce tre diritti fondamentali (condividere, creare e adattare) e tre principi (attribuire, condividere allo stesso modo e mantenere aperto).

Dal 2010, diversi paesi del mondo hanno creato delle licenze aperte pensate appositamente per i loro territori e per offrire alle pubbliche amministrazioni degli strumenti chiari e certificati con cui facilitare la pubblicazione degli open data. I primi paesi che hanno creato queste licenze sono stati il Regno Unito con la *Open Government Licence* (OGL) e la Francia con la *Licence information publique librement réutilisable* (LIP). La licenza italiana *Italian Open Data Licence* (IODL), sviluppata a fine 2010 da *Formez PA*, permette di leggere, modificare, pubblicare i dati e creare opere derivate, citando la fonte delle informazioni [23].

L'*Open Knowledge International* raccomanda l'utilizzo delle licenze aperte conformi ai principi degli open data, tali licenze sono raccolte nella seguente tabella [24]:

Legenda: BY = richiesta di attribuzione, SA = richiesta di condivisione allo stesso modo

Licenza	BY	SA
Creative Commons CC Zero (CC0)	No	No
Open Data Commons Public Domain Dedication and Licence (PDDL)	No	No
Creative Commons Attribution 4.0 (CC-BY-4.0)	Si	No
Open Data Commons Attribution License (ODC-BY)	Si	No
Creative Commons Attribution Share-Alike 4.0 (CC-BY-SA-4.0)	Si	Si
Open Data Commons Open Database License (ODbL)	Si	Si

Tabella 2.1: Licenze aperte raccomandate dall'Open Knowledge International

## 2.5 Tipologie e formati degli open data

Come descritto all'inizio del capitolo, gli open data rappresentano informazioni aperte relative a molteplici tematiche differenti, ad esempio: cartografia, ambiente, genetica, composti chimici, formule matematiche e scientifiche, medicina, popolazione, politica, statistiche ed economia. È possibile suddividere gli open data in alcune categorie principali elencate di seguito:

- **Dati geospaziali:** open data utilizzati per realizzare mappe, localizzare strade, edifici e punti di interesse di ogni tipo tramite le coordinate geografiche. Uno dei principali esempi di utilizzo di questi dati è il progetto OpenStreetMap.
- **Istruzione e cultura:** open data riferiti a opere e prodotti culturali, generalmente conservati da biblioteche, gallerie, archivi e musei. Oltre a informazioni relative a scuole e università.
- **Scienza e tecnologia:** open data prodotti da ogni tipo di ricerca scientifica e tecnologica, dall'astronomia alla zoologia.
- **Economia e finanza:** open data relativi ai conti pubblici, come entrate e spese, oltre a informazioni sui mercati finanziari.
- **Governo e società:** open data relativi alla pubblica amministrazione, al suo operato, oltre che alla politica e alla società che ci circonda.

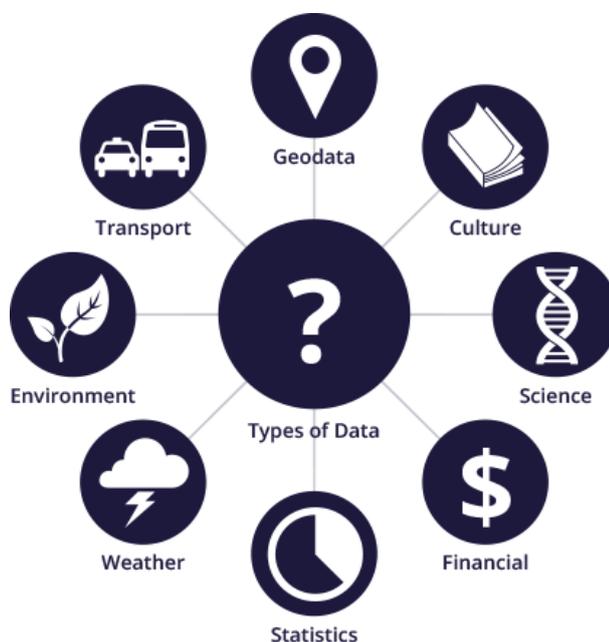


Figura 2.4: Principali tipologie di open data

- **Statistica:** open data prodotti da uffici ed enti statistici, come indicatori sociali, economici, demografici e turistici. Uno dei principali esempi italiani di pubblicazione di questi dati è l'Istat.
- **Meteo:** open data relativi alle previsioni meteorologiche e al clima, come temperature medie, precipitazioni e altri indicatori.
- **Ambiente:** open data relativi all'ambiente, ad esempio informazioni sulla presenza di fattori inquinanti, sulla qualità delle acque e sulla gestione dei rifiuti.
- **Salute e sport:** open data relativi a salute, medicina e sport, ad esempio tassi e cause di mortalità, incidenza delle malattie e svolgimento di attività sportive.
- **Trasporti:** open data relativi a tutti i tipi di mezzi di trasporto terrestri, aerei e acquatici, ad esempio orari, tragitti e tempi di percorrenza.

Come descritto nelle caratteristiche e nella classificazione degli open data, tali tipologie di dati devono essere pubblicati in formati aperti e facilmente utilizzabili da un computer

per eseguire operazioni di lettura ed elaborazione. Ad esempio, se si utilizzasse il formato PDF per la codifica dei dati sarebbe comoda la lettura da parte di un essere umano, ma non per un computer che deve elaborarli. I formati principali in cui sono disponibili gli open data sono elencati di seguito:

- **JSON:** il formato JSON (JavaScript Object Notation) è adatto all'interscambio di dati fra applicazioni client/server, di facile lettura da parte di un computer tramite apposite funzioni di *parsing*.
- **CSV:** il formato CSV (Comma Separated Values) si basa su file di testo formattati in un modo specifico, in grado di memorizzare tabelle di database o fogli di calcolo.
- **RDF:** il formato RDF (Resource Description Framework) serve per la codifica, lo scambio e il riutilizzo di metadati strutturati. Consente l'interoperabilità semantica tra applicazioni che condividono le informazioni sul web, tramite i linked data e i linked open data.
- **XML:** il formato XML (Extensible Markup Language) è un metalinguaggio per la definizione di linguaggi di markup. Consente di definire e controllare le proprietà dei dati contenuti in un file.
- **HTML:** il formato HTML (Hypertext Markup Language) è un linguaggio di markup nato per realizzare documenti ipertestuali. Spesso capita che open data semplici siano diffusi in questo formato.
- **TXT:** il formato TXT (Text File) è molto semplice e contiene testo puro che i computer sono in grado di leggere facilmente. Generalmente non include metadati strutturali, di conseguenza servono parser specifici per la lettura automatica.
- **GeoJSON:** il formato GeoJSON è utilizzato per archiviare collezioni di geometrie spaziali (come punti, linee spezzate e poligoni) i cui attributi sono descritti attraverso il formato JSON. Si tratta di uno dei principali formati per la codifica di open data geospaziali, per creare mappe interattive o individuare punti di interesse.
- **KML:** il formato KML (Keyhole Markup Language) è un linguaggio basato su XML per la gestione di open data geospaziali. I file KML specificano insiemi di

elementi, come punti, poligoni, modelli 3D e descrizioni per creare ad esempio mappe interattive. Anche Google Maps e Google Earth utilizzano questo formato.

- **XLS:** il formato XLS indica i fogli di calcolo realizzati tramite Microsoft Excel. Pur non essendo un formato aperto, è talmente diffuso che spesso gli open data vengono pubblicati in tale formato, violando di conseguenza i loro principi. Se il file contiene macro o formule matematiche può non essere di facile lettura.

Oltre ai principali formati appena descritti, esistono molti altri formati aperti con cui vengono pubblicati gli open data. Talvolta però, come nel caso del formato XLS, capita che i dati vengano pubblicati in formati proprietari o che non rispettano le linee guida e i principi degli open data. Questo avviene poiché tali formati sono largamente diffusi o l'ente che pubblica i dati non ha provveduto a convertirli in formati adeguati. Esempi di questo tipo sono file DOC, ODF e PDF, ma anche immagini o file multimediali utilizzati poiché i supporti originali non erano digitali. La logica dell'open data suggerisce comunque di pubblicare i dati anche se non soddisfano tutti i requisiti necessari piuttosto che non pubblicarli affatto [17].

## 2.6 Come si utilizzano gli open data

Gli open data possono essere utilizzati per semplice consultazione, ricerche oppure per essere implementati e utilizzati all'interno di progetti software di ogni genere, come siti web e applicazioni mobili. Per prima cosa è necessario individuare gli open data di cui si ha bisogno, in modo che possano fornire un valore aggiunto al progetto. Gli open data possono trovarsi su appositi portali web realizzati dalle amministrazioni pubbliche come comuni, regioni o sui portali nazionali che raccolgono buona parte dei dati. Esistono anche altri siti di terze parti che mettono a disposizione open data originali, oppure repository che raccolgono dati provenienti da più fonti. Solitamente, nei portali web sui quali sono raccolti gli open data, è presente un *database management system* che permette di archiviare, catalogare e distribuire i dati. Uno dei principali software di questo tipo è il *Comprehensive Knowledge Archive Network* (CKAN), sviluppato dall'Open Knowledge International.



Figura 2.5: Logo di CKAN

Successivamente, dopo aver individuato i dati di interesse, può iniziare la fase di recupero di tali informazioni. Solitamente gli open data sono disponibili in due modi:

- **Download di dataset:** questa operazione può avvenire in modi diversi a discrezione della fonte che rende disponibili i dati. Le alternative principali sono: download diretto dal sito web, utilizzo di connessioni FTP oppure tramite file torrent.
- **Utilizzo di API<sup>2</sup>:** anche in questo caso è l'ente che distribuisce i dati a decidere se offrirli o meno tramite API. Utilizzando questo sistema chi ha necessità di utilizzare i dati può eseguire delle *query*, ovvero interrogazioni specifiche sul dataset messo a disposizione dalla fonte. In questo modo non è necessario eseguire il download dell'intero insieme di dati, ma è possibile richiedere solo quelli di interesse.

Nel caso in cui si utilizzino le API per recuperare i dati, le risposte alle query conterranno già le informazioni specifiche di cui si ha bisogno in un particolare formato, e si potranno eseguire richieste HTTP ogni volta che si necessita di nuovi open data. Se invece si esegue il download di dataset, è necessario caricare i file sul proprio server o in un altro tipo di archiviazione. Una soluzione efficace è quella di importare il contenuto del dataset all'interno del database (MySQL o di altro genere) che si utilizza, in questo modo è possibile eseguire delle query per estrapolare dinamicamente i dati di interesse dal dataset. Tali open data potranno essere elaborati e visualizzati in modi diversi, ad esempio come testo o su una mappa interattiva, nel sito web, nell'applicazione mobile o nel software che si sta sviluppando.

---

<sup>2</sup>Una API (Application Programming Interface) è un insieme di procedure, funzioni e strumenti software sviluppati appositamente per facilitare l'esecuzione di un determinato compito, favorendo la comunicazione con altre applicazioni.

Un aspetto importante da considerare, quando si sceglie di utilizzare gli open data rispetto ad altri dati non aperti, è che molto probabilmente tali dati saranno sempre fruibili liberamente e gratuitamente anche in futuro. Se invece si scelgono sistemi chiusi, non esistono garanzie che un servizio attualmente libero e gratuito rimanga tale. Ad esempio, l'utilizzo delle API di Google Maps (che è un sistema chiuso) al momento è gratuito, ma con notevoli limiti introdotti recentemente e nulla vieta che in futuro tali limiti diventino ancora più stringenti. Invece, le API di OpenStreetMap (che è un sistema aperto) è molto probabile che non diventeranno mai a pagamento. Un aspetto negativo nell'uso degli open data rispetto ad altri dati chiusi è che questi ultimi, essendo a pagamento, sono soggetti a meno limiti tecnici utilizzandoli tramite API. Questo poiché gli enti che li pubblicano dispongono di maggiori risorse, ad esempio Google Maps rispetto a OpenStreetMap offre i dati senza problemi di velocità di caricamento o limiti sul numero di query che si possono eseguire.

## 2.7 Principali fonti di open data

In questa sezione sono descritte in modo più dettagliato le principali fonti che possono essere utilizzate per recuperare gli open data. In particolare, per ogni fonte sono raccolte una serie di informazioni utili:

- Una panoramica della fonte che pubblica gli open data;
- Come si utilizzano gli open data pubblicati dalla fonte;
- In quali formati sono disponibili gli open data;
- Quali sono le limitazioni degli open data.

Gli open data analizzati in dettaglio sono quelli forniti da: OpenStreetMap, Istat, Dati.gov.it, DatiOpen.it, Comune di Bologna, Arpa e DBpedia. Inoltre, nella sezione *Altre fonti* sono raccolte ulteriori sorgenti di open data.

### 2.7.1 OpenStreetMap

OpenStreetMap (OSM) è un progetto aperto e collaborativo che ha lo scopo di creare mappe e cartografie a contenuto libero di tutto il mondo. Il progetto mira a raccogliere dati geografici, la cui caratteristica fondamentale è che possiedono la licenza libera Open Database License. Di conseguenza è possibile utilizzare liberamente gli open data di OpenStreetMap per qualsiasi scopo, con il solo vincolo di citare la fonte e usare la stessa licenza per eventuali opere derivate. Inoltre chiunque può contribuire al progetto, aggiungendo o modificando i dati.



Figura 2.6: Logo di OpenStreetMap

Tuttavia la quantità e qualità di informazioni non sono distribuite omogeneamente in tutto il mondo, poiché dipende dal lavoro svolto dai volontari che mappano ogni luogo. I risultati più evidenti di OpenStreetMap sono le mappe online, che però rappresentano solo la punta dell'iceberg di ciò che si può realizzare con questi open data. Ad esempio è possibile implementarli in siti web e applicazioni mobili, usarli come base nei videogiochi, calcolare distanze oppure utilizzarli per scopi umanitari in caso di catastrofi naturali e guerre.

Un'alternativa proprietaria alle pratiche di *crowdsourcing* di OpenStreetMap era *Google Map Maker*, nel quale i contributi dell'utente potevano essere inseriti su Google Maps a fronte di validazione, impedendone però il riutilizzo da parte di terzi e diventando di proprietà di Google. Le mappe di OpenStreetMap sono create usando i dati registrati da dispositivi GPS mobili, fotografie aeree e dati provenienti da altre fonti libere. Sia le immagini renderizzate che i dati vettoriali, oltre che lo stesso database di dati geospaziali sono pubblicati con la licenza libera Open Database License. OpenStreetMap ha preso

spunto da progetti come Wikipedia, dove gli utenti sono liberi di modificare i dati con appositi editor e caricare nuove tracce GPS nel database per migliorare costantemente il servizio, inoltre è presente un archivio storico delle modifiche apportate [25].

### Come si utilizzano gli open data di OpenStreetMap

Per utilizzare gli open data forniti da OpenStreetMap ci sono due possibilità: eseguire il download dei dataset o utilizzare apposite API per richiedere solo i dati di interesse.

- **Download di dataset:** si può eseguire il download degli open data pubblicati da OpenStreetMap in vari modi. Dal sito web ufficiale *openstreetmap.org* premendo sul bottone *Esporta*, per porzioni ridotte di un luogo, ad esempio una singola città o un'area geografica ristretta. Oppure, sempre tramite il sito web ufficiale *Planet OSM* o altri siti web, come *Geofabrik*, e tool di terze parti come *Osmosis*, è possibile scaricare mappe di dimensioni maggiori, ad esempio l'intera mappa del mondo, di singoli continenti, nazioni o regioni.
- **Utilizzo di API:** si possono utilizzare le API di OpenStreetMap sia per eseguire il download dei dataset che per fare richieste più specifiche. OpenStreetMap mette a disposizione diversi tipi di API. Le *Main API* permettono di recuperare gli open data di piccole aree ma soprattutto di modificare i dati presenti nel database di OpenStreetMap, sono quindi utilizzate soprattutto per gli editor. Le *XAPI* sono pensate per la sola lettura di dati e permettono di recuperare porzioni più grandi di aree a livello di città. Infine le *Overpass API* sono anch'esse pensate per la sola lettura di dati e permettono query più complesse su aree più vaste.

Il formato con cui sono rappresentate le mappe di OpenStreetMap è *OSM XML*, mentre le risposte alle query utilizzando le varie API possono essere in formato XML, JSON o CSV. Come descritto precedentemente, se si sceglie di eseguire il download dell'insieme dei dati, è poi necessario importare tali dati in formato OSM XML in un database per poter eseguire delle query su di essi e creare la propria istanza privata di OpenStreetMap. Possono essere utilizzati vari tipi di database, come: *PostgreSQL*, *MySQL*, *SQLite*, *MonetDB*, *CouchDB*, *MongoDB*, *Hadoop* e *Hive*.

Per creare la mappa e visualizzarla sono disponibili vari metodi. Il modo più semplice è quello di eseguire un *embed* della mappa tramite un *iframe* configurabile direttamente dal sito web di OpenStreetMap. Se invece si vuole modificare la mappa ed eseguire operazioni più avanzate sulla stessa si possono utilizzare librerie dedicate come *OpenLayers* e *Leaflet*. Con queste librerie è possibile ad esempio aggiungere *marker* alla mappa e scegliere da quale fonte (i server di OpenStreetMap o il nostro server privato) recuperare le *tiles*, cioè le piastrelle che unite tra loro formano la mappa stessa.

Utilizzando invece le API, le query sono eseguite direttamente sui server di OpenStreetMap o dei suoi partner. Di conseguenza è possibile recuperare direttamente i dati di nostro interesse, ad esempio in formato JSON, per poi visualizzarli in forma testuale, su una mappa oppure elaborarli. Per aiutare gli sviluppatori a creare le query è disponibile un apposito *tool* chiamato *Overpass Turbo*, che con il suo *wizard* è in grado di tradurre un comando scritto con un linguaggio comprensibile da un essere umano in una specifica query, adatta per essere utilizzata con le Overpass API.

### Limitazioni degli open data di OpenStreetMap

I limiti relativi alle richieste verso OpenStreetMap non sono precisi ma è sconsigliato eseguirne un numero elevato, poiché non è sempre garantita una risposta. Essendo OpenStreetMap un servizio gratuito e aperto, viene consigliato di eseguire il download dei dataset per poi operare sui propri server in modo da non gravare su quelli di OpenStreetMap (poiché si tratta di server donati), specialmente se si prevede un utilizzo massiccio dei dati offerti. Come descritto precedentemente, eseguire il download dell'insieme di dati comporta un notevole utilizzo dello storage sul server proprietario. Ad esempio, il download dell'intera mappa italiana consiste in circa 1.3 GB di dati. Una delle operazioni che utilizza molte risorse è la creazione e l'interazione con una mappa embed di OpenStreetMap all'interno di un sito web, questo poiché la mappa è costruita unendo tra loro tante tiles. Utilizzando invece le API di OpenStreetMap per richiedere open data specifici, i limiti di query variano in base all'API che si utilizza, ad esempio le Overpass API garantiscono 10.000 query al giorno [26] [27].

### 2.7.2 Istat

L'Istat (Istituto nazionale di statistica) è un ente di ricerca pubblico italiano. Si tratta dell'organismo che produce e diffonde il maggior numero di dati e statistiche in Italia, messi a disposizione di cittadini, imprese e istituzioni tramite la licenza Creative Commons. Le ricerche dell'Istat vertono su molteplici tematiche relative a tutto il territorio nazionale, tra le quali:

- Censimenti su popolazione, industria, servizi e agricoltura;
- Indagini campionarie sulle famiglie e indagini economiche.

L'Istat dispone di due differenti piattaforme web, nelle quali pubblica gli open data prodotti e consente il loro libero utilizzo da parte di cittadini, aziende e istituzioni:

- **Linked Open Data:** la piattaforma sperimentale Linked Open Data (LOD) dell'Istat consente di utilizzare dati collegati in formato aperto (solo una parte dell'intero database dell'Istat), sulla base di tecnologie e standard del web semantico. I linked open data, interrogabili direttamente da qualsiasi applicazione, rispondono alle esigenze degli utenti di disporre di dati standardizzati e interoperabili.
- **I.Stat:** la piattaforma I.Stat consente di accedere al database costantemente aggiornato delle statistiche prodotte dall'Istat. Tali statistiche sono ricercabili per tema e il sistema è interrogabile anche per parola chiave. I dati sono presentati sotto forma di tavole multidimensionali che gli utenti possono esportare in formato CSV o XLS. Inoltre è possibile creare tabelle personalizzate agendo sulle variabili e il periodo di riferimento. Attraverso un web service SDMX (Statistical Data and Metadata eXchange), che consente l'interrogazione diretta *machine-to-machine*, è possibile formulare specifiche query sui dati ed effettuare il download dei risultati. Inoltre, con il progetto *LinkedStat*, è in atto la traduzione in formato linked open data dell'intero insieme di dati pubblicati su I.Stat per renderli accessibili tramite un endpoint SPARQL.



Figura 2.7: Loghi di I.Stat e Linked Open Data

### Come si utilizzano gli open data di Istat

Per utilizzare gli open data forniti dall'Istat ci sono due possibilità: eseguire il download dei dataset o utilizzare apposite API per richiedere solo i dati di interesse.

- **Download di dataset:** per eseguire il download dei dataset dalla piattaforma Linked Open Data, bisogna raggiungere la sezione *Dataset/Ontologie* del sito web e selezionare i file di interesse. Invece per scaricare i dati dalla piattaforma I.Stat, è necessario selezionare il dataset di interesse dalla colonna a sinistra e successivamente premere sul bottone *Esportazioni* sulla destra. Si può scegliere in quale formato scaricare i dati, ad esempio CSV o XLS, in quest'ultimo caso si potrà scaricare direttamente il file, se invece si sceglie il formato CSV sarà necessario inserire un indirizzo email. Dopo alcuni minuti, quando il file da scaricare sarà pronto, si riceverà un'email contenente il link per eseguire il download del dataset.
- **Utilizzo di API:** per utilizzare le API fornite dalla piattaforma Linked Open Data ci sono varie possibilità, è infatti possibile eseguire interrogazioni guidate o interrogazioni libere. Nel primo caso bisogna raggiungere le sezioni *Basi territoriali* o *Variabili censuarie* del sito web. Successivamente è possibile seguire un processo automatico che guida lo sviluppatore nella creazione delle query. Nella sezione *Basi territoriali* si possono costruire query per il recupero di dati territoriali selezionando la partizione di territorio che si preferisce. Invece nella sezione *Variabili censuarie* si possono costruire query per il recupero di determinate variabili censuarie, selezionando il livello di aggregazione di interesse. Per costruire interrogazioni libere bisogna raggiungere la sezione *SPARQL endpoint*, dalla quale si ha piena libertà di eseguire interrogazioni specifiche, inoltre sulla destra sono disponibili alcuni esempi. Per utilizzare le API del portale I.Stat è possibile accedere ad un apposito web service SDMX in grado di eseguire query sui database dell'Istat.

Nonostante ciò risulta più comodo utilizzare le API di Linked Open Data, poichè è possibile costruire le query utilizzando lo standard SPARQL.

Eseguendo il download dei dataset dalla piattaforma Linked Open Data si otterranno i dati in formato CSV, mentre dalla piattaforma I.Stat è possibile scegliere tra CSV e XLS. Invece utilizzando le API della piattaforma Linked Open Data è possibile scegliere tra i formati CSV, JSON, XML, TXT e TSV.

### **Limitazioni degli open data di Istat**

Sono presenti diversi limiti relativi alle query eseguibili verso il portale Linked Open Data di Istat, poichè ancora in fase sperimentale. Le interrogazioni permettono il download di file molto grandi, anche superiori a 10 MB, poichè è stato deciso di non limitare la grandezza dei dataset restituiti, ma di limitare il tempo di download ad un massimo di 10 minuti per ogni interrogazione. Se si richiedono open data relativi a comuni di grandi dimensioni, tale tempo può essere insufficiente per interrogazioni a livello di sezione di censimento, in questo caso è consigliato eseguire il download dell'intero dataset. Inoltre alcune interrogazioni complesse possono avere tempi di risposta di diversi minuti, di conseguenza è necessario attendere il termine dell'operazione prima di lanciare un'altra query [28] [29].

### **2.7.3 Dati.gov.it**

Dati.gov.it è il portale web italiano dedicato alla raccolta di buona parte degli open data distribuiti dalle pubbliche amministrazioni di tutto il territorio nazionale. Disponibile dal 18 ottobre 2011, raccoglie in un unico portale la maggior parte degli open data messi a disposizione da comuni, regioni, ministeri e alti tipi di enti e pubbliche amministrazioni tramite un processo collaborativo. La piattaforma web, che nel 2017 è stata rinnovata, è promossa dal Governo Italiano e gestita dall'Agenzia per l'Italia digitale con il supporto di Formez PA. Inoltre Dati.gov.it è lo strumento con il quale l'Agenzia per l'Italia digitale, centro di competenza nazionale sul tema degli open data, promuove le politiche di valorizzazione del patrimonio informativo pubblico nazionale.



Figura 2.8: Logo di Dati.gov.it

Per questo motivo, Dati.gov.it mette a disposizione degli sviluppatori e delle amministrazioni una serie di risorse utili per approfondire la tematica degli open data, migliorarne la qualità e favorirne il riutilizzo. Infine Dati.gov.it, come i portali web dedicati ai dati aperti degli altri Stati europei, contribuisce ad alimentare lo *European Data Portal*, ovvero il portale web che raccoglie tutti gli open data degli Stati membri dell'Unione Europea.

### Come si utilizzano gli open data di Dati.gov.it

Per utilizzare gli open data forniti da Dati.gov.it ci sono due possibilità: eseguire il download dei dataset o utilizzare apposite API per richiedere solo i dati di interesse.

- **Download di dataset:** si può eseguire il download degli open data pubblicati da Dati.gov.it raggiungendo la sezione *Dati* del portale web. Successivamente è possibile cercare i dati di interesse tramite l'apposita barra di ricerca anche selezionando tema, organizzazione, formato e licenza per una ricerca più specifica. Una volta selezionato il dataset di interesse, è possibile scaricarlo in vari formati.
- **Utilizzo di API:** si possono utilizzare le API di Dati.gov.it per fare richieste più specifiche sui dataset. Il portale web supporta un sottoinsieme delle API di CKAN che possono essere utilizzate attraverso delle richieste HTTP in grado di fornire i risultati.

Eseguendo il download dei dataset da Dati.gov.it è possibile ottenere gli open data nei formati JSON, CSV, RDF, XML, HTML, GeoJSON, KML, XLS e altri in base alla disponibilità. Invece utilizzando le API si otterranno dataset in formato JSON.

### Limitazioni degli open data di Dati.gov.it

Non sono specificati limiti precisi relativi al numero di query che è possibile eseguire in un determinato intervallo di tempo. Tuttavia è suggerita l'esecuzione di chiamate di massimo 100 record per volta tramite il parametro *limit=100* per questioni di performance [30].

### 2.7.4 DatiOpen.it

DatiOpen.it è un portale web che raccoglie centinaia di dataset in formato open data messi a disposizione da pubbliche amministrazioni, enti e aziende di vario genere. L'idea di base di DatiOpen.it è quella di dare una spinta decisa al fenomeno degli open data in Italia con le seguenti azioni:

- Raccogliere e documentare la maggiore quantità possibile di open data italiani. La redazione di DatiOpen.it monitora quotidianamente il panorama open data italiano alla ricerca di nuovi dati da catalogare ed inserire nel portale. L'obiettivo è quello di diventare il punto di riferimento nazionale per chiunque ricerchi un dato aperto.
- Permettere la visualizzazione direttamente dal web degli open data pubblicati. Grazie alla tecnologia *StatPortal OpenData*, i dati non sono solo scaricabili, come avviene nella maggior parte dei portali web di open data, ma anche visualizzabili in tabelle, grafici e mappe interattive.

Lo scopo principale del progetto DatiOpen.it è quello di rendere l'enorme patrimonio di open data che si sta costituendo accessibile facilmente al grande pubblico, in modo da valorizzare tali dati e diffonderli. Inoltre enti pubblici, aziende private, cittadini, professionisti e studenti possono caricare i dati in loro possesso sulla piattaforma.

### Come si utilizzano gli open data di DatiOpen.it

Per utilizzare gli open data forniti da DatiOpen.it ci sono due possibilità: eseguire il download dei dataset o utilizzare apposite API per richiedere solo i dati di interesse.



Figura 2.9: Logo di DatiOpen.it

- **Download di dataset:** si può eseguire il download degli open data pubblicati da DatiOpen.it raggiungendo la sezione *Ricerca open data* del portale web. Successivamente è possibile cercare i dati di interesse tramite l'apposita barra di ricerca anche selezionando fonte, tema, formato, autore, licenza e parole chiave per una ricerca più specifica. Una volta selezionato il dataset di interesse, è possibile scaricarlo in vari formati.
- **Utilizzo di API:** si possono utilizzare le API di DatiOpen.it per fare richieste più specifiche sui dataset. L'intero catalogo degli open data di DatiOpen.it è accessibile tramite un sottoinsieme delle API di CKAN e tramite un endpoint SPARQL. Questo consente l'accesso e l'interrogazione delle informazioni contenute nel catalogo in un'ottica di riuso, integrazione con altri portali, applicazioni e sviluppo di nuovi servizi.

Eseguendo il download dei dataset da DatiOpen.it è possibile ottenere gli open data nei formati JSON, CSV, RDF, XML, XLS, OData, SHP, MDB, DBF e altri in base alla disponibilità. Invece utilizzando le API si otterranno dataset in formato JSON.

### Limitazioni degli open data di DatiOpen.it

Non sono specificati limiti precisi relativi al numero di query che è possibile eseguire in un determinato intervallo di tempo. Su DatiOpen.it sono presenti per la maggior parte dati a livello locale, come comuni e regioni, ma anche alcuni dataset a livello nazionale, come ad esempio quelli di OpenStreetMap, di enti nazionali e di alcuni ministeri [31].

### 2.7.5 Comune di Bologna

Il Comune di Bologna, come altre città e regioni, ha realizzato un portale web chiamato *Open Data Comune di Bologna* nel quale pubblica i propri dati in formato aperto. Si tratta di uno dei comuni italiani più virtuosi nel campo degli open data, pubblicando sul suo portale una consistente quantità di dati. In questo modo aumenta trasparenza e accessibilità nei confronti di cittadini, associazioni e imprese, che possono così partecipare attivamente nei processi decisionali relativi al territorio e costruire un nuovo modello di sviluppo economico seguendo i principi dell'open government. Il Comune di Bologna promuove quindi il riutilizzo dei suoi open data per la creazione di applicazioni e servizi innovativi, che rappresentano un valore aggiunto a beneficio dell'intera comunità.

#### Come si utilizzano gli open data del Comune di Bologna

Per utilizzare gli open data forniti dal Comune di Bologna ci sono due possibilità: eseguire il download dei dataset o utilizzare apposite API per richiedere solo i dati di interesse.

- **Download di dataset:** si può eseguire il download degli open data pubblicati dal Comune di Bologna raggiungendo la sezione *Dati* del portale web oppure cercare i dati di interesse tramite l'apposita barra di ricerca. Inoltre è possibile scegliere argomento, tag e responsabile per una ricerca più specifica. Una volta selezionato il dataset di interesse, si può scaricare in vari formati.
- **Utilizzo di API:** si possono utilizzare le API del Comune di Bologna per fare richieste più specifiche sui dataset, non tutti però sono interrogabili tramite API. Inoltre nella sezione *Linked data* del portale web sono raccolti tutti i linked open data del Comune di Bologna, interrogabili tramite un apposito endpoint SPARQL.

Eseguendo il download dei dataset del Comune di Bologna o utilizzando le API è possibile ottenere gli open data nei formati JSON, CSV, RDF, XML, XLS, SHP e altri in base alla disponibilità.



Figura 2.10: Logo di Open Data Comune di Bologna

### Limitazioni degli open data del Comune di Bologna

Non sono specificati limiti precisi relativi al numero di query che è possibile eseguire in un determinato intervallo di tempo [32] [33].

#### 2.7.6 Arpa

L'Arpa (Agenzia regionale per la protezione ambientale) è un ente della pubblica amministrazione italiana, gestito direttamente dalle varie regioni d'Italia. L'Arpa svolge soprattutto le seguenti funzioni per il controllo e la tutela dell'ambiente:

- Controllo delle fonti e dei fattori di inquinamento di aria, acqua e suolo, oltre all'inquinamento acustico ed elettromagnetico;
- Monitoraggio di tutte le componenti ambientali: meteo, clima, qualità di aria, acqua e suolo oltre al livello sonoro ambientale;
- Controllo e vigilanza del rispetto della normativa vigente relativa all'ambiente;
- Supporto tecnico-scientifico, strumentale e analitico agli enti che hanno funzioni amministrative in ambito ambientale (come comuni, province e regioni);
- Sviluppo di un sistema informativo ambientale che sia di supporto e a disposizione di istituzioni e organizzazioni interessate.

Le agenzie Arpa delle varie regioni italiane offrono spesso portali web realizzati appositamente per la pubblicazione di open data ambientali di vario genere, compresi dati climatici e meteorologici. Tali dati sono molto utili per essere implementati all'interno di progetti software in grado di sfruttarli e utilizzarli in modi innovativi.



Figura 2.11: Logo di Arpa Emilia-Romagna

### Come si utilizzano gli open data di Arpa

Per utilizzare gli open data forniti dalle agenzie Arpa ci sono due possibilità: eseguire il download dei dataset o utilizzare apposite API per richiedere solo i dati di interesse.

- **Download di dataset:** si può eseguire il download degli open data pubblicati dalle varie agenzie Arpa accedendo ai relativi portali regionali. Successivamente è possibile cercare il dataset di interesse e una volta selezionato scaricarlo in vari formati.
- **Utilizzo di API:** nei portali web delle agenzie Arpa che lo prevedono si possono utilizzare le API per fare richieste più specifiche sui dataset. Ad esempio l'agenzia Arpa dell'Emilia Romagna consente l'utilizzo delle API di CKAN, poiché il portale web è realizzato con questo sistema.

Eseguendo il download dei dataset dai portali Arpa regionali o utilizzando le API solitamente è possibile ottenere gli open data nei formati CSV, JSON, RDF, XML, XLS e altri in base alla disponibilità.

### Limitazioni degli open data di Arpa

Se presenti, i limiti precisi relativi al numero di query che è possibile eseguire in un determinato intervallo di tempo sono riportati nelle pagine dedicate sui vari portali Arpa regionali. I maggiori limiti relativi agli open data forniti da Arpa sono che si tratta ovviamente di dati regionali e non nazionali. Non essendo ancora presente uno standard comune tra le regioni, lo stesso tipo di dato può essere disponibile in un'area geografica e assente in un'altra, oppure non essere aggiornato in modo omogeneo [34].

#### 2.7.7 DBpedia

DBpedia è un progetto nato nel 2007 dalla collaborazione dell'Università Libera di Berlino, dell'Università di Lipsia e di OpenLink Software. Il progetto mira ad estrarre dati strutturati da Wikipedia, la più vasta enciclopedia online del mondo, e alla pubblicazione degli stessi sul web tramite linked open data in formato RDF.



Figura 2.12: Logo di DBpedia

DBpedia è disponibile in 125 lingue e in totale la base di conoscenza estratta da Wikipedia descrive 38.3 milioni di entità tra cui persone, luoghi, opere creative, istituzioni, aziende, organizzazioni, specie naturali e malattie. Le pagine di Wikipedia sono costituite soprattutto da testo libero, ma contengono anche informazioni strutturate presenti nei cosiddetti *infobox*, ovvero gli schemi che si trovano sulla destra di molte pagine, nei

quali sono riassunte le informazioni fondamentali relative all'argomento trattato. Ad esempio dati relativi a coordinate geografiche, categoria di appartenenza, immagini, link ad altre pagine web e reindirizzamenti a pagine simili o in un'altra lingua. Periodicamente, DBpedia estrae tramite un apposito framework tali informazioni strutturate da Wikipedia per trasformarle in una vasta *knowledge base* interrogabile con le tecniche del web semantico.

### Come si utilizzano gli open data di DBpedia

Per utilizzare gli open data forniti da DBpedia ci sono due possibilità: eseguire il download dei dataset o utilizzare apposite API per richiedere solo i dati di interesse.

- **Download di dataset:** si può eseguire il download degli open data pubblicati da DBpedia raggiungendo la sezione *Downloads* del portale web, dal quale è possibile scaricare l'intera base di conoscenza. Altrimenti si possono cercare gli open data di interesse tramite un apposito browser e scaricarli in vari formati.
- **Utilizzo di API:** si possono utilizzare le API di DBpedia per fare richieste più specifiche sui dataset. Essendo il progetto realizzato tramite gli standard del web semantico, è possibile eseguire specifiche query sull'intero catalogo di informazioni mediante un endpoint SPARQL.

Eseguendo il download dei dataset da DBpedia o utilizzando le API è possibile ottenere gli open data nei formati CSV, JSON, RDF, XML, HTML e altri. Se invece si esegue il download dell'intero database di DBpedia i dati sono in formato RDF.

### Limitazioni degli open data di DBpedia

DBpedia raccomanda un utilizzo appropriato della piattaforma, imponendo alcuni limiti. È possibile eseguire al massimo 100 query al secondo e il tempo di esecuzione di ogni query non può superare 120 secondi. Inoltre ogni query può generare una risposta di massimo 10.000 righe. Nel caso in cui tali limiti di utilizzo non dovessero essere sufficienti è possibile creare una propria istanza privata di DBpedia [35].

### 2.7.8 Altre fonti

Oltre alle fonti di open data descritte precedentemente, esistono tanti altri portali web che si occupano di pubblicare open data e che funzionano in maniera analoga. Un'importante realtà è rappresentata da Open Data Hub Italia, piattaforma ricca di contenuti, ospitata da *SciamLab* e realizzata con il sistema CKAN [36]. Inoltre tutte le regioni italiane dispongono di portali propri o di terzi sui quali pubblicano open data, così come le principali città italiane. Anche diversi ministeri ed enti pubblici utilizzano piattaforme proprie per la diffusione degli open data, la stessa Università di Bologna ha lanciato a fine 2017 il suo portale web sperimentale chiamato *Opendata beta, University of Bologna*.



Figura 2.13: Loghi di alcune fonti di open data

Buona parte dei dataset di tutte le pubbliche amministrazioni italiane è comunque raccolta su *Dati.gov.it*, che funge da aggregatore [37]. Inoltre, tutti gli open data raccolti dai vari portali web nazionali dei singoli Stati dell'Unione Europea alimentano lo *European Data Portal* [38]. Invece nello *European Union Open Data Portal* sono raccolti tutti gli open data prodotti direttamente dalle istituzioni europee [39]. Altra importante piattaforma è *Data.gov*, nella quale confluiscono tutti gli open data diffusi dalle pubbliche amministrazioni degli Stati Uniti d'America e che funziona in modo molto simile all'italiana *Dati.gov.it* [40]. Infine, il portale *DataHub* raccoglie dataset aperti provenienti da tutto il mondo [41].

## 2.8 Statistiche e valore degli open data

Il fenomeno degli open data è in continua espansione e un numero sempre crescente di dataset vengono pubblicati e resi accessibili sul web, risulta quindi difficile eseguire una stima precisa sulla quantità totale di open data disponibili. Come si può osservare dalla figura 2.14, che mostra il grafico realizzato da Dati.gov.it aggiornato al 2015, il numero di dataset disponibili in Italia è sempre in crescita. Si nota però che all'aumentare della quantità dei dati disponibili non migliora anche la qualità del formato, infatti i dati classificati con cinque stelle di apertura, cioè i linked open data sono una minima parte del totale.

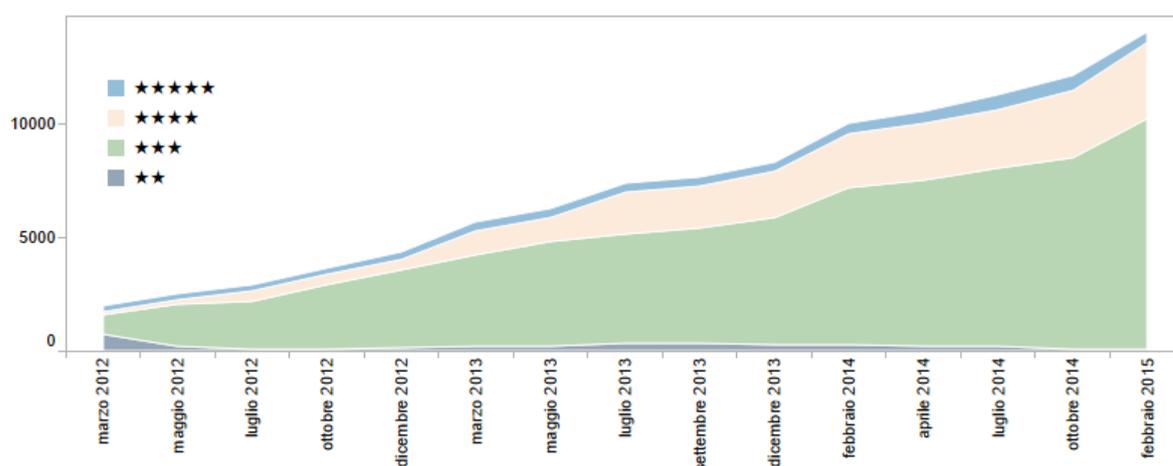


Figura 2.14: Numero di dataset pubblicati in Italia tra il 2012 e il 2015

Certamente rispetto a qualche anno fa la situazione è migliorata e anche in Italia il fenomeno degli open data si sta espandendo, senza raggiungere però il livello degli Stati Uniti d'America. Nella tabella 2.2 è riportato il numero di dataset pubblicati sui principali portali web descritti nella sezione precedente. Nella tabella 2.3 invece sono riportati i 10 dataset più scaricati dallo European Union Open Data Portal e le rispettive visualizzazioni. In entrambe le tabelle le informazioni sono aggiornate a settembre 2018. Secondo il report *Open Data Maturity in Europe 2017*, pubblicato dallo European Data Portal, il valore di mercato degli open data in Europa raggiungerà 75.7 miliardi di euro entro il

2020. Sempre secondo il rapporto, che analizza la situazione degli open data in Europa, l'Italia è passata in un anno dal gruppo degli Stati cosiddetti *followers* a quello dei *trend-setters*, saltando i *fast tracker* e collocandosi così nel gruppo degli Stati più virtuosi in ambito open data al settimo posto (nel 2016 si trovava in ventesima posizione). La figura 2.15 mostra il grafico rappresentante la classifica degli Stati europei in ambito open data e i relativi gruppi di appartenenza [42].

Fonte	Dataset
European Data Portal	832.558
Data.gov	302.925
Open Data Hub Italia	49.379
Dati.gov.it	22.476
European Union Open Data Portal	12.595
DataHub	11.314
DatiOpen.it	3.482
Comune di Bologna	2.589
Open Data Emilia-Romagna	1.192
Open Data Arpae Emilia-Romagna	23
Opendata beta, University of Bologna	16

Tabella 2.2: Numero di dataset disponibili nelle principali fonti di open data

L'apertura dei dati non determina solo effetti positivi sulle amministrazioni pubbliche di molti Stati nel mondo, ma produce anche rilevanti impatti sull'economia complessiva. Gli open data portano vantaggi al sistema economico, ad esempio utilizzando dati geospaziali, sui trasporti, sui beni culturali, demografici, immobiliari, occupazionali e sociali per produrre applicazioni e servizi, oltre che per generare conoscenza. Gli open data, come ho avuto modo di spiegare, possiedono un valore intrinseco, anche quando il singolo dato è poco utile, se combinato con altri dati provenienti da fonti diverse può generare un notevole valore aggiunto. In questo modo, ad esempio, le informazioni geografiche sono alla base delle mappe, le quali possono servire a fornire servizi online di ogni tipo, mentre

i dati immobiliari possono servire agli operatori che vogliono fare incontrare la domanda e l'offerta di abitazioni. I dati meteorologici in tempo reale possono essere utilizzati sia per applicazioni mobili che per alimentare i siti web meteo tradizionali. Stesso discorso anche per quanto riguarda i dati sui trasporti che alimentano servizi sulla mobilità. O ancora i dati sui bilanci delle imprese possono offrire importanti informazioni di mercato per fare analisi della concorrenza e studi economici. Gli open data possono quindi trasformarsi in contenuti generatori di nuove attività e servizi a valore aggiunto per gli utenti, rappresentati da cittadini, amministrazioni e imprese.

Dataset	Visualizzazioni	Download
CORDIS - EU research projects under Horizon 2020	33.775	39.693
Memoria di traduzione della DGT	26.268	30.895
TED - public procurement notices	12.042	17.941
CORDIS - reference data	9.873	13.776
List of entities subject to EU financial sanctions	24.278	10.207
CORDIS - EU research projects under FP7	18.262	8.755
Transparency Register	15.982	8.730
TED - avvisi di appalti pubblici da UE e mondo	43.126	7.722
EuroVoc - il thesaurus multilingue dell'UE	18.873	6.033
DORIE - ricerca sulle questioni istituzionali UE	5.593	4.810

Tabella 2.3: Top 10 dataset scaricati dallo European Union Open Data Portal

Anche i linked open data sono in espansione e il cosiddetto *Linked Open Data Cloud*, cioè il diagramma che mostra i collegamenti tra i vari dataset, diventa sempre più vasto. Al centro di tale diagramma, rappresentato nella figura 2.16 e aggiornato ad agosto 2018, si posiziona DBpedia, che è connessa ad altri dataset attraverso oltre 50 milioni di collegamenti RDF, a loro volta interconnessi ad altri linked open data [43].

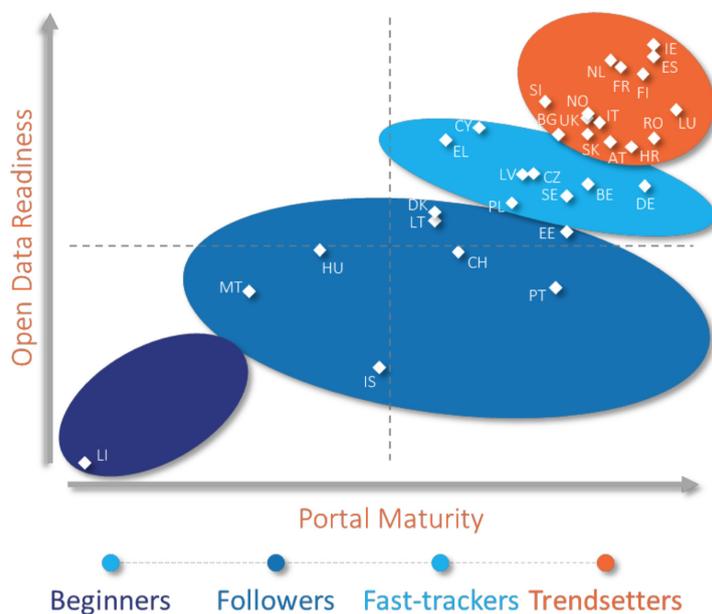


Figura 2.15: Classifica degli Stati europei in ambito open data

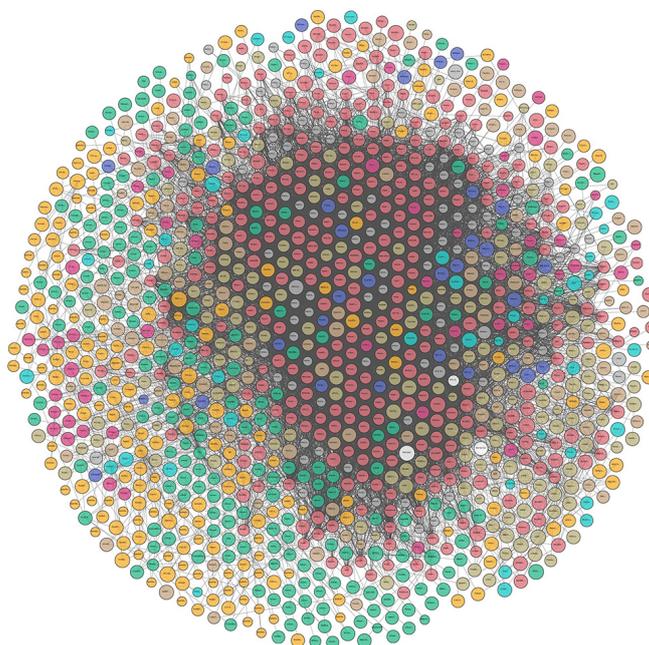


Figura 2.16: Il Linked Open Data Cloud ad agosto 2018

## 2.9 Esempi di applicazione degli open data

Ricerca, siti web, applicazioni per smartphone, tablet e computer, ma anche infografiche, sono solo alcuni degli esempi di utilizzo degli open data. Per capire in che ambiti e progetti possono essere applicati, è giusto conoscere alcuni esempi virtuosi nei quali sono stati utilizzati con successo per creare servizi innovativi.

- **OpenStreetMap:** nell'ambito dei dati cartografici l'utilizzo degli open data ha dato vita a progetti importantissimi come *openstreetmap.org*, costruito da una comunità di mappatori che contribuiscono e mantengono i dati su strade, sentieri, e punti di interesse di ogni tipo in tutto il mondo. Inoltre, come descritto in precedenza, OpenStreetMap è a sua volta una fonte di open data.
- **Google Traduttore:** anche un colosso del web come Google utilizza l'enorme volume di documenti in formato open dell'Unione Europea, disponibili in tutte le lingue del continente, per allenare gli algoritmi di traduzione automatica. In questo modo *translate.google.it* aumenta costantemente la precisione del servizio offerto agli utenti.
- **Where does my money go:** si tratta di un sito web, raggiungibile all'indirizzo *app.wheredoesmymoneygo.org*, che opera nell'ambito della trasparenza nel Regno Unito. Il servizio permette di capire in tempo reale come è impiegato dal governo il denaro che i cittadini spendono in tasse.
- **OpenParlamento:** si tratta di un sito web, raggiungibile all'indirizzo *openparlamento.it*, che monitora le attività del parlamento italiano e il processo di formazione delle leggi. In modo da mostrare ai cittadini cosa succede esattamente e quali parlamentari sono coinvolti nelle varie attività.
- **Mapumental:** è un servizio innovativo, raggiungibile tramite il sito web *mapumental.com*, che permette agli utenti di capire dove è possibile andare ad abitare impostando i tempi massimi di percorrenza tra la casa e il posto di lavoro. Inoltre si può anche scegliere in base ai prezzi delle case e alla bellezza del quartiere.

- **Husets Web:** in Danimarca, il sito web *husetsweb.dk* aiuta gli utenti a trovare i modi migliori per risparmiare energia elettrica in casa, inclusa la pianificazione finanziaria e la possibilità di contattare gli artigiani che potranno eseguire il lavoro. Funziona grazie al riutilizzo di dati catastali, informazioni sugli incentivi governativi e al registro delle imprese locali.
- **Safecast:** si tratta di una organizzazione internazionale di volontari, con lo scopo di raccogliere informazioni e dati relativi al livello di radioattività dell'ambiente. Dal sito web *safecast.org/tilemap* è possibile visualizzare la mappa del pianeta che mostra in che misura un territorio è radioattivo, inoltre tutti i dati pubblicati possono essere riutilizzati da chiunque.
- **Air Quality in Scotland:** è un portale web, raggiungibile all'indirizzo *scottishairquality.co.uk*, che permette agli utenti di monitorare costantemente la qualità dell'aria in Scozia. Il sistema prende in considerazione diversi agenti inquinanti e avvisa gli utenti registrati se i valori superano determinati limiti.
- **Accessible UK Train Timetables:** si tratta di un portale web, raggiungibile all'indirizzo *traintimes.org.uk*, che permette agli utenti di consultare l'orario dei treni della metropolitana di Londra e il prezzo del biglietto. Inoltre è possibile visualizzare su una mappa interattiva la posizione dei treni in tempo reale.
- **VeloBÒ:** si tratta di un'applicazione mobile per smartphone e tablet con sistema operativo Android. L'app utilizza gli open data pubblicati dal Comune di Bologna e mostra le piste ciclabili distribuite nell'area urbana della città, oltre alle postazioni per il noleggio delle biciclette.
- **Comune di Bologna:** utilizzando gli open data pubblicati dal Comune di Bologna sono state realizzate decine di infografiche e mappe interattive che mostrano ad esempio dove sono localizzati distributori di carburante, hotspot Wi-Fi, bagni pubblici, aree pedonali, parchimetri, colonnine di ricarica per veicoli elettrici e molto altro. Tutti i progetti sono raccolti nella sezione *Casi d'uso* del sito web *dati.comune.bologna.it*.

Numerosi esempi di come le aziende italiane utilizzano gli open data nelle loro attività per generare prodotti e servizi si possono trovare sul portale *Open Data 200*, raggiungibile all'indirizzo *italy.opendata500.com* e sviluppato dal *GovLab - New York University* in collaborazione con la *Fondazione Bruno Kessler*.



Figura 2.17: L'*Open Data Button* indica i progetti che utilizzano gli open data

Come si è potuto osservare sono molti gli esempi pratici in cui gli open data stanno già creando vantaggi economici e sociali, e non è possibile prevedere quali saranno gli utilizzi innovativi che arriveranno in futuro. Il *mesh-up* dei dati, cioè l'integrazione dinamica delle informazioni provenienti da più fonti, potrà portare a campi di applicazione attualmente inimmaginabili. Il potenziale degli open data potrà essere sfruttato appieno se pubbliche amministrazioni, organizzazioni e aziende pubblicheranno i dati in formato aperto e senza limitazioni di utilizzo degli stessi.





## Parte II

# Progettazione ed implementazione di open data in un portale web turistico

*"Siate affamati. Siate folli."*

Steve Jobs



# Capitolo 3

## Progettazione

Nei primi due capitoli si è approfondito il concetto di *dati* e sottolineata la loro importanza nella società, per poi descrivere i principi e le caratteristiche degli *open data*. L'obiettivo di questo capitolo è di esporre un caso pratico di progettazione e successiva implementazione di open data all'interno del portale web turistico *OkBed&Breakfast*. Ho svolto il lavoro descritto durante un tirocinio della durata di due mesi presso l'agenzia web *Magic* di Funo di Argelato (Bologna), tra aprile e giugno 2018.

### 3.1 Panoramica su OkBed&Breakfast

Il portale web OkBed&Breakfast, raggiungibile all'indirizzo *okbedandbreakfast.it*, è nato nel 2009 come progetto interno dell'azienda Magic. Si tratta di un sito web attivo su tutto il territorio italiano il cui scopo è quello di mettere in contatto turisti e proprietari di bed and breakfast, in modo da fare incontrare domanda e offerta, creando visibilità per i primi e soluzioni per i secondi. OkBed&Breakfast è stato sviluppato da Magic nel corso di nove anni, durante i quali è stato costantemente migliorato e arricchito grazie al contributo di diverse figure professionali che hanno partecipato al progetto. Il team si è occupato di programmazione, contenuti e grafica lavorando in modo collaborativo e sinergico per realizzare OkBed&Breakfast con tecnologie di ultima generazione. Il progetto è stato sviluppato tecnicamente utilizzando il CMS open source Drupal, oltre ai linguaggi di programmazione HTML, CSS, JavaScript e PHP.

La piattaforma ospita oltre 15.300 strutture tra cui l'utente può scegliere, inizialmente le liste dei bed and breakfast sono state acquistate da *registroimprese.it*, che offre i dati ufficiali delle camere di commercio.



Figura 3.1: Logo di OkBed&Breakfast

Come accennato precedentemente, OkBed&Breakfast si rivolge a due categorie di utenti, che utilizzano il portale in modo diverso: viaggiatori e proprietari.

- **Viaggiatori:** gli utenti viaggiatori possono utilizzare il sito web per cercare il bed and breakfast più adatto alle loro esigenze consultando un ampio catalogo di strutture, inoltre registrandosi gratuitamente al portale possono usufruire di tutti i servizi. È possibile cercare le strutture inserendo il nome della destinazione nell'apposita barra di ricerca in home page anche selezionando particolari filtri di ricerca. Oppure si possono selezionare regione, provincia, comune o tipo di luogo da visitare. L'utente, dopo aver trovato le strutture di suo interesse e confrontato prezzi e caratteristiche, può richiedere la disponibilità al proprietario del bed and breakfast compilando un apposito form di contatto nel quale si indicano i giorni di permanenza e il numero degli ospiti. Inoltre i viaggiatori possono scrivere delle recensioni sulle strutture in cui hanno soggiornato.
- **Proprietari:** gli utenti proprietari di una struttura possono registrarsi gratuitamente al portale per aggiungere il proprio bed and breakfast alla piattaforma. Nel

caso in cui la struttura sia già presente su OkBed&Breakfast, poiché negli elenchi della camera di commercio, il proprietario può rivendicarne il possesso tramite un apposito processo di verifica. Il proprietario può inoltre gestire e personalizzare la pagina della sua struttura inserendo tutte le informazioni e completandole con i servizi disponibili, le caratteristiche delle camere, i prezzi e le immagini, oltre a poter rispondere alle recensioni lasciate dai clienti.

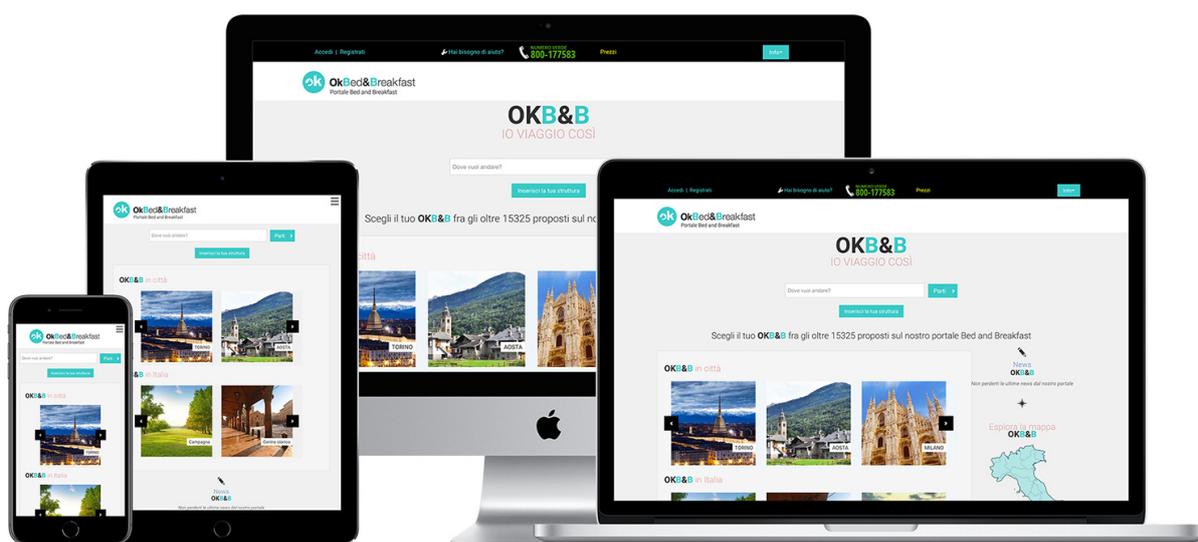


Figura 3.2: Home page di OkBed&Breakfast su computer, tablet e smartphone

L'utilizzo della piattaforma da parte dei viaggiatori è totalmente gratuito, mentre i proprietari possono gestire la prima struttura gratuitamente ma con diverse limitazioni utilizzando il piano *Free*. Per usufruire di tutte le funzionalità aggiuntive e gestire più di una struttura possono sottoscrivere un abbonamento a pagamento ai piani *Premium* o *Pro*. Inizialmente OkBed&Breakfast si è sostenuto solo tramite advertising, con l'inserimento di banner pubblicitari nelle pagine del portale. Successivamente sono stati rimossi in favore degli abbonamenti a pagamento appena descritti. Tramite OkBed&Breakfast non è possibile prenotare direttamente un soggiorno in una struttura, dato che l'obiettivo del progetto è di rendere possibile l'incontro tra domanda e offerta. Come descritto, sarà l'utente a richiedere la disponibilità di una camera al proprietario della struttura per

poi accordarsi con esso, in questo modo, a differenza di altri portali, gli utenti potranno usufruire dei prezzi originali poiché OkBed&Breakfast non trattiene alcuna provvigione né dopo il contatto, né dopo la prenotazione.

Altre funzionalità di OkBed&Breakfast sono la ricerca di luoghi da visitare in base ai piatti tipici e la pubblicazione di notizie relative ad eventi di ogni genere in tutta Italia, per suggerire all'utente quali località visitare. Inoltre sono presenti un servizio di messaggistica interno al portale e una newsletter per gli utenti registrati. OkBed&Breakfast è anche presente sui principali social network, dove vengono pubblicati aggiornamenti relativi al portale e alle strutture presenti nel catalogo [44].

## 3.2 Architettura e specifiche del progetto

Lo scopo del progetto svolto durante il tirocinio è stato di implementare gli open data all'interno di OkBed&Breakfast. Per raggiungere questo obiettivo, dopo una prima fase di studio sugli open data e sul loro utilizzo, è iniziata la fase progettuale, nella quale si è deciso insieme al team dell'azienda quali dati utilizzare, da dove recuperarli, dove implementarli e in che modo.

Innanzitutto sono state individuate le tipologie di dati più utili da mostrare all'utente che sta navigando sul sito web per trovare il bed and breakfast adatto alle sue esigenze. Inoltre è stato importante capire quale può essere l'utenza di OkBed&Breakfast, per mostrare dati che si avvicinino il più possibile alle necessità di pubblici diversi, ad esempio utenti giovani che cercano divertimento e utenti anziani che vogliono fare turismo. In particolare, si è scelto di implementare gli open data di OpenStreetMap e dell'Istat nelle pagine dei singoli bed and breakfast presenti sul portale e nelle pagine dedicate a comuni, province e luoghi. Inoltre si è sempre preferito utilizzare gli open data tramite API, piuttosto che eseguire il download dei dati. In questo modo gli open data mostrati sono sempre i più aggiornati che la fonte mette a disposizione e vengono richiesti dinamicamente senza occupare lo storage offerto dall'*hosting*.

Un requisito fondamentale è che tutte le nuove funzionalità introdotte si comportassero correttamente sia quando il portale web è utilizzato da desktop che quando è utilizzato da mobile. Di conseguenza sono state eseguite costantemente operazioni di *debugging* e *testing* delle nuove funzioni nei sistemi operativi (Windows, macOS, Android e iOS) e browser (Chrome, Safari, Firefox, Edge, Internet Explorer e Opera) principali.

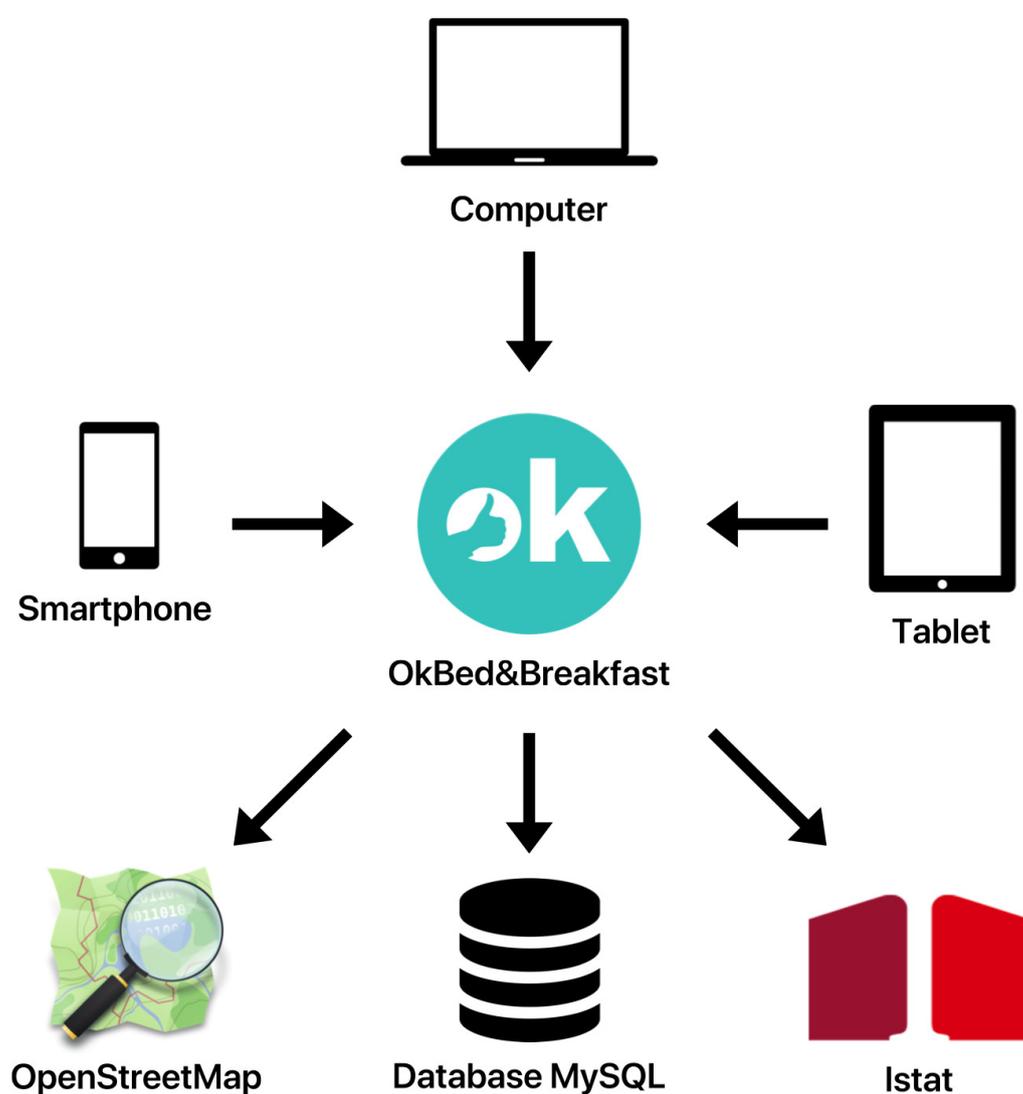


Figura 3.3: Architettura del progetto

La figura 3.3 mostra lo schema che rappresenta l'architettura del progetto. Come si può osservare, i vari client (computer, tablet e smartphone) accedono ad OkBed&Breakfast navigando tra le sue pagine e inviando richieste. Il server che ospita il portale web risponde alle domande dei client tramite il CMS open source Drupal, in modo che essi possano visualizzare i contenuti richiesti. OkBed&Breakfast invia a sua volta richieste al database MySQL sul quale sono memorizzati i contenuti del portale, ad esempio le coordinate geografiche di tutte le strutture. Inoltre, tramite chiamate AJAX, invia richieste a OpenStreetMap e all'Istat, le due fonti scelte per il recupero degli open data. Nel primo caso le richieste avvengono utilizzando le Overpass API per il recupero degli open data dei punti di interesse e tramite OpenLayers per la costruzione delle mappe interattive. Nel secondo caso invece, le richieste avvengono mediante l'endpoint SPARQL del portale Linked Open Data di Istat. Ogni volta che OpenStreetMap, Istat e il database MySQL rispondono alle richieste lato client (tramite JavaScript) o lato server (tramite PHP) a seconda della complessità, questi ricevono le informazioni ad esempio in formato JSON, che vengono elaborate per visualizzare sullo schermo degli utenti i risultati (come mappe interattive, punti di interesse o semplice testo).

### 3.2.1 Pagine dei bed and breakfast

In questo tipo di pagine vengono mostrate informazioni, servizi disponibili, prezzi e immagini relative al bed and breakfast che l'utente ha selezionato. Inoltre, l'utente può richiedere la disponibilità al proprietario della struttura e scrivere una recensione. In fondo alla pagina è stata creata un'apposita sezione nella quale sono visualizzate le informazioni recuperate tramite gli open data, con le quali l'utente può interagire. In particolare sono disponibili tre pannelli diversi:

- **Pannello luogo:** il primo pannello, chiamato *Luogo*, mostra tramite un marker fisso giallo la posizione del bed and breakfast su una mappa di OpenStreetMap, creata tramite la libreria OpenLayers. In questo modo l'utente può capire subito dove si trova esattamente la struttura. I dati di latitudine e longitudine di ogni struttura sono stati recuperati dal dataset acquistato dalla camera di commercio, oppure inseriti dal proprietario della struttura nel caso in cui non fosse presente.

- **Pannello nelle vicinanze:** il secondo pannello, chiamato *Nelle vicinanze*, mostra tramite un marker fisso giallo la posizione del bed and breakfast su una mappa di OpenStreetMap, creata tramite la libreria OpenLayers. Inoltre, l'utente può selezionare i punti di interesse da mostrare sulla mappa in base ai propri interessi, per capire dove si trovano rispetto alla struttura. Sono disponibili 60 tipologie diverse di punti di interesse tra i quali l'utente può scegliere, suddivisi in 6 categorie di 10 punti di interesse ciascuna: turismo, cibo, divertimento, trasporti, servizi e negozi. L'utente può premere sui titoli delle 6 categorie per aprire il relativo elenco di punti di interesse, realizzato tramite l'elemento *collapse* della libreria *Bootstrap*, gestito in modalità *accordion* che permette un solo elenco aperto alla volta.

Quando l'utente preme sulle *checkbox* dei punti di interesse che vuole visualizzare, verranno invocate delle chiamate in AJAX che richiedono gli open data di OpenStreetMap (disponibili nel raggio di 30 Km dalla struttura) tramite le Overpass API, costruendo dinamicamente le query. Viene letto il file JSON ricevuto come risposta, e infine vengono creati i marker e i relativi *popup* da visualizzare sulla mappa utilizzando JavaScript e OpenLayers. I marker che l'utente ha selezionato appariranno dopo qualche secondo di attesa, sottolineato dalla comparsa della rotellina di caricamento al centro dello schermo. I marker visualizzati saranno rappresentati da specifiche immagini in base al tipo di dato scelto, inoltre al *mouseover* sui punti di interesse comparirà un popup con il nome del marker selezionato (per i punti di interesse nei quali è presente l'informazione).

Al click per selezionare la checkbox verrà mostrato anche il numero di punti di interesse caricati per quella categoria, mentre sopra alla mappa verrà aggiornato il numero totale dei punti visualizzati. Quando invece l'utente deselecta la checkbox, i punti scompariranno dalla mappa e il numero totale dei punti verrà aggiornato di conseguenza. Inoltre è anche presente il bottone *Pulisci la mappa* che se premuto riporta la mappa alla configurazione iniziale. Quindi vengono rimossi tutti i punti di interesse selezionati, chiuso il collapse eventualmente aperto e reimpostato il centro e lo zoom della mappa.

- **Pannello distanze:** il terzo pannello, chiamato *Distanze*, mostra quanto dista in linea d'aria il bed and breakfast da un altro luogo, in particolare, dal centro cittadino, dalla stazione ferroviaria/metro più vicina, dall'aeroporto più vicino e dall'ospedale più vicino. Quando l'utente preme sul pannello viene invocata una chiamata in AJAX che calcola in un file PHP (quindi lato server, per non pesare sul client, specialmente se si sta navigando da dispositivi mobili) la distanza in linea d'aria tra la struttura e i punti di interesse nel raggio di 30 o 50 Km. Una volta eseguiti i calcoli, sottolineati dalla comparsa della rotellina di caricamento al centro dello schermo, l'utente visualizzerà quanto distano in chilometri i punti di interesse più vicini per ogni categoria elencata precedentemente. Ogni tipo di punto è rappresentato da un'immagine identificativa e dove presente è mostrato anche il nome del punto di interesse. Per il recupero delle coordinate del centro cittadino si è utilizzata la Nominatim API di OpenStreetMap, mentre per il recupero degli altri punti si è utilizzata la Overpass API.

### 3.2.2 Pagine di comuni, province e luoghi

In questo tipo di pagine vengono mostrate le informazioni e la mappa del comune, provincia o luogo (come campagna, centro storico, mare e montagna) che l'utente ha selezionato, con l'elenco delle relative strutture. In cima alla pagina è stata creata un'apposita sezione nella quale sono visualizzate le informazioni recuperate tramite gli open data e l'utente può interagire con esse. In particolare, in tutte e tre le tipologie di pagine è mostrata una mappa di OpenStreetMap, creata tramite la libreria OpenLayers, che mostra l'area geografica selezionata.

Sulla mappa vengono visualizzati, anche tramite cluster usando la libreria JavaScript *AnimatedCluster*, i marker delle strutture, recuperando le coordinate degli stessi dal database MySQL (anche con eventuali filtri di ricerca applicati). I marker delle strutture senza prezzo sono rappresentati da un'icona bianca, mentre quelle con il prezzo da un'icona rossa e una *label* fissa che mostra il costo minimo. Quando l'utente preme sui marker, essi diventano verdi e posizionati al centro della mappa, così è più semplice capire quali sono i bed and breakfast già visualizzati. Inoltre sono stati aggiunti i popup al

click sui marker con eventuale nome, immagine, prezzo e link della struttura selezionata. L'utente può interagire con la mappa, premendo sui cluster verrà eseguito uno *zoom-in*, il centro della mappa verrà aggiornato e i cluster verranno ricalcolati dinamicamente e riposizionati con un'animazione. Nelle pagine di comuni e province, se non sono presenti nel database le coordinate per settare il centro della mappa, queste vengono recuperate tramite la Nominatim API di OpenStreetMap. Nelle pagine dei luoghi invece, le coordinate per centrare la mappa sono impostate di *default* al centro dell'Italia poiché viene sempre visualizzata la mappa di tutto il territorio nazionale.

Inizialmente, nelle pagine relative ai comuni sono stati aggiunti tre bottoni, posizionati appena sopra alla mappa descritta precedentemente, per richiedere gli open data forniti dall'Istat relativi a popolazione, famiglie e alloggi presenti nel comune selezionato. Quando l'utente preme su uno dei tre bottoni, verrà invocata la relativa chiamata in AJAX che calcola in un file PHP i dati richiesti. Il calcolo avviene quindi lato server, per non pesare sul client, ma soprattutto a causa dell'attuale mancanza di supporto al protocollo HTTPS sul portale Linked Open Data di Istat. Questo limite bloccherebbe la richiesta HTTP se fatta direttamente da AJAX e JavaScript poiché OkBed&Breakfast supporta il protocollo HTTPS. Purtroppo le richieste fatte al portale Linked Open Data di Istat sono molto lente da soddisfare, poiché la piattaforma è ancora in versione sperimentale. Di conseguenza, oltre alla comparsa della rotellina di caricamento al centro dello schermo, compariva anche un messaggio per avvisare l'utente dell'attesa. Gli open data forniti da Istat sono stati rimossi successivamente proprio a causa dell'eccessiva lentezza nel caricamento (almeno 30 secondi di attesa per ogni richiesta), in fase di applicazione delle modifiche dalla versione *dev* alla versione *live* di OkBed&Breakfast.

### 3.3 Open data utilizzati

Come descritto precedentemente, gli open data che si è scelto di utilizzare sono stati quelli forniti da OpenStreetMap e dal portale Linked Open Data di Istat. Durante la fase iniziale di analisi della situazione attuale sugli open data e progettazione del lavoro da svolgere si è riscontrato un fenomeno determinante. Poiché OkBed&Breakfast

opera a livello nazionale, è stato necessario trovare open data disponibili in tutta Italia e non solo a livello locale. Di conseguenza si è scelto di utilizzare, nel modo descritto precedentemente, gli open data pubblicati da due importanti fonti come OpenStreetMap e Istat, poiché relativi a tutto il territorio nazionale. In particolare, gli open data forniti da OpenStreetMap sono stati utilizzati per:

- Realizzare le mappe interattive tramite la libreria OpenLayers nelle pagine dei bed and breakfast e in quelle di comuni, province e luoghi.
- Visualizzare tramite marker sulla mappa interattiva i punti di interesse scelti dall'utente nelle pagine dei bed and breakfast e calcolarne il numero, visualizzando il risultato tramite contenuto testuale.
- Calcolare le distanze minime tra la struttura e alcuni punti di interesse nelle pagine dei bed and breakfast. In particolare centro cittadino, stazione ferroviaria/metro, aeroporto e ospedale, visualizzando il risultato tramite contenuto testuale.

Invece gli open data forniti dal portale Linked Open Data di Istat, rimossi nella fase finale a causa dell'eccessiva lentezza nel caricamento, sono stati utilizzati per:

- Calcolare alcuni indicatori statistici relativi al comune selezionato nelle pagine dei comuni. In particolare popolazione, numero di famiglie e numero di alloggi presenti, visualizzando il risultato tramite contenuto testuale.

Di seguito sono elencate nel dettaglio le tipologie di open data che l'utente può selezionare, per visualizzare i punti di interesse sulla mappa nel pannello *Nelle vicinanze* all'interno delle pagine dei bed and breakfast (tutti i dati sono recuperati nel raggio di 30 Km dalla struttura):

- **Turismo:** informazioni turistiche, attrazioni turistiche, monumenti, siti archeologici, castelli, musei, punti panoramici, parchi naturali, zoo, acquari.
- **Cibo:** ristoranti, pizzerie, fast food, pub, birrerie, bar, gelaterie, pasticcerie, aree picnic, fontane potabili.
- **Divertimento:** cinema, teatri, stadi, discoteche, casinò, stabilimenti balneari, piscine, parchi tematici, parchi acquatici, parchi giochi.

- **Trasporti:** parcheggi, posteggi taxi, fermate autobus, stazioni autobus, stazioni ferroviarie/metro, aeroporti, porti, noleggio auto, noleggio biciclette, noleggio imbarcazioni.
- **Servizi:** banche, chiese, uffici postali, ospedali, veterinari, parrucchiere, centri sportivi, bagni pubblici, parchi pubblici, aree di sgambamento.
- **Negozi:** centri commerciali, supermercati, negozi di abbigliamento, negozi di articoli sportivi, negozi di elettronica, negozi di articoli per animali, librerie, lavanderie, tabaccherie, farmacie.

### 3.4 Scelte progettuali

Per implementare gli open data di OpenStreetMap e Istat su OkBed&Breakfast, sono state intraprese varie scelte progettuali, evolute mano a mano che si definivano possibilità di sviluppo, esigenze aziendali e limitazioni tecniche esistenti:

- Come accennato precedentemente OkBed&Breakfast opera a livello nazionale, ma alcune tipologie di open data che sarebbero potute essere utili creando un valore aggiunto (come ad esempio i dati ambientali pubblicati da Arpa), sono disponibili soltanto a livello locale per alcuni comuni, province o regioni. Per questo motivo una soluzione poteva essere quella di utilizzare questo tipo di open data solo nei luoghi dove sono presenti, ad esempio nei comuni principali. Così facendo però si sarebbe persa omogeneità tra i dati mostrati nelle varie pagine, per questo motivo si è scelto di usare dati disponibili in tutti i luoghi e non solo in alcuni. Tipologie di open data che rientrano in questa categoria, cioè disponibili solo localmente, sono ad esempio i dati relativi a meteo, clima e temperatura media del luogo, solitamente pubblicati in alcune regioni sui portali web delle agenzie Arpa.
- I limiti di query eseguibili verso OpenStreetMap non sono precisi, ma è solo sconsigliato eseguirne un numero elevato, inoltre non è sempre garantita una risposta, dato che vengono utilizzati server donati e il servizio è open e gratuito. Poiché gli open data offerti da OpenStreetMap sono utilizzati in tutte le pagine dedicate ai

bed and breakfast e nelle pagine di comuni, province e luoghi è previsto un utilizzo piuttosto intenso di tali dati, specialmente per la creazione delle mappe interattive. Di conseguenza, come suggerito anche da OpenStreetMap, inizialmente si era pensato di non utilizzare i loro server, ma di eseguire il download dell'insieme dei dati e importarli nel database MySQL installato sul server proprietario. In questo modo si sarebbero eseguite le interrogazioni direttamente da lì, con l'aggiunta di un sistema che aggiornasse i dati a intervalli regolari.

Ragionando su quale fosse la scelta migliore per l'implementazione degli open data, è sorto il primo problema. Dato che OkBed&Breakfast opera su tutto il territorio nazionale, sarebbero stati necessari i dati e la mappa di tutta Italia, di oltre 1.3 GB. Questo fattore avrebbe influito negativamente sull'occupazione di storage, sull'esecuzione dei backup automatici giornalieri e sull'utilizzo di risorse dedicate ad aggiornare periodicamente l'intero insieme di dati.

Successivamente, per questi motivi si è pensato di continuare ad utilizzare le mappe di Google Maps, che erano già usate in precedenza, tramite le Embed API (utilizzo gratuito e illimitato). Al posto dei punti di interesse offerti da Google sarebbero stati visualizzati sulla mappa quelli recuperati da OpenStreetMap, per la creazione dei marker, eseguendo il download dell'insieme dei dati. In questo caso però si sarebbe eseguito il download solo dei dati relativi ai punti di interesse necessari, e non della mappa italiana completa, in modo da non occupare troppo storage e risorse per l'aggiornamento dei dati stessi.

In seguito, si è scoperto che utilizzare gli open data di OpenStreetMap su MySQL è sconsigliato, poiché è suggerito l'utilizzo di PostgreSQL e non esistono soluzioni ufficiali per usare MySQL, oltre ad occupare comunque storage e risorse. Per questi motivi si è deciso di utilizzare le Overpass API. Queste API sono progettate proprio per recuperare i dati dal database di OpenStreetMap e rispondono con un file GeoJSON contenente tutti gli elementi trovati con le relative coordinate (latitudine e longitudine). Tali dati sarebbero poi stati visualizzati sulla mappa di Google

Maps, come deciso in precedenza. I limiti di utilizzo delle Overpass API non sono precisi, ma vengono garantite le 10.000 query al giorno, che sono sufficienti per OkBed&Breakfast.

Successivamente, dopo la comunicazione dei cambi delle politiche di Google (avvenuta il 3 maggio 2018), che introducevano limiti più stringenti e molte funzionalità a pagamento, si è deciso di abbandonare del tutto Google Maps. Di conseguenza si è optato per utilizzare interamente le mappe di OpenStreetMap sui loro server, create tramite la libreria OpenLayers e le Overpass API per il recupero dei punti di interesse. Nel caso in cui fossero state troppo lente nella visualizzazione (problema che si era posto inizialmente) si sarebbe proceduto eseguendo il download dell'intera mappa italiana, gestendo tutto sul server proprietario. Fortunatamente, sia in fase di sviluppo nella versione *dev* del sito web, che dopo aver applicato tutte le modifiche anche nella versione *live*, le mappe si caricavano ad una velocità adeguata. Di conseguenza non è stato necessario eseguire il download dell'intera mappa italiana, evitando quindi l'occupazione eccessiva di storage e il consumo di risorse per aggiornare periodicamente i dati.

- A causa dell'eccessiva lentezza di risposta (almeno 30 secondi a richiesta), in fase di pubblicazione delle modifiche dal sito web *dev* alla versione *live*, si è deciso di rimuovere l'implementazione degli open data forniti da Istat. Il codice è comunque rimasto commentato, nella speranza che in futuro il portale Linked Open Data di Istat, che ora è in versione sperimentale, diventi più veloce e reattivo di come è attualmente. Per risolvere il problema si era anche pensato di eseguire il download dell'insieme dei dati e importarli nel database del portale, per poi interrogarli direttamente da lì. Ma per evitare l'occupazione eccessiva di storage e risorse dedicate ad aggiornare periodicamente i dati, si è preferito utilizzare sempre gli open data tramite le API messe a disposizione (come avvenuto anche nel caso di OpenStreetMap tramite Overpass API e Nominatim API).
- Nel pannello *Distanze* delle pagine dedicate ai singoli bed and breakfast vengono calcolate le distanze della struttura da alcuni punti di interesse. Durante la fase

progettuale si sono compiute alcune scelte, per determinare quale fosse la soluzione migliore in grado di realizzare questa funzionalità. In particolare si sono considerate due alternative, la prima alternativa era il calcolo del percorso reale. Per questo scopo potevano essere usate le OSRM API, che calcolano il percorso tra due luoghi. In particolare la posizione del bed and breakfast sarebbe stata impostata di default come punto di partenza e l'utente avrebbe potuto selezionare un punto di arrivo premendo sulla mappa. Il problema è che si sarebbe dovuto necessariamente installare la libreria sul proprio server, poichè quello demo si può utilizzare solo per dei test. L'aspetto positivo è che si sarebbe ricevuta come risposta il file JSON con il percorso reale tra i due punti.

La seconda alternativa era il calcolo delle distanze in linea d'aria. In particolare, la posizione del bed and breakfast sarebbe stata impostata di default come punto di partenza e l'utente poteva selezionare un punto di arrivo premendo sulla mappa. In questo modo sarebbe stato possibile creare percorsi ma sempre in linea d'aria, eseguendo il doppio click sul punto di arrivo si poteva uscire dalla modalità di input. L'aspetto negativo è che si sarebbero calcolate distanze in linea d'aria e non percorsi reali tra due punti, l'aspetto positivo è che si sarebbe potuto realizzare tutto tramite OpenLayers senza installare nulla di aggiuntivo sul proprio server.

In conclusione si è optato per una versione modificata della seconda alternativa, quindi il calcolo delle distanze in linea d'aria. Si recuperano i punti di destinazione tramite Overpass API e Nominatim API, si calcola la distanza minima dalla struttura selezionata e si visualizzano direttamente i risultati senza che l'utente debba interagire con la mappa creando autonomamente il percorso.

- Nei tre pannelli presenti delle pagine dedicate ai bed and breakfast, si è scelto di utilizzare le immagini scaricate gratuitamente dal sito web *Map Icons Collection*, raggiungibile all'indirizzo *mapicons.mapsmarker.com*, per rappresentare i vari tipi di punti di interesse. Quindi si è scelto di non utilizzare le immagini fornite da OpenStreetMap o di creare immagini personalizzate.

### 3.4.1 Download o API: vantaggi e svantaggi

Esistono dei vantaggi e degli svantaggi nella scelta di quale metodologia utilizzare per il recupero degli open data: download o API. Solitamente in tutte le fonti di open data è presente la possibilità di eseguire il download degli stessi, mentre l'accesso tramite API non è sempre permesso, poiché più costoso da mantenere. Se presenti entrambe le metodologie è importante considerare diversi fattori per fare la scelta migliore, anche in base al tipo di dati che si vogliono utilizzare. Durante le fase di progettazione si sono considerati i seguenti vantaggi e svantaggi, preferendo il recupero di tutti gli open data tramite API [17] [45].

- **Vantaggi nel download di dataset:** eseguendo il download dell'intero dataset e importandolo nel proprio database, si evitano gli eventuali limiti imposti dalla fonte che pubblica gli open data (ad esempio il massimo numero di query giornaliera che possono essere eseguite). Inoltre la velocità di risposta è direttamente proporzionale alle prestazioni del database che si utilizza e dell'eventuale server su cui risiede. Si hanno quindi più possibilità di controllo degli aspetti tecnici.
- **Svantaggi nel download di dataset:** eseguendo il download dell'intero dataset e importandolo nel proprio database, si incide sullo storage occupato, con tutti i problemi che ne conseguono. Le dimensioni di un dataset possono infatti andare da pochi KB a parecchi GB. Altro aspetto negativo riguarda l'aggiornamento dei dati, poiché eseguendo il download dell'intero dataset non avremo sempre l'ultima versione disponibile. Per aggirare questo problema si può prevedere un download automatico dell'intero dataset a intervalli regolari (ad esempio ogni settimana o ogni mese). La gestione di tale meccanismo è però più complessa e costosa in termini di risorse rispetto all'uso delle API.
- **Vantaggi nell'utilizzo di API:** l'intero dataset si trova sul server dell'ente che mette a disposizione gli open data, di conseguenza non viene occupato ulteriore spazio sul nostro server proprietario o altro tipo di archiviazione. Inoltre i dati prelevati sono sempre aggiornati alla versione più recente disponibile. Risulta comodo eseguire direttamente query specifiche sul dataset di interesse per recuperare solo gli open data di cui si ha davvero bisogno.

- **Svantaggi nell'utilizzo di API:** spesso l'utilizzo di questo tipo di servizi è regolato da limiti imposti dalla fonte che pubblica gli open data. Possono essere presenti limiti sul numero di query eseguibili (ad esempio ogni giorno o ogni ora), sulla quantità di dati scambiati (in termini di volume di traffico) oppure limiti temporali tra una richiesta e la successiva (ad esempio una richiesta ogni minuto). Spesso per utilizzare le API sono necessarie maggiori competenze tecniche. Inoltre è presente un problema riguardante il tempo di risposta necessario per soddisfare ogni richiesta, che può variare in base alla complessità di quest'ultima e alla potenza dell'API. Tali limiti possono essere più o meno stringenti, a discrezione della fonte di open data, e sono presenti poiché essendo tutto gratuito l'ente non garantisce di soddisfare sempre e comunque qualsiasi richiesta. Per questo motivo, in alcuni casi è consigliato dalla fonte stessa eseguire il download dei dataset, proprio per non gravare sui loro server (come nel caso di OpenStreetMap).

### 3.5 Screenshot

Di seguito sono raccolti una serie di screenshot che mostrano il funzionamento e il modo in cui sono visualizzati gli open data all'interno di OkBed&Breakfast:

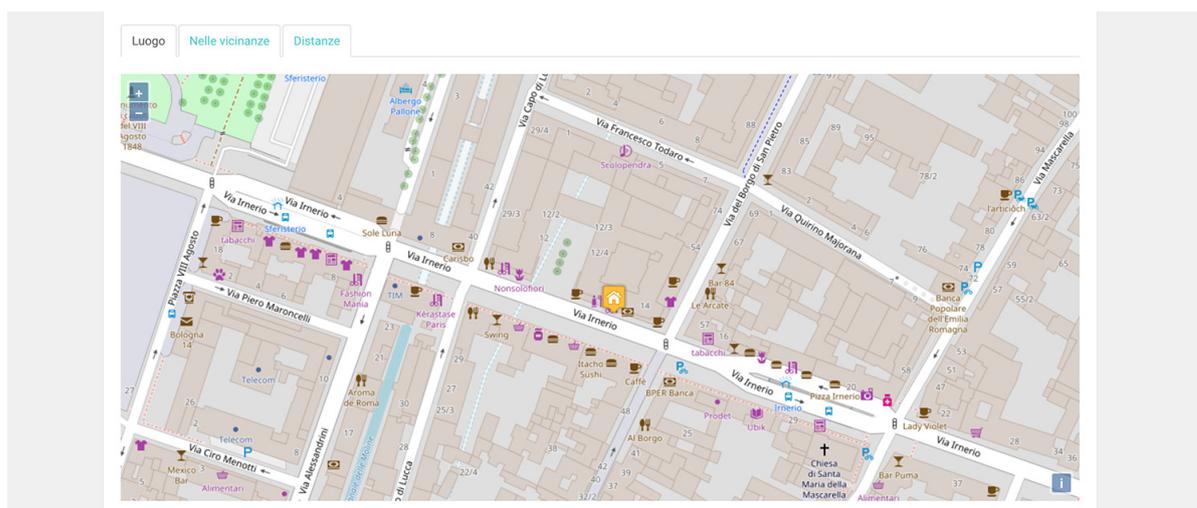


Figura 3.4: Pagine dei bed and breakfast, pannello *Luogo*



Figura 3.5: Pagine dei bed and breakfast, pannello *Nelle vicinanze*

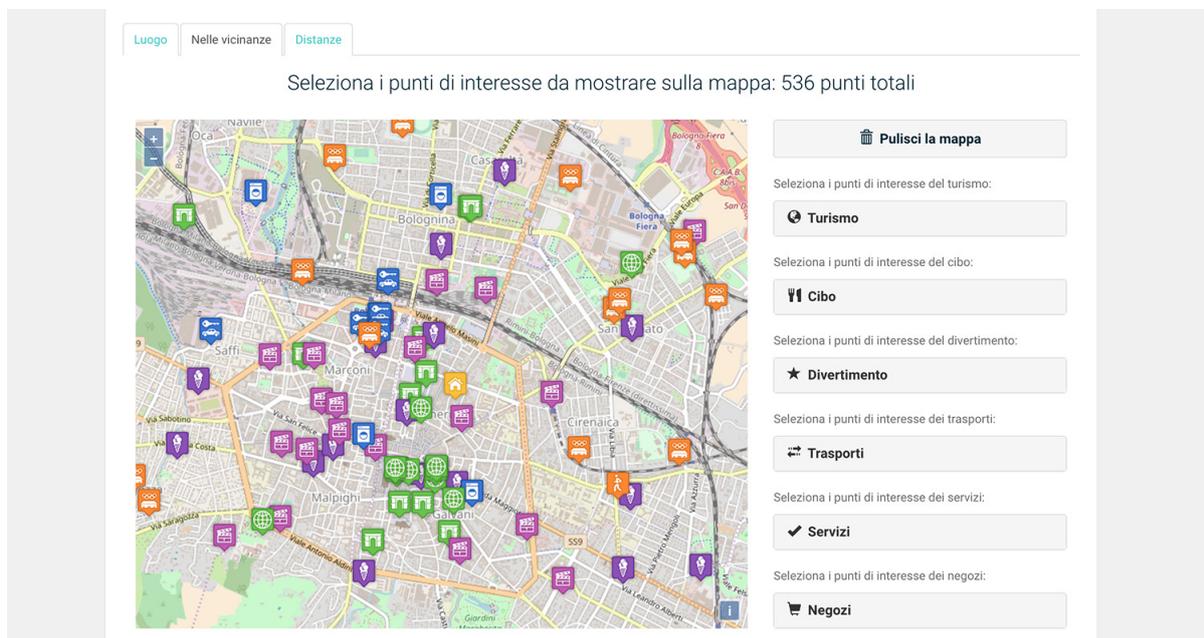


Figura 3.6: Pagine dei bed and breakfast, pannello *Nelle vicinanze* e punti selezionati

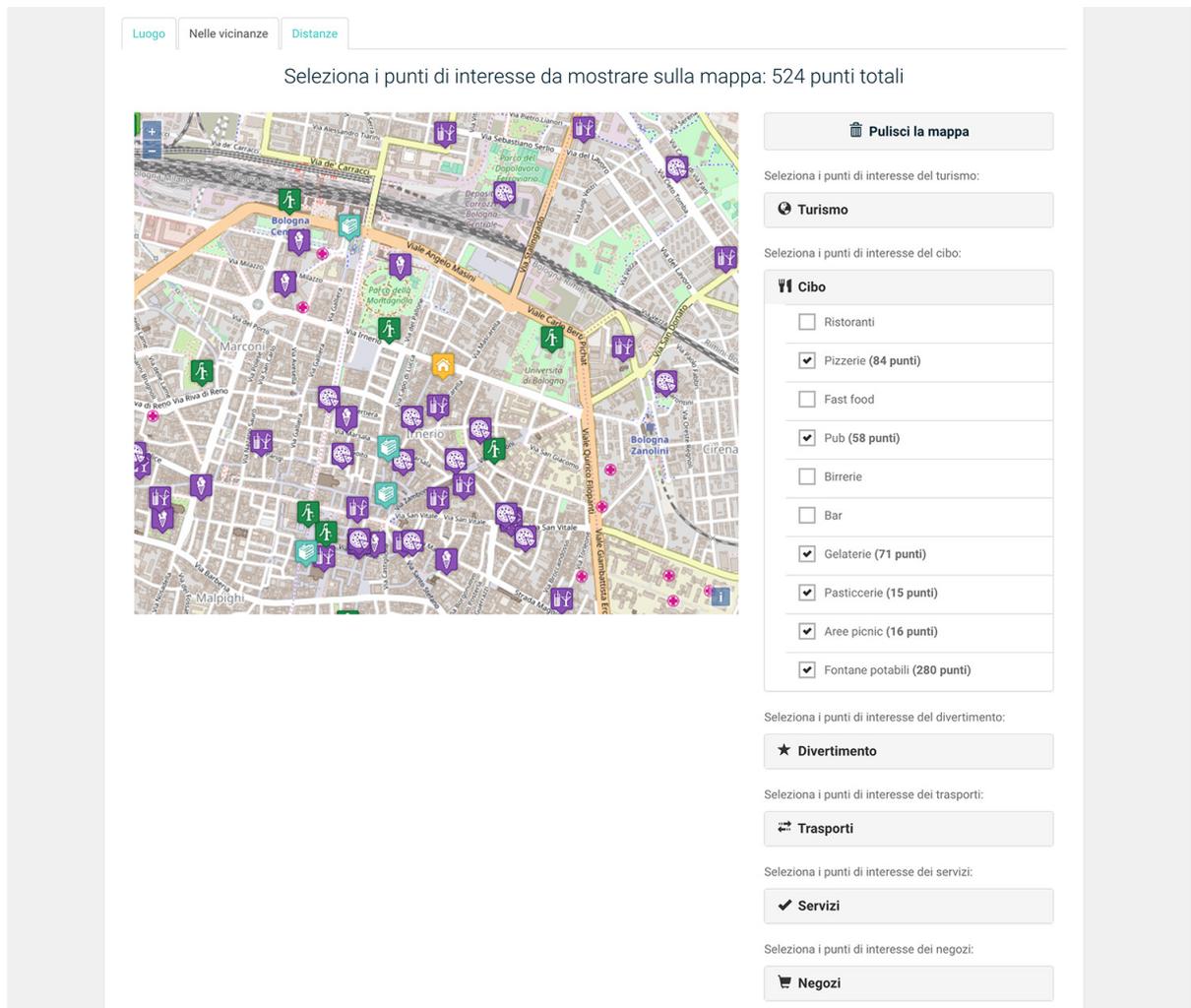


Figura 3.7: Pagine dei bed and breakfast, pannello *Nelle vicinanze* e punti selezionabili

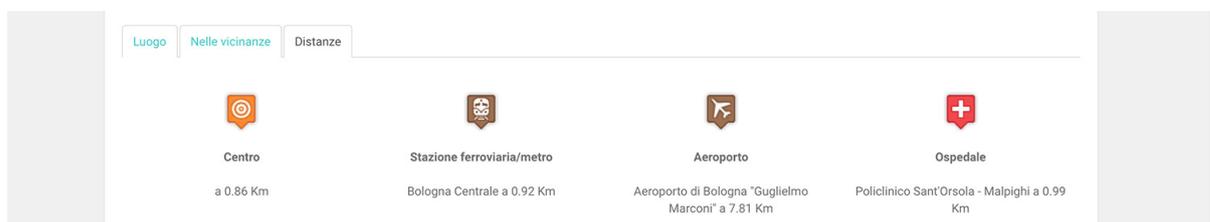


Figura 3.8: Pagine dei bed and breakfast, pannello *Distanze*



Figura 3.9: Pagine dei comuni con marker e cluster sulla mappa



Figura 3.10: Pagine dei comuni con marker selezionati e popup

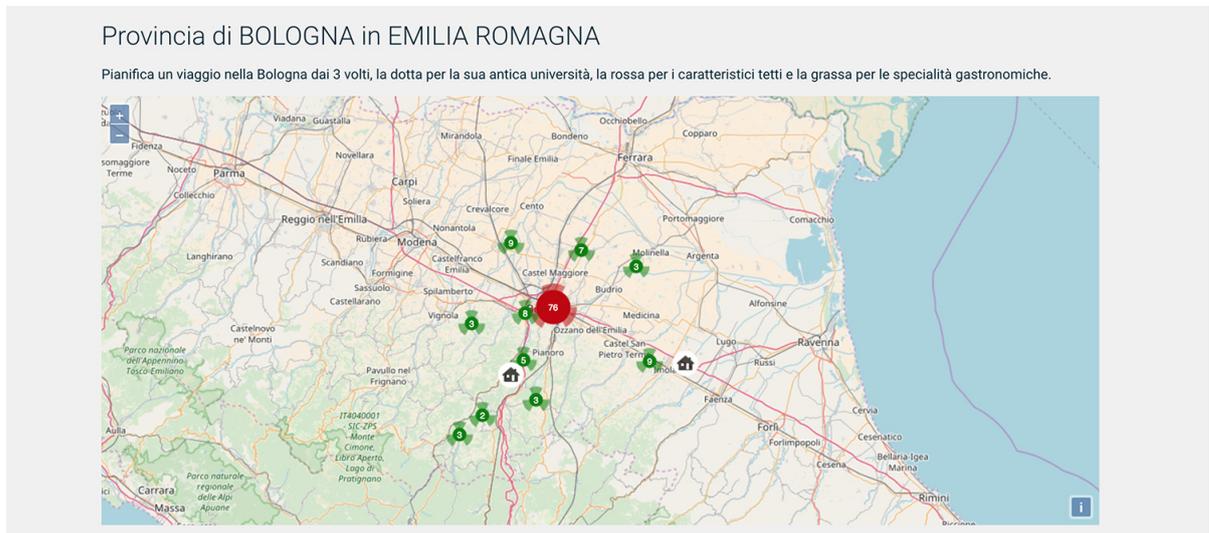


Figura 3.11: Pagine delle province con marker e cluster sulla mappa

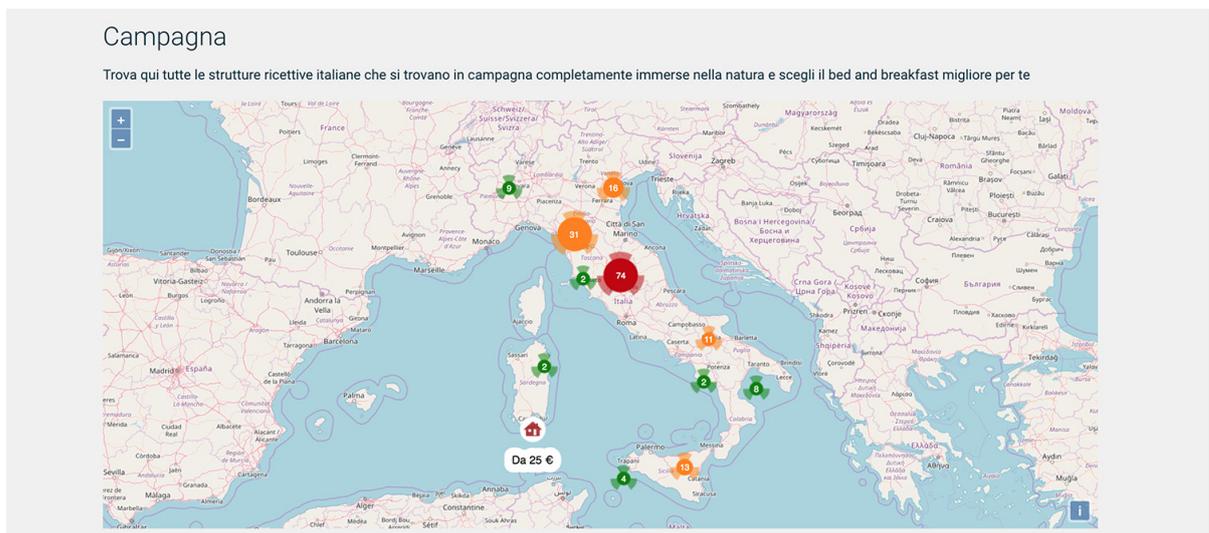


Figura 3.12: Pagine dei luoghi con marker e cluster sulla mappa

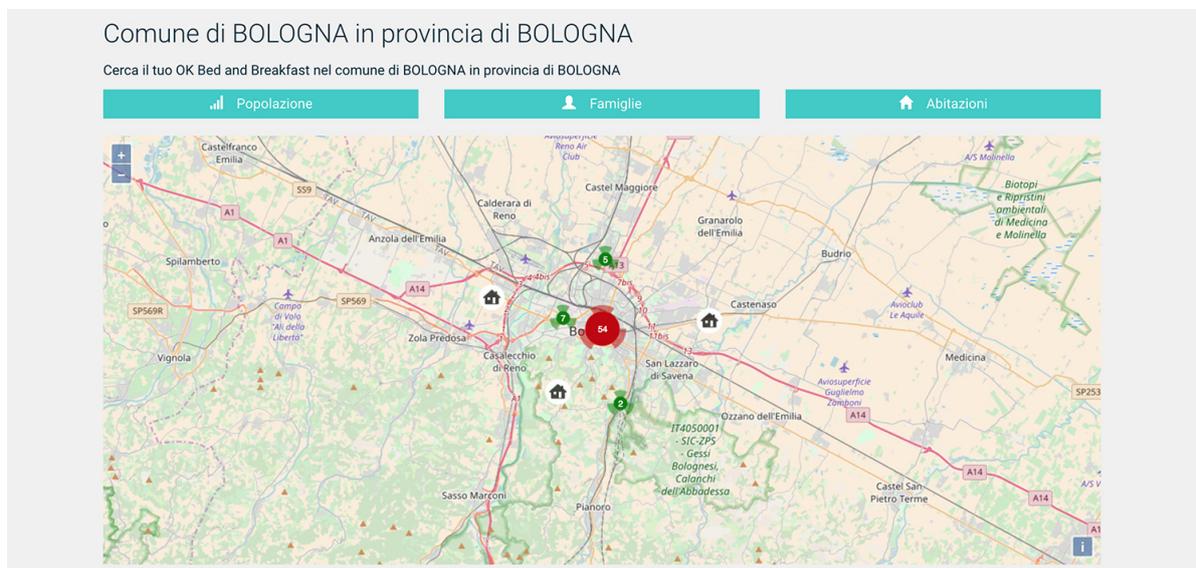


Figura 3.13: Pagine dei comuni con gli open data di Istat

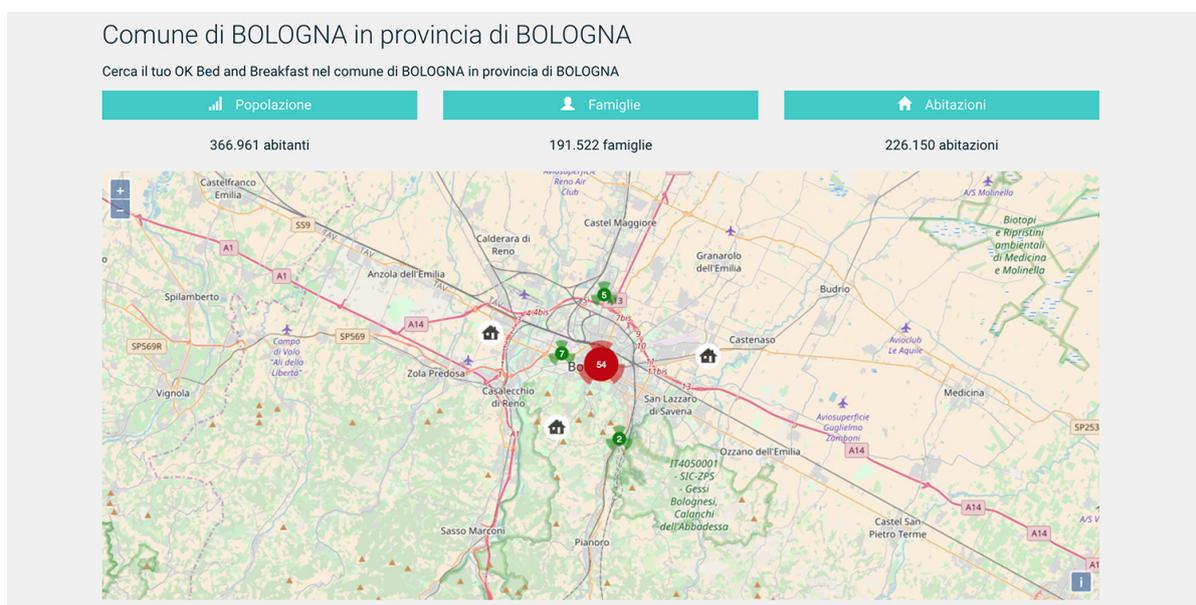


Figura 3.14: Pagine dei comuni con gli open data di Istat calcolati



# Capitolo 4

## Implementazione

Nel capitolo precedente si è descritta la fase di progettazione del lavoro svolto, per definire quali open data implementare nel portale web OkBed&Breakfast e in che modo mostrarli all'utente che utilizza la piattaforma. L'obiettivo di questo capitolo è invece quello di approfondire la fase di implementazione del progetto, quindi gli aspetti più tecnici relativi alle tecnologie utilizzate, al codice scritto in diversi linguaggi di programmazione e agli errori e problemi riscontrati.

### 4.1 Tecnologie utilizzate

Per lo svolgimento del progetto descritto è stato necessario utilizzare diverse tecnologie e linguaggi di programmazione, che insieme hanno consentito l'implementazione degli open data all'interno di OkBed&Breakfast. Specialmente in ambito web, è necessario combinare tra loro e utilizzare contemporaneamente più tecnologie, ad esempio per eseguire query al database o alle API e creare pagine dinamiche, con le quali l'utente può interagire.

#### 4.1.1 Drupal

Drupal è una piattaforma software open source, che rientra nella categoria dei content management system (CMS) [46]. Nato nel 2001, Drupal permette di pubblicare, amministrare e organizzare siti web dinamici e complessi, che utilizzano grandi quantità

e varietà di contenuti, con lo scopo di semplificare e velocizzare il lavoro del *webmaster*. Come accennato, si tratta di un software open source mantenuto e sviluppato da una community di centinaia di utenti e sviluppatori, distribuito con licenza GNU GPL. Di conseguenza chiunque è libero di scaricare ed utilizzare Drupal, anche partecipando attivamente all'evoluzione e al miglioramento del progetto.



Figura 4.1: Logo di Drupal

Drupal è scritto interamente con il linguaggio PHP e viene eseguito lato server, supportando nativamente i database MySQL e PostgreSQL. Il webmaster ha accesso al pannello di controllo che permette la gestione *back-end* del sito web, tramite la quale si definisce anche il *front-end*, cioè i contenuti e l'interfaccia grafica che gli utenti utilizzano. Drupal è una piattaforma modulare e flessibile, che può quindi essere estesa facilmente. Esistono centinaia di moduli che permettono di aggiungere nuove funzionalità al sito web, in modo simile a come avviene per i *plugin* di WordPress, altro CMS open source molto diffuso ma più limitato tecnicamente. Inoltre sono disponibili centinaia di temi grafici, facilmente installabili e modificabili, che permettono la totale indipendenza tra contenuto e presentazione. OkBed&Breakfast è realizzato con Drupal 7, di conseguenza si sono utilizzate diverse funzionalità del CMS per implementare gli open data. Ad esempio la creazione e modifica dei file relativi a moduli, contenuti, nodi e tassonomie presenti sul portale o la gestione dei file minimizzati e delle cache durante lo sviluppo.

#### 4.1.2 HTML

HTML, acronimo di *HyperText Markup Language*, è il linguaggio standard di markup utilizzato per creare pagine web [47]. Nato nel 1993, è di pubblico dominio e deriva

dall'SGML, un metalinguaggio per definire linguaggi di markup. L'HTML permette di creare documenti ipertestuali, definendone struttura, aspetti grafici, testi, immagini e link attraverso l'inserimento di specifici *tag* che vengono interpretati dal browser, il quale genera il DOM (Document Object Model). La sintassi è stabilita dal W3C, che nel corso degli anni ha cercato di definire uno standard comune, insieme alle maggiori aziende e organizzazioni informatiche riunite nel WHATWG. Ogni documento HTML ha una struttura che definisce l'*header*, contenente informazioni di controllo, e il *body*, all'interno del quale è scritto il contenuto vero e proprio della pagina web. Attualmente l'ultima versione pubblicata è HTML 5, che integra funzionalità prima utilizzabili solo tramite estensioni del browser, oltre ad essere pensata per il corretto funzionamento delle pagine sui dispositivi mobili. Anche le pagine web presenti su OkBed&Breakfast sono scritte in HTML, durante l'implementazione del progetto si è utilizzato per definire il *layout* degli elementi visualizzati sullo schermo.

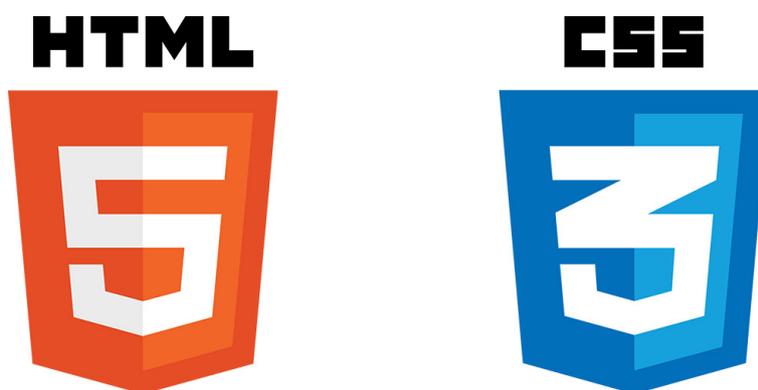


Figura 4.2: Loghi di HTML e CSS

### 4.1.3 CSS

CSS, acronimo di *Cascading Style Sheets*, è un linguaggio utilizzato per descrivere lo stile di pagine web realizzate in HTML o altri linguaggi di markup [48]. Nato nel 1996, le specifiche del linguaggio sono definite dal W3C, con il CSS è possibile separare il contenuto dei documenti HTML dalla loro presentazione. In questo modo l'HTML

è di più facile lettura e risulta più semplice riutilizzare e mantenere il codice dedicato alla presentazione, che può essere inserito in un foglio di stile esterno richiamabile dal documento HTML. È anche possibile scrivere codice CSS direttamente all'interno del file HTML o nei singoli elementi presenti nello stesso. Quando si applicano nuove proprietà di stile agli elementi HTML, questi ultimi sono richiamati tramite vari tipi di selettori, ad esempio di *classe* o *id*. La versione attuale del linguaggio è CSS 3. Durante l'implementazione del progetto, il linguaggio CSS è stato utilizzato per definire lo stile degli elementi visualizzati sullo schermo.

#### 4.1.4 JavaScript

JavaScript è un linguaggio di scripting utilizzato principalmente in ambito web per creare pagine dinamiche [49]. Nato nel 1995, si è evoluto nel tempo e consente di aumentare notevolmente l'interattività delle pagine web con numerosi effetti dinamici. JavaScript è un linguaggio orientato agli oggetti e agli eventi, che funziona lato client, quindi viene eseguito direttamente sul browser dell'utente. Quando quest'ultimo esegue determinate operazioni, ad esempio con mouse e tastiera, vengono invocate le relative funzioni che eseguono il codice collegato e interagiscono con il DOM. Le funzioni JavaScript possono essere raccolte in appositi file esterni oppure scritte direttamente all'interno del codice HTML. Javascript è un linguaggio interpretato dal browser, la sua sintassi deriva da C e Java e dispone di tutte le funzionalità tipiche dei linguaggi di programmazione ad alto livello. Durante l'implementazione del progetto, il linguaggio JavaScript è stato utilizzato insieme ad alcune librerie per realizzare le funzionalità lato client. Ad esempio la creazione delle mappe interattive, la gestione dei cluster e dei punti di interesse.

#### 4.1.5 jQuery

jQuery è una delle librerie JavaScript più utilizzate al mondo [50]. Con jQuery vengono semplificate numerose operazioni, ad esempio la selezione di elementi del DOM, la gestione di eventi e animazioni, oltre all'implementazione di AJAX. Nata nel 2006, la libreria jQuery è distribuita con la licenza libera MIT. Si tratta di un framework estremamente versatile in grado di gestire in maniera efficace aspetti grafici, strutturali e di

manipolazione delle pagine web, inoltre garantisce la compatibilità con tutti i moderni browser. Per utilizzare jQuery è necessario includere i file della libreria all'interno delle pagine HTML. Durante l'implementazione del progetto, jQuery è stato utilizzato frequentemente in vari ambiti. Ad esempio per invocare le chiamate AJAX, inoltre sono state utilizzate diverse sue funzioni e costrutti per la gestione dei punti di interesse e degli elementi da visualizzare sullo schermo.



Figura 4.3: Loghi di JavaScript e jQuery

#### 4.1.6 AJAX

AJAX, acronimo di *Asynchronous JavaScript and XML*, è una tecnica di sviluppo software che permette la creazione di pagine e applicazioni web interattive [51]. Nato nel 2005, la principale caratteristica della tecnologia AJAX consiste nell'aggiornare dinamicamente il contenuto di una pagina web senza un reale *refresh* della stessa. Utilizzando AJAX avviene uno scambio di dati in *background* tra il browser e il server in modo asincrono, senza interferire con il comportamento della pagina web in uso. Il funzionamento di AJAX è reso possibile dall'utilizzo congiunto di più tecnologie: HTML e CSS per la presentazione, DOM per l'interazione, XML, JSON o altri formati per lo scambio dei dati, l'oggetto *XMLHttpRequest* per la comunicazione asincrona e JavaScript che unisce le tecnologie elencate. Solitamente, le funzioni richiamate con le richieste HTTP di AJAX sono scritte in JavaScript, ma questo non è obbligatorio. Durante l'implementazione del progetto, AJAX è stato utilizzato spesso tramite jQuery, per richiamare l'esecuzione di funzioni JavaScript, PHP e per richiedere gli open data con le Overpass API.

### 4.1.7 PHP

PHP, acronimo di *PHP: Hypertext Preprocessor*, è un linguaggio di scripting open source utilizzato principalmente per creare pagine web dinamiche [52]. Nato nel 1995, ciò che distingue PHP da altri linguaggi di scripting è il fatto di essere *server-side*, questo significa che il codice viene eseguito nel server, generando documenti HTML che saranno inviati al client, il quale non potrà conoscere il codice di esecuzione ma solo il risultato finale. Si tratta di un linguaggio interpretato e Turing completo, che offre molte possibilità di utilizzo. Permette la creazione di funzioni e l'utilizzo di tutti i principali costrutti, inoltre è in grado di interfacciarsi con molte tipologie di database. Il codice PHP, la cui sintassi deriva dal linguaggio C, può essere scritto in appositi file o anche inserito all'interno di file HTML. Sono disponibili in rete numerose librerie che ne estendono le funzionalità e la versione attuale del linguaggio è PHP 7. Durante l'implementazione del progetto il linguaggio PHP è stato utilizzato per definire numerose funzioni, specialmente quelle dal maggiore costo computazionale. Ad esempio si è utilizzato per il calcolo delle distanze, in modo che l'esecuzione avvenisse nel server e permettesse anche ai dispositivi mobili meno performanti di visualizzare i risultati.

### 4.1.8 MySQL

MySQL è un *database management system* (DBMS) open source, basato sul modello relazionale [53]. Nato nel 1995, è stato sviluppato dalla *MySQL AB*, acquisita nel 2008 dalla *Sun Microsystems*, a sua volta acquisita nel 2010 dalla *Oracle*. MySQL è uno dei più diffusi database open source del mondo, utilizzato da numerose applicazioni e importanti siti web, oltre ad essere compatibile con i sistemi operativi Windows, macOS e Linux. È disponibile in diverse versioni anche a pagamento ed è composto da un terminale SQL (*Structured Query Language*) a riga di comando che esegue operazioni sul server. MySQL supporta quindi molti costrutti del linguaggio SQL, come viste, query annidate, trigger, transazioni, stored procedures e diversi tipi di dati numerici, testuali, temporali e binari. Inoltre esistono molti software per la gestione dei database MySQL tramite interfaccia grafica, come *MySQL Workbench* e *HeidiSQL*, oppure *phpMyAdmin*, che invece è accessibile da browser. OkBed&Breakfast, come spiegato precedentemente,

è realizzato con Drupal, il quale si appoggia ad un database MySQL. Durante l'implementazione del progetto si è interrogato il database MySQL tramite apposite query, per il recupero delle coordinate e delle altre informazioni relative alle strutture da visualizzare sulle mappe interattive realizzate con OpenLayers. Per la gestione del database tramite interfaccia grafica si è utilizzato il software HeidiSQL.



Figura 4.4: Loghi di PHP e MySQL

#### 4.1.9 SPARQL

SPARQL, acronimo di *SPARQL Protocol and RDF Query Language*, è un linguaggio di interrogazione semantica di dati strutturati nel formato RDF (*Resource Description Framework*) sviluppato dal W3C [54]. Come spiegato nei capitoli precedenti, SPARQL è uno dei pilastri fondamentali su cui si basa il web semantico 3.0, insieme a RDF e OWL (*Web Ontology Language*). Nato nel 2008, con SPARQL si possono estrarre informazioni dalle basi di conoscenza pubblicate sul web, come i linked data e i linked open data. SPARQL è in grado di recuperare e manipolare dati rappresentati in RDF, il quale è composto da *statement* (le unità di base che rappresentano le informazioni) che descrivono concetti e relazioni. Ogni statement è una tripla nella forma *soggetto* (risorsa), *predicato* (proprietà) e *oggetto* (valore collegato ad altre risorse tramite un URI). Per rappresentare i dati definiti mediante il modello RDF si utilizzano apposite serializzazioni come RDF/XML, Turtle, N-Triples, Notation3 e JSON-LD. Tuttavia RDF non esegue assunzioni sul dominio del discorso, per questo tipo di operazioni si utilizza OWL, un linguaggio per la creazione di ontologie, caratterizzate da un vocabolario di concetti e relazioni

tra questi. SPARQL consente la costruzione di query che si basano su congiunzioni e disgiunzioni logiche, triple patterns e pattern opzionali. Durante l'implementazione del progetto, SPARQL è stato utilizzato per formulare le query dedicate al recupero dei dati presenti sul portala Linked Open Data di Istat. In particolare, venivano eseguite richieste HTTP POST contenenti la query, verso l'endpoit SPARQL. Quest'ultimo rispondeva con un file JSON, che una volta ricevuto veniva opportunamente elaborato per estrarre le informazioni necessarie e visualizzare il risultato sul client.



Figura 4.5: Logo W3C del semantic web

#### 4.1.10 Bootstrap

Bootstrap è un framework open source per la creazione di siti e applicazioni web di vario tipo [55]. Nato inizialmente come progetto interno a Twitter, con lo scopo di uniformare l'interfaccia grafica del social network, nel 2011 è stato rilasciato come open source, per permettere ad altri sviluppatori di contribuire al progetto. Bootstrap è pensato per un utilizzo front-end e contiene strumenti e modelli basati su HTML e CSS come form, bottoni, tabelle, menu e tanti altri elementi tipografici e dedicati all'interfaccia grafica, comodamente implementabili all'interno delle pagine web. Sono presenti alcune estensioni opzionali di JavaScript e inoltre il framework supporta il *responsive design*, in modo che il layout del sito web si adatti automaticamente se visualizzato su computer, tablet o smartphone. Bootstrap è compatibile con tutti i moderni browser, utilizza un sistema a griglia per posizionare gli elementi nella pagina web e per utilizzarlo è sufficiente includerlo nel codice HTML. Durante l'implementazione del progetto, Bootstrap è stato utilizzato per impostare il layout dei componenti dell'interfaccia grafica, in modo

che fossero ottimizzati anche per dispositivi mobili. Ad esempio si sono utilizzati gli elementi *collapse* e *list group* per la selezione dei punti di interesse da visualizzare sulla mappa e l'elemento *tabs* per realizzare i pannelli nelle pagine dei bed and breakfast.



Figura 4.6: Loghi di Bootstrap e OpenLayers

#### 4.1.11 OpenLayers

OpenLayers è una delle librerie open source JavaScript più utilizzate al mondo per la creazione di mappe interattive, che l'utente può visualizzare direttamente sul browser [56]. Nata nel 2006, la libreria è stata sviluppata da *MetaCarta*, mentre dal 2007 il progetto è gestito dall'organizzazione no profit *Open Source Geospatial Foundation*. Con OpenLayers si possono realizzare mappe da inserire in pagine web, in grado di accedere tramite API a molte fonti cartografiche sia libere, come OpenStreetMap, che proprietarie, come Google Maps, Bing Maps, Yahoo! Maps e Here. Le mappe create sono formate dall'unione di numerose *tiles*, ovvero singole immagini richieste alla fonte dei dati. OpenLayers è distribuito con la licenza libera BSD e supporta i formati KML, GML e GeoJSON. Permette quindi la visualizzazione di dati vettoriali, marker, popup e cluster sulla mappa, gestendo gli elementi tramite livelli. Inoltre è possibile personalizzare ed estendere la mappa tramite librerie di terze parti, l'ultima versione della libreria è OpenLayers 5. Durante l'implementazione del progetto, OpenLayers 4 è stato utilizzato per creare tutte le mappe interattive mostrate nelle pagine dei bed and breakfast e nelle pagine di comuni, province e luoghi. In particolare si sono utilizzate le mappe di OpenStreetMap, sostituendo tutte le mappe che in precedenza utilizzavano Google Maps ed erano create con le Embed API. Sulle mappe create con OpenLayers si sono aggiunti marker, popup e cluster gestiti dinamicamente, con i quali l'utente può interagire.

### 4.1.12 AnimatedCluster

AnimatedCluster è un'estensione open source della libreria OpenLayers, con la quale è possibile realizzare *cluster*, cioè raggruppamenti di marker in base alla loro vicinanza [57]. AnimatedCluster è compatibile con OpenLayers 3 e 4, e implementa un layer in grado di creare cluster animati e dinamici. Quando si esegue uno zoom-in o uno zoom-out sulla mappa, i cluster vengono espansi o raggruppati attraverso un'animazione grafica. Durante l'implementazione del progetto, AnimatedCluster è stato utilizzato per creare i cluster che raggruppavano i marker delle strutture nelle pagine di comuni, province e luoghi. Dopo aver eseguito la query al database MySQL, per richiedere le informazioni delle strutture da visualizzare sulla mappa con eventuali filtri di ricerca scelti dall'utente, ogni marker veniva creato e aggiunto all'array dei *features*. Successivamente si è creato il layer dedicato ai cluster, i quali contenevano tutti i marker. Inoltre sono state fatte diverse modifiche al funzionamento originale di AnimatedCluster, ad esempio l'aggiunta dello zoom-in al click sui cluster, la comparsa del popup con le informazioni sulla struttura al click sui marker e il riposizionamento del centro della mappa al click sui cluster e sui marker. Altre modifiche hanno riguardato lo stile dei marker, che cambia dinamicamente in base all'interazione con l'utente o se sono presenti informazioni sul prezzo.

### 4.1.13 Overpass API

Overpass API è una potente API (*Application Programming Interface*) di sola lettura, che permette il recupero degli open data dalle mappe di OpenStreetMap e funziona in modo molto simile ad un database raggiungibile via web [58]. Tramite richieste HTTP GET e due appositi linguaggi, chiamati *Overpass QL* e *Overpass XML*, si eseguono query all'API, che recupera gli open data richiesti e li restituisce al client in formato GeoJSON o XML. A differenza della Main API, pensata per la modifica dei dati, e della XAPI, pensata per la lettura dei dati ma sprovvista del supporto a query complesse, l'Overpass API è ottimizzata per gestire query più elaborate che richiedono da zero a 10 milioni di elementi. Il tempo di risposta di Overpass API varia da pochi secondi ad alcuni minuti, a seconda della complessità della query e del numero di elementi contenuti nella risposta. È possibile scrivere query per il recupero dei dati geospaziali che soddisfino

determinati criteri di ricerca, come posizione, tipo di oggetti, proprietà dei tag, prossimità o combinazioni più articolate, inoltre si possono impostare i limiti massimi di *timeout* e dimensione della risposta. È possibile testare il funzionamento delle query scritte in Overpass QL tramite il front-end interattivo Overpass Turbo e il suo wizard. Le richieste a Overpass API possono essere eseguite verso diversi server endpoint pubblici messi a disposizione, ciascuno con differenti limiti di utilizzo.



Figura 4.7: Logo di Overpass API

Gli open data in formato GeoJSON, che Overpass API invia come risposta al client, contengono le informazioni recuperate da OpenStreetMap. Tali dati geospaziali sono gli elementi che rappresentano i componenti fondamentali di OpenStreetMap stesso. Ogni elemento può possedere un certo numero di proprietà chiamate etichette, in inglese *tags*, contraddistinte da coppie chiave-valore che descrivono l'elemento nel dettaglio. Gli elementi possono essere di tre tipologie diverse:

- **Nodo (node):** un nodo rappresenta un punto specifico sulla superficie terrestre definito dalla sua latitudine e longitudine. Ogni nodo comprende almeno un numero ID e una coppia di coordinate geografiche. I nodi sono utilizzati per definire punti di interesse autonomi, ad esempio una panchina, un ristorante o una scuola, in questo caso devono possedere almeno una etichetta che indichi cosa rappresenta il nodo. I nodi servono anche per definire la forma degli elementi *way*, quando vengono usati come punti lungo i percorsi, in questo caso i nodi solitamente non hanno tag, anche se alcuni potrebbero. Ad esempio per contrassegnare i segnali

stradali su una via. Infine, un nodo può anche essere incluso come membro di un elemento *relation*, il quale può indicarne il ruolo, cioè la sua funzione all'interno della relazione.

- **Percorso (way):** un percorso è un'interconnessione ordinata di un insieme di nodi (da 2 a 2.000), che descrive polilinee come strade o fiumi. Nel caso in cui si raggiunga il limite massimo di nodi è sufficiente suddividere il percorso in più percorsi, raggruppandoli all'interno di una relazione. Un percorso è caratterizzato da proprietà omogenee e i nodi contenuti in esso possono appartenere a più percorsi. Un percorso può anche essere chiuso, in questo caso si parla di aree, che ad esempio rappresentano edifici e foreste.
- **Relazione (relation):** una relazione ha il compito di raggruppare più elementi, come nodi, percorsi e anche altre relazioni. I singoli elementi rappresentano i membri della relazione, ognuno di essi ha un ruolo e possono apparire più volte. Le relazioni sono utilizzate per rappresentare entità più complesse, come percorsi ciclabili e linee di autobus, divieti di svolta o multipoligoni che descrivono aree con buchi.

Tramite le Overpass API e il linguaggio di interrogazione Overpass QL si possono richiedere gli elementi appena descritti presenti in una determinata area geografica, e costruire query che restituiscano tali elementi in base al contenuto delle loro etichette. In questo modo, ad esempio, è possibile richiedere le coordinate geografiche di tutti i bar presenti nel comune di Bologna o che si trovano nel raggio di qualche chilometro da un determinato punto.

Durante l'implementazione del progetto, Overpass API è stata utilizzata per il recupero di tutti gli open data relativi ai 60 punti di interesse che l'utente può visualizzare sulla mappa interattiva. Inoltre si è utilizzata per richiedere coordinate e informazioni relative a centro cittadino, stazione ferroviaria/metro, aeroporto e ospedale intorno al bed and breakfast selezionato, per calcolare successivamente i più vicini in linea d'aria. In tutti i casi, si sono sempre realizzate query che richiedono i punti di interesse disponibili nel raggio di 30 o 50 Km dalla struttura, anziché richiedere i punti presenti nello

stesso comune della struttura. Questo per mostrare all'utente una risposta più precisa, anche nel caso in cui il bed and breakfast si trovi al confine con comuni diversi. Tutte le query sono state scritte utilizzando il linguaggio Overpass QL, sono state costruite dinamicamente in base alla checkbox selezionata dall'utente e richiedono i dati tramite chiamate AJAX.

#### 4.1.14 Nominatim API

Nominatim è un tool accessibile sia tramite il sito web *nominatim.openstreetmap.org* sia tramite una API [59]. Si tratta di un potente motore di ricerca con il quale è possibile cercare luoghi e punti di interesse all'interno del database di OpenStreetMap e funziona in modo simile alla Geocoding API di Google Maps. Tramite la ricerca diretta, inserendo nella barra di ricerca sul sito web o nella richiesta tramite API il nome del luogo o un indirizzo, si possono cercare le coordinate di città, regioni e stati, oppure punti di interesse di ogni genere come ristoranti, cinema e musei. Utilizzando invece la ricerca inversa, inserendo nella richiesta al sito web o all'API delle coordinate geografiche precise, il servizio le trasformerà in un indirizzo completo.



Figura 4.8: Logo di Nominatim API

Eseguendo sia la richiesta diretta sia la richiesta inversa, si otterranno risposte nei formati HTML, XML e JSON. Per il buon funzionamento del servizio è consigliato eseguire al massimo una richiesta al secondo. È possibile personalizzare le richieste utilizzando diversi parametri, che consentono ad esempio di scegliere il formato della risposta, il limite massimo di risultati, o di restringere la ricerca a determinate aree geografiche. Nominatim ordina i risultati in base al più pertinente, inoltre non è necessario che il nome

del luogo sia scritto in modo preciso, mentre utilizzando le Overpass API per cercare luoghi in base al loro nome sono presenti più limitazioni. Durante l'implementazione del progetto, l'API di Nominatim è stata utilizzata per il recupero delle coordinate geografiche di determinati luoghi, ottenendo i risultati in formato JSON. In particolare si è utilizzata per determinare le coordinate del centro cittadino, necessarie per il calcolo della distanza dalla struttura selezionata. Inoltre si è utilizzata per impostare il centro delle mappe interattive nelle pagine di comuni e provincie, nel caso in cui non fossero presenti le relative coordinate nel database MySQL.

## 4.2 Codice implementato e file

Per implementare gli open data all'interno di OkBed&Breakfast si sono utilizzati sei diversi *file* (oltre alle librerie incluse) scritti prevalentemente in JavaScript e PHP, nei quali si è fatto uso di tutte le tecnologie descritte in precedenza. In alcuni casi i file esistevano già, di conseguenza si sono modificati direttamente, mentre in altri casi sono stati creati. Nelle sezioni successive si riportano gli *snippet* di codice più rilevanti, suddivisi in base al file a cui appartengono.

### 4.2.1 node-article.tpl.php

Il file rappresenta il nodo Drupal delle pagine dei singoli bed and breakfast ed esisteva già in precedenza. All'interno di esso sono stati realizzati i pannelli *Luogo*, *Nelle vicinanze* e *Distanze*, che utilizzano gli open data forniti da OpenStreetMap e sostituiscono le precedenti mappe realizzate con Google Maps, le quali offrivano meno funzionalità. In questo file si è utilizzato OpenLayers per la creazione delle mappe interattive e Bootstrap per definire il layout. Dopo aver incluso i file CSS e JavaScript di OpenLayers e aver definito lo stile grafico, si sono create le mappe interattive. Di seguito lo snippet che mostra l'implementazione della mappa nel pannello *Luogo*, con il marker fisso che indica la posizione del bed and breakfast selezionato. Tutte le altre mappe realizzate nelle pagine di OkBed&Breakfast sono state implementate in modo simile.

```
1 <style> // Stile della mappa
2 #maposmluogo {
```

```
3     position: relative;
4     height: 500px;
5     width: 100%;
6   }
7 </style>
8 <script> // Mappa pannello luogo
9   jQuery(window).load(function(){
10     var iconFeature = new ol.Feature({
11       geometry: new ol.geom.Point(ol.proj.transform([
12         <?php echo $coordinate_struttura['lng']; ?>,
13         <?php echo $coordinate_struttura['lat']; ?>],
14         'EPSG:4326', 'EPSG:3857')),
15     });
16     var iconStyle = new ol.style.Style({
17       image: new ol.style.Icon(({
18         anchor: [0, 5],
19         anchorXUnits: 'fraction',
20         anchorYUnits: 'pixels',
21         src: '/sites/default/themes/realia/img/markers/home.png'
22       }))
23     });
24     iconFeature.setStyle(iconStyle);
25     var vectorSource = new ol.source.Vector({
26       features: [iconFeature]
27     });
28     var vectorLayer = new ol.layer.Vector({
29       source: vectorSource
30     });
31     var rasterLayer = new ol.layer.Tile({
32       source: new ol.source.OSM()
33     });
34     var maposm_luogo = new ol.Map({
35       layers: [rasterLayer, vectorLayer],
36       target: document.getElementById('maposmluogo'),
37       interactions: ol.interaction.defaults({mouseWheelZoom:false}),
38       view: new ol.View({
39         center: ol.proj.transform([
40         <?php echo $coordinate_struttura['lng']; ?>,
```

```

41         <?php echo $coordinate_struttura['lat']; ?>),
42         'EPSG:4326', 'EPSG:3857'),
43         zoom: 18
44     })
45 });
46 });
47 </script>

```

Di seguito lo snippet di codice che mostra l'implementazione dei tre pannelli e il relativo contenuto. Come esempio si riporta la creazione di uno dei 60 punti di interesse disponibili, in particolare l'elemento *Informazioni turistiche*, appartenente alla categoria *Turismo*.

```

1 <div class="mappe">
2   <ul class="nav nav-tabs">
3     <li class="active"><a data-toggle="tab" href="#luogo">Luogo</a></li>
4     <li><a data-toggle="tab" href="#vicini">Nelle vicinanze</a></li>
5     <li><a data-toggle="tab" href="#distanze">Distanze</a></li>
6   </ul>
7   <div class="tab-content">
8     <div id="luogo" class="tab-pane fade in active"> // Pannello luogo
9       <div id="maposmluogo"></div>
10    </div>
11    <div id="vicini" class="tab-pane fade"> // Pannello nelle vicinanze
12      <center><div class="ex-h2">Seleziona i punti di interesse da mostrare sulla
13        mappa: <span id="puntiTotali"><span></div></center>
14      <br>
15      <div class="col-sm-8">
16        <div id="maposmvicinanze">
17          <div id="popup"></div>
18        </div>
19      <div class="col-sm-4">
20        <div class="panel-group" id="accordion">
21          <a href="#pulisci_mappa" class="btn btn-info btn-lg"><span class="
22            glyphicon glyphicon-trash"></span>Pulisci la mappa</a>
23          <br>
24          <p>Seleziona i punti di interesse del turismo:</p>

```

```

24     <div class="panel panel-default">
25         <div class="panel-heading">
26             <h4 class="panel-title">
27                 <a data-toggle="collapse" data-parent="#accordion" href="#"
                    collapseTurismo"><span class="glyphicon glyphicon-globe"
                    ></span>Turismo</a>
28             </h4>
29         </div>
30     <div id="collapseTurismo" class="panel-collapse collapse">
31         <ul class="list-group">
32             <li class="list-group-item"><input type="checkbox" class="
                    carica-poi" id="informazioniTuristiche" data-nome="
                    informazioni turistiche" data-numero="0" data-citta="<?
                    php echo ucwords(strtolower($content['field_localita'] [' #
                    items'] [0] ['taxonomy_term']->name)) ?>" data-cat="tourism
                    " data-tipo="information" data-immagine="information"
                    data-lat="<?php echo $coordinate_struttura['lat']; ?>"
                    data-lon="<?php echo $coordinate_struttura['lng']; ?>" <
                    label for="informazioniTuristiche">Informazioni
                    turistiche <span id="informazioni turistiche numero"><
                    span></label></li>
33         </ul>
34     </div>
35     </div> // Nel file sono presenti tutti i 60 punti di interesse
36 </div>
37 </div>
38 </div>
39 <div id="distanze" class="tab-pane fade"> // Pannello distanze
40     <div id="distanzeDati" data-citta="<?php echo ucwords(strtolower($content['
                    field_localita'] [' #items'] [0] ['taxonomy_term']->name)) ?>" data-tid="<?
                    php echo $content['field_localita'] [' #items'] [0] ['taxonomy_term']->tid
                    ?>" data-lon="<?php echo $coordinate_struttura['lng']; ?>" data-lat="<?
                    php echo $coordinate_struttura['lat']; ?>" data-tipologie="center,
                    station,aerodrome,hospital"></div>
41     <div id="distanzeContenuto"></div>
42 </div>
43 </div>
44 </div>

```

### 4.2.2 taxonomy-term-location\_istat.tpl.php

Il file definisce il template Drupal delle pagine di comuni, province e regioni ed esisteva già in precedenza. Nelle pagine di comuni e province si sono sostituite le mappe di Google Maps con quelle di OpenStreetMap utilizzando OpenLayers. Sono stati aggiunti i popup, e con la libreria AnimatedCluster i marker delle strutture vengono visualizzati anche tramite cluster, recuperando le coordinate degli stessi dal database MySQL. In precedenza i cluster erano presenti sulle mappe di Google Maps, ma senza la comparsa dei popup al click sui marker. Questo perché in città molto grandi come Roma, Drupal non riusciva a gestire la grande quantità di caratteri generata e la pagina non veniva caricata correttamente. Inoltre, nelle pagine dei comuni, sono stati aggiunti tre bottoni per richiedere gli open data Istat relativi a popolazione, famiglie e alloggi. Di seguito lo snippet che mostra l'implementazione dei cluster con la libreria AnimatedCluster. Creando i marker con la funzione *addFeatures()* e definendone lo stile dinamico in base al livello di aggregazione con la funzione *getStyle()*, si aggiungono i cluster alla mappa.

```
1 var clusterSource=new ol.source.Cluster({
2   distance: 50,
3   source: new ol.source.Vector()
4 });
5 var clusterLayer = new ol.layer.AnimatedCluster({
6   name: 'Cluster',
7   source: clusterSource,
8   animationDuration: 700,
9   style: getStyle
10 });
11 map.addLayer(clusterLayer); // Aggiunta del livello cluster alla mappa
12 addFeatures();
```

Di seguito lo snippet di codice che mostra l'implementazione della funzione *addFeatures()* per le pagine delle province. Scorrendo con un ciclo *foreach* tutti i comuni della provincia, si estraggono dal database MySQL le informazioni delle strutture presenti (con eventuali filtri di ricerca impostati dall'utente), creando i relativi marker.

```
1 function addFeatures(){
2   var features=[];
3   <?php
```

```

4   $padri=taxonomy_get_children($id_provincia);
5   foreach($padri as $padre){ // Query al database MySQL
6       $risultato = db_query("SELECT DISTINCT field_data_field_localita.entity_id
          AS Bid, node.title, file_managed.uri, field_data_field_priceminnotte.
          field_priceminnotte_value AS price, field_data_field_geolocalizzazione.
          field_geolocalizzazione_lat AS lat, field_data_field_geolocalizzazione.
          field_geolocalizzazione_lng AS lng FROM field_data_field_localita LEFT
          JOIN node ON field_data_field_localita.entity_id=node.nid ".$filtri['
          con_foto']." JOIN field_data_field_image ON field_data_field_image.
          entity_id=node.nid AND field_data_field_image.delta=0 LEFT JOIN
          file_managed ON file_managed.fid=field_data_field_image.field_image_fid
          LEFT JOIN field_data_field_geolocalizzazione ON
          field_data_field_geolocalizzazione.entity_id=field_data_field_localita.
          entity_id ".$filtri['camere']." ".$filtri['servizi']." ".$filtri['
          ospiti']." ".$filtri['struttura']." ".$filtri['luoghi']." ".$filtri['
          prezzo']." ".$filtri['contatti_presenti'][0]." WHERE field_localita_tid
          =".$padre->tid." AND node.status=1".$filtri['contatti_presenti'][1]."
          ".$filtri['group_by']." ORDER BY field_data_field_image.field_image_fid
          DESC");
7       foreach ($risultato as $key => $record){ // Creazione dei marker
8           $img=$link='';
9           if(isset($record->lat) && isset($record->lng)){
10              $nome=str_replace(" ", "", $record->title);
11              $nome=str_replace("'", '', $nome);
12              if(isset($record->uri))
13                  $img=theme('image_style', array('style_name'=>'thumbnail', 'path'=>
                    $record->uri));
14              else
15                  $img='';
16              $link=''. $base_url.' /'. drupal_get_path_alias('node/'. $record->Bid).'
                    ;
17              echo "var marker = new ol.Feature(new ol.geom.Point(ol.proj.transform
                    ([".$record->lng.", ".$record->lat."], 'EPSG:4326', 'EPSG:3857')));"
                    ;
18              echo "marker.set('prezzo', ". number_format($record->price, 0). ");";
19              echo "marker.set('immagine', '". $img. "')";";
20              echo "marker.set('link', '". $link. "')";";

```

```
21         echo "marker.set('visitato','0');";
22         echo "features.push(marker);";
23     }
24 }
25 }
26 ?>
27 clusterSource.getSource().clear();
28 clusterSource.getSource().addFeatures(features);
29 }
```

### 4.2.3 taxonomy-term-luoghi.tpl.php

Il file definisce il template Drupal delle pagine dei luoghi (come campagna, centro storico, mare e montagna) ed esisteva già in precedenza. Il suo funzionamento è molto simile al file che gestisce le pagine di comuni, province e regioni. Anche in questo caso la mappa di Google Maps è stata sostituita con quella di OpenStreetMap e sono stati implementati cluster e popup. Di seguito lo snippet di codice che mostra l'implementazione dei popup al click sui marker.

```
1 var element = document.getElementById('popup');
2 var popup = new ol.Overlay({
3     element: element,
4     positioning: 'bottom-center',
5     stopEvent: false,
6     offset: [12, -4]
7 });
8 map.addOverlay(popup);
9 map.on('click', function(evt) { // Funzione attivata al click sulla mappa
10     var element = document.getElementById('popup');
11     var feature = map.forEachFeatureAtPixel(evt.pixel,
12         function(feature) {
13             return feature;
14         });
15     if(feature) { // Se il click e' avvenuto su un marker crea il popup
16         var size = feature.get('features').length;
17         if(size == 1) {
18             jQuery(element).popover('destroy');
```

```
19     var coordinates = feature.getGeometry().getCoordinates();
20     popup.setPosition(coordinates);
21     var featureArray = feature.get('features');
22     var featureee = featureArray[0];
23     var contenuto = '';
24     featureee.set('visitato', '1');
25     if(featureee.get('prezzo') == 0){
26         contenuto = '<center><a class="pbeb" href="' + featureee.get('link') + '
27             ">' + featureee.get('immagine') + '</a><br><a class="pbeb" href="' +
28             featureee.get('link') + '">' + featureee.get('nome') + '</a><center>';
29     } else {
30         contenuto = '<center><a class="pbeb" href="' + featureee.get('link') + '
31             ">' + featureee.get('immagine') + '</a><br><a class="pbeb" href="' +
32             featureee.get('link') + '">' + featureee.get('nome') + '</a><br>Da '
33             + featureee.get('prezzo') + ' €<center>';
34     }
35     jQuery(element).popover({
36         'placement': 'top',
37         'animation': false,
38         'html': true,
39         'content': contenuto
40     });
41     jQuery(element).popover('show');
42     map.getView().setCenter([window.coordinate[0], window.coordinate[1]]);
43 } else { // Altrimenti il popup viene distrutto
44     jQuery(element).popover('destroy');
45     popup.setPosition(undefined);
46 }
47 });
```

#### 4.2.4 markerOpenLayers.js

Il file creato contiene la chiamata in AJAX invocata al click sulle checkbox nel pannello *Nelle vicinanze*, con la richiesta dei dati tramite le Overpass API. Gestisce la lettura degli open data dal file GeoJSON ricevuto in risposta, come mostrato nello snippet di codice seguente. Leggendo nodi, percorsi e relazioni crea i marker e i relativi popup al

mouseover. Inoltre, il file contiene la funzione collegata al bottone *Pulisci la mappa* e le quattro chiamate in AJAX che lanciano l'esecuzione del codice presente nei relativi quattro nodi PHP (all'interno del file *page-ajax.tpl.php*). Tali chiamate si occupano del calcolo delle distanze nel pannello omonimo, oltre al recupero e alla visualizzazione degli open data Istat relativi a popolazione, famiglie e alloggi.

```
1 var punti_totali = 0;
2 jQuery('.carica-poi').each(function() {
3     jQuery(this).change(function() {
4         var riferimento = jQuery(this);
5         if (this.checked) { // Se la checkbox e' stata selezionata
6             jQuery('body').append('');
9             jQuery.ajax({ // Esecuzione chiamata AJAX con query a Overpass API
10                type: 'GET',
11                url: 'https://overpass-api.de/api/interpreter',
12                dataType: "json",
13                data: 'data=[out:json][timeout:40];(node["' + riferimento.attr('data-cat
14                    ') + '=' + riferimento.attr('data-tipo') + '](around:30000, ' +
15                    riferimento.attr('data-lat') + ', ' + riferimento.attr('data-lon') +
16                    ');way["' + riferimento.attr('data-cat') + '=' + riferimento.attr
17                    ('data-tipo') + '](around:30000, ' + riferimento.attr('data-lat') +
18                    ', ' + riferimento.attr('data-lon') + ');relation["' + riferimento.
19                    attr('data-cat') + '=' + riferimento.attr('data-tipo') + '](
20                    around:30000, ' + riferimento.attr('data-lat') + ', ' + riferimento.
21                    attr('data-lon') + '););out body;>;out skel qt;',
22                success: function(response) { // Gestione della risposta in GeoJSON
23                    var iconStyle = new ol.style.Style({
24                        image: new ol.style.Icon(({
25                            anchor: [0, 5],
26                            anchorXUnits: 'fraction',
27                            anchorYUnits: 'pixels',
28                            src: '/sites/default/themes/realia/img/markers/interesse/' +
29                                riferimento.attr('data-immagine') + '.png'
30                        })
31                    })
32                });
33                var vectorSource = new ol.source.Vector({});
```

```
22     var nodicaricati = 0;
23     var waycaricati = 0;
24     var relationcaricati = 0;
25     var punti_attuali = 0;
26     jQuery.each(response.elements, function(index, element) {
27         if ((element.type == 'node') && (typeof element['tags'] !== '
28             undefined')) {
29             nodicaricati = nodicaricati + 1;
30             var indexString = index.toString();
31             window['iconFeature' + indexString] = new ol.Feature({
32                 geometry: new ol.geom.Point(ol.proj.transform([element.lon,
33                     element.lat], 'EPSG:4326', 'EPSG:3857')),
34                 nome: element.tags.name
35             });
36             window['iconFeature' + indexString].setStyle(iconStyle);
37             vectorSource.addFeature(window['iconFeature' + indexString]);
38         } else if ((element.type == 'way') && (typeof element['tags'] !==
39             'undefined')) {
40             var id_primo_nodo = element.nodes[0];
41             jQuery.each(response.elements, function(index2, element2) {
42                 if ((element2.type == 'node') && (element2.id ==
43                     id_primo_nodo)) {
44                     waycaricati = waycaricati + 1;
45                     var lat_way = element2.lat;
46                     var lon_way = element2.lon;
47                     var indexString = index.toString();
48                     window['iconFeature' + indexString] = new ol.Feature({
49                         geometry: new ol.geom.Point(ol.proj.transform([
50                             element2.lon, element2.lat], 'EPSG:4326', 'EPSG
51                             :3857')),
52                         nome: element.tags.name
53                     });
54                     window['iconFeature' + indexString].setStyle(iconStyle);
55                     vectorSource.addFeature(window['iconFeature' +
56                         indexString]);
57                 }
58             });
59         } else if (element.type == 'relation') {
```

```
53     var id_primo_way = element.members[0].ref;
54     jQuery.each(response.elements, function(index3, element3) {
55         if ((element3.type == 'way') && (element3.id ==
56             id_primo_way)) {
57             var id_primo_nodo_way = element3.nodes[0];
58             jQuery.each(response.elements, function(index4, element4
59                 ) {
60                 if ((element4.type == 'node') && (element4.id ==
61                     id_primo_nodo_way)) {
62                     relationcaricati = relationcaricati + 1;
63                     var lat_relation = element4.lat;
64                     var lon_relation = element4.lon;
65                     var indexString = index.toString();
66                     window['iconFeature' + indexString] = new ol.
67                         Feature({
68                             geometry: new ol.geom.Point(ol.proj.transform
69                                 ([element4.lon, element4.lat], 'EPSG:4326
70                                 ', 'EPSG:3857')),
71                             nome: element.tags.name
72                         });
73                     window['iconFeature' + indexString].setStyle(
74                         iconStyle);
75                     vectorSource.addFeature(window['iconFeature' +
76                         indexString]);
77                 }
78             });
79         }
80     });
81     punti_attuali = nodicaricati + waycaricati + relationcaricati;
82     window['vectorLayer' + riferimento.attr('data-tipo')] = new ol.layer.
83         Vector({
84             source: vectorSource
85         });
86     maposm_vicinanze.addLayer(window['vectorLayer' + riferimento.attr('
87         data-tipo')]);
88     riferimento.attr('data-numero', punti_attuali);
```

```
81     punti_totali = punti_totali + punti_attuali;
82     jQuery('#loading_img').remove()
83     document.getElementById("puntiTotali").innerHTML = punti_totali.
        toString() + ' punti totali';
84     var idNumero = riferimento.attr('data-nome') + ' numero';
85     document.getElementById(idNumero).innerHTML = '<strong>' +
        punti_attuali.toString() + ' punti</strong>';
86     },
87     error: function() {
88         jQuery('#loading_img').remove()
89         console.log('Errore inatteso!');
90     }
91 });
92 } else { // Se la checkbox e' stata deselezionata
93     maposm_vicinanze.removeLayer(window['vectorLayer' + riferimento.attr('data-
        tipo')]);
94     punti_totali = punti_totali - riferimento.attr('data-numero');
95     document.getElementById("puntiTotali").innerHTML = punti_totali.toString()
        + ' punti totali';
96     var idNumero = riferimento.attr('data-nome') + ' numero';
97     document.getElementById(idNumero).innerHTML = '';
98     }
99 });
100 });
```

Di seguito è mostrato uno snippet di file GeoJSON ricevuto come risposta dall'interrogazione delle Overpass API. In particolare richiedendo le stazioni ferroviarie/metro nel raggio di 30 Km dalla struttura selezionata.

```
1 {
2   "version": 0.6,
3   "generator": "Overpass API 0.7.55.4 3079d8ea",
4   "osm3s": {
5     "timestamp_osm_base": "2018-09-18T15:08:02Z",
6     "copyright": "The data included in this document is from www.openstreetmap.
        org. The data is made available under ODbL."
7   },
8   "elements": [{
9     "type": "node",
```

```
10     "id": 360437215,  
11     "lat": 44.5936789,  
12     "lon": 11.3709258,  
13     "tags": {  
14         "name": "Funo Centergross",  
15         "operator": "RFI",  
16         "railway": "station",  
17         "wikipedia": "it:Stazione di Funo Centergross"  
18     }  
19 },  
20 {  
21     "type": "way",  
22     "id": 447736873,  
23     "nodes": [  
24         1628372674,  
25         4455261835,  
26         1628372585,  
27         1628372651,  
28         4455261837,  
29         1628372663,  
30         1628372799  
31     ],  
32     "tags": {  
33         "building": "train_station",  
34         "name": "Bologna Centrale",  
35         "operator": "RFI, Grandi Stazioni",  
36         "railway": "station",  
37         "wikipedia": "it:Stazione di Bologna Centrale"  
38     }  
39 },  
40 {  
41     "type": "node",  
42     "id": 1628372674,  
43     "lat": 44.5057171,  
44     "lon": 11.3417651  
45 }  
46 ]  
47 }
```

### 4.2.5 page-ajax.tpl.php

Il file contiene funzioni scritte in PHP gestite tramite uno *switch case*, tali funzioni sono invocate da numerose chiamate AJAX eseguite su OkBed&Breakfast. Ogni *case* rappresenta un contenuto di tipo AJAX creato tramite Drupal. Il file esisteva già in precedenza ed è stato aggiunto il codice invocato dalle quattro chiamate AJAX presenti all'interno di *markerOpenLayers.js*. Il codice implementato calcola la distanza tra la struttura selezionata e il centro cittadino e costruisce le query per richiedere i punti di interesse (stazioni ferroviarie/metro, aeroporti e ospedali) nelle vicinanze della struttura, per calcolare successivamente le distanze minime. Infine richiede gli open data Istat su popolazione, famiglie e abitazioni, li legge e li elabora per stampare il risultato. Si è scelto di realizzare queste funzioni in PHP e non in JavaScript per eseguire i calcoli sul server anziché sul client e per aggirare l'attuale mancanza di supporto al protocollo HTTPS da parte del portale Linked Open Data di Istat. Di seguito è riportato lo snippet di codice che mostra il recupero degli open data da Nominatim API e il calcolo della distanza dal centro cittadino. Inoltre si mostra la costruzione della query per richiedere alle Overpass API gli ospedali nelle vicinanze della struttura.

```
1 case 20530:
2 $citta = $_POST['citta'];
3 $tid_comune = $_POST['tid'];
4 $lat = $_POST['lat'];
5 $lon = $_POST['lon'];
6 $tipologie = $_POST['tipologie'];
7 $arrayTipologie = explode(',', $tipologie);
8 foreach($arrayTipologie as $tipologia) {
9     if ($tipologia == 'center'){ // Calcolo distanze dal centro cittadino
10         $nome_citta_convertito = str_replace(" ", "%20", $citta);
11         $options = array(
12             'method' => 'GET',
13             'timeout' => 40,
14         ); // Richiesta a Nominatim API
15         $result = drupal_http_request('https://nominatim.openstreetmap.org/search/'.
16             $nome_citta_convertito.'?format=json&limit=1', $options);
17         $json_response = drupal_json_decode($result -> data);
```

```

17     $distanza = (6370 * 3.1415926 * sqrt(($json_response[0]['lat'] - $lat) * (
        $json_response[0]['lat'] - $lat) + cos($json_response[0]['lat'] /
        57.29578) * cos($lat / 57.29578) * ($json_response[0]['lon'] - $lon) * (
        $json_response[0]['lon'] - $lon)) / 180);
18     echo '<div class="col-sm-3"><center><br><br>';
19     echo '<strong>Centro</strong><br><br>a '.number_format($distanza, 2).' Km<br>
        <br></center></div>';
20 }
21 if ($tipologia == 'hospital'){ // Costruzione query Overpass QL ospedali vicini
22     $query3 = '[out:json][timeout:40];(node["amenity"="'.$tipologia.'"]["
        emergency"="yes"](around:30000, '.$lat.', '.$lon.);way["amenity"="'.
        $tipologia.'"["emergency"="yes"](around:30000, '.$lat.', '.$lon.);
        relation["amenity"="'.$tipologia.'"["emergency"="yes"](around:30000, '.
        $lat.', '.$lon.););out body;>;out skel qt;';
23     $risultato3 = calcola_distanze($query3, $lat, $lon); // Chiamata funzione per
        il calcolo delle distanze minime passando la query costruita
24     echo '<div class="col-sm-3"><center><br><br>';
25     echo '<strong>Ospedale</strong><br><br>'.$risultato3['nome'].' a '.
        number_format($risultato3['distanza'], 2).' Km<br><br></center></div>';
26 }
27 }
28 break;

```

Di seguito lo snippet di codice che mostra il recupero degli open data Istat relativi alla popolazione di un comune, interrogando l'endpoint SPARQL del portale Linked Open Data di Istat. Dopo aver ricevuto il file JSON di risposta, i dati vengono letti per calcolare il totale degli abitanti, poiché i dataset sono suddivisi in base alle località.

```

1 case 20531:
2 $codiceistat = $_POST['codiceistat'];
3 $options = array(
4     'method' => 'POST',
5     'timeout' => 150,
6 ); // Query all'endpoint SPARQL
7 $query = urlencode('PREFIX ORACLE_SEM_HT_NS: <http://oracle.com/semtech#
        ALL_LINK_HASH> PREFIX ORACLE_SEM_FS_NS: < http: //oracle.com/semtech#timeout

```

```

=600,allow_dup=t,strict_default=f> PREFIX rdf: < http://www.w3.org/1999/02/22-
rdf-syntax-ns#> PREFIX rdfs: < http://www.w3.org/2000/01/rdf-schema#> PREFIX
ter: < http://datiopen.istat.it/odi/ontologia/territorio/> PREFIX cen: < http:
//datiopen.istat.it/odi/ontologia/censimento/> PREFIX qb: < http://purl.org/
linked-data/cube#> SELECT ? pop WHERE { ? comune ter: haNome ? nomecom. ?
comune ter : haCodIstat ? cod_istat_com. ? comune ter : localita_di_COM ?
entitaTerritoriale. ? entitaTerritoriale ter : haNome ? nomelocalita. ?
entitaTerritoriale ter : haIndicatoreCensimento ? o. ? o cen :
haPopolazioneResidente ? pop. ? o cen : haClassiEta16Categorie ? CE16Categorie.
? o rdfs : label ? desc_o. ? CE16Categorie rdfs : comment ? Eta. FILTER(lang(
? Eta) = "it") FILTER(lang( ? desc_o) = "it") FILTER( ? Eta = "Totale di
classificazione"@it) FILTER(regex( ? desc_o, "Popolazione residente Totale"))
FILTER( ? cod_istat_com = "'. $codiceistat.'")}});
8 $querytot = 'http://datiopen.istat.it/sparql/oracle?query='.$query.'&format=json';
9 $result = drupal_http_request($querytot, $options); $json_response =
    drupal_json_decode($result -> data);
10 $popolazione_totale = 0;
11 // Lettura e calcolo della risposta in JSON
12 foreach($json_response['results']['bindings'] as $key => $elemento) {
13     $popolazione_totale = $popolazione_totale + $elemento['pop']['value'];
14 }
15 echo '<br>'.str_replace(",",".", (string) number_format($popolazione_totale, 0, ".
    ", ",")).' abitanti';
16 break;

```

Di seguito è mostrato uno snippet di file JSON ricevuto come risposta dall'interrogazione dell'endpoint SPARQL del portale Linked Open Data di Istat. Nell'esempio sono stati richiesti gli open data relativi alla popolazione del comune di Bologna, i quali sono suddivisi per località (sono mostrate le prime due).

```

1 {
2   "head": {
3     "vars": ["pop", "nomelocalita"]
4   },
5   "results": {
6     "bindings": [
7       {
8         "pop": {

```

```
9         "type": "literal",
10        "datatype": "http://www.w3.org/2001/XMLSchema#decimal",
11        "value": "366133"
12    },
13    "nomelocalita": {
14        "type": "literal",
15        "xml:lang": "it",
16        "value": "Bologna"
17    }
18 },
19 {
20     "pop": {
21         "type": "literal",
22         "datatype": "http://www.w3.org/2001/XMLSchema#decimal",
23         "value": "306"
24     },
25     "nomelocalita": {
26         "type": "literal",
27         "xml:lang": "it",
28         "value": "Lavino di Mezzo"
29     }
30 }
31 ]
32 }
33 }
```

Di seguito è mostrato uno snippet di file GeoJSON ricevuto come risposta dall'interrogazione della Nominatim API, per il recupero delle coordinate del centro cittadino, necessarie per il calcolo della distanza dalla struttura. Nell'esempio sono stati richiesti gli open data relativi alla posizione del comune di Bologna. Il risultato è stato limitato ad uno solo con il parametro *limit=1*, che restituisce il primo elemento più pertinente alla ricerca.

```
1 [
2   {
3     "place_id": "198025018",
4     "licence": "Data c OpenStreetMap contributors, ODbL 1.0. https://osm.org/
      copyright",
```

```
5   "osm_type": "relation",
6   "osm_id": "43172",
7   "boundingbox": [
8     "44.421033",
9     "44.556094",
10    "11.2296206",
11    "11.4336295"
12  ],
13  "lat": "44.4936714",
14  "lon": "11.3430347",
15  "display_name": "Bologna, BO, EMR, Italia",
16  "class": "place",
17  "type": "city",
18  "importance": 0.801066118511801,
19  }
20 ]
```

### 4.2.6 okbedbreakfast.module

Il file è un modulo personalizzato di Drupal, contiene funzioni scritte in PHP utilizzate per la gestione di OkBed&Breakfast ed esisteva già in precedenza. Nella funzione *okbedbreakfast\_init()* è stata aggiunta l'inclusione automatica del file *markerOpenLayers.js* nel *footer* di tutte le pagine di tipo *node article* e *taxonomy term*. Inoltre è stata scritta la funzione *calcola\_distanze()*, invocata dal file *page-ajax.tpl.php*, per il calcolo delle distanze dalla struttura all'aeroporto, stazione ferroviaria/metro e ospedale più vicini. Di seguito è mostrato uno snippet di tale funzione, che invia alle Overpass API la query ricevuta come parametro in ingresso, successivamente elabora il file GeoJSON ricevuto in risposta e ritorna il punto di interesse a distanza minima. La funzione è stata realizzata per consentire in futuro un'eventuale aggiunta più comoda di ulteriori distanze minime tra la struttura e punti di interesse di altro tipo.

```
1 function calcola_distanze($query, $lat, $lon)
2 {
3   $data = array('data' => ''.$query.'');
4   $options = array(
5     'method' => 'GET',
```

```

6     'data' => drupal_http_build_query($data) ,
7     'timeout' => 40,
8 ); // Esecuzione query a Overpass API
9 $result = drupal_http_request('https://overpass-api.de/api/interpreter',
    $options);
10 $json_response = drupal_json_decode($result->data);
11 $datiTipologia = array();
12 // Gestione della risposta in GeoJSON con calcoli distanze
13 foreach($json_response['elements'] as $key => $elemento) {
14     if (($elemento['type'] == 'node') && (array_key_exists('tags', $elemento))) {
15         $distanza = (6370 * 3.1415926 * sqrt(($elemento['lat'] - $lat) * ($elemento
            ['lat'] - $lat) + cos($elemento['lat'] / 57.29578) * cos($lat /
            57.29578) * ($elemento['lon'] - $lon) * ($elemento['lon'] - $lon)) /
            180);
16         $datiTipologia[number_format($distanza, 2)] = array('nome' => $elemento['
            tags']['name'], 'distanza' => $distanza);
17     }
18     else if ($elemento['type'] == 'way') {
19         if ($elemento['tags']['name']) {
20             $id_primo_nodo = $elemento['nodes'][0];
21             foreach($json_response['elements'] as $key2 => $elemento2) {
22                 if (($elemento2['type'] == 'node') && ($elemento2['id'] ==
                    $id_primo_nodo)) {
23                     $lat_way = $elemento2['lat'];
24                     $lon_way = $elemento2['lon'];
25                 }
26             }
27             $distanza = (6370 * 3.1415926 * sqrt(($lat_way - $lat) * ($lat_way -
                $lat) + cos($lat_way / 57.29578) * cos($lat / 57.29578) * ($lon_way
                - $lon) * ($lon_way - $lon)) / 180);
28             $datiTipologia[number_format($distanza, 2)] = array('nome' => $elemento[
                'tags']['name'], 'distanza' => $distanza);
29         }
30     }
31     else if ($elemento['type'] == 'relation') {
32         $id_primo_way = $elemento['members'][0]['ref'];
33         foreach($json_response['elements'] as $key3 => $elemento3) {
34             if (($elemento3['type'] == 'way') && ($elemento3['id'] == $id_primo_way)){

```

```
35     $id_primo_nodo_way = $elemento3['nodes'][0];
36     foreach($json_response['elements'] as $key4 => $elemento4) {
37         if (($elemento4['type'] == 'node') && ($elemento4['id'] ==
38             $id_primo_nodo_way)) {
39             $lat_relation = $elemento4['lat'];
40             $lon_relation = $elemento4['lon'];
41         }
42     }
43 }
44 $distanza = (6370 * 3.1415926 * sqrt(($lat_relation - $lat) * (
45     $lat_relation - $lat) + cos($lat_relation / 57.29578) * cos($lat /
46     57.29578) * ($lon_relation - $lon) * ($lon_relation - $lon)) / 180);
47 $datiTipologia[number_format($distanza, 2)] = array('nome' => $elemento[
48     'tags']['name'], 'distanza' => $distanza);
49 }
50 ksort($datiTipologia);
51 $datiPoiMin = reset($datiTipologia);
52 return $datiPoiMin; // Ritorna l'elemento alla distanza minima
53 }
```

## 4.3 Errori e soluzioni

In questa sezione sono riportati gli errori e le problematiche riscontrate durante l'implementazione degli open data su OkBed&Breakfast. Per ogni errore sono specificate descrizione del problema, possibili cause ed eventuali soluzioni applicate.

- Nella mappa delle pagine di comuni, province e luoghi, può sorgere un problema relativo alla corretta visualizzazione dei marker delle strutture. In alcuni casi, eseguendo uno zoom-in non appaiono tutti i singoli marker ma rimane il cluster con un unico gruppo di marker nello stesso punto. Questo problema è dovuto ad un malfunzionamento quando venne importato nel database MySQL il dataset della camera di commercio contenente i dati dei bed and breakfast. Più strutture hanno assunto le stesse coordinate, venendo visualizzate tramite cluster nella stessa

posizione. Il problema era presente anche in precedenza con le mappe di Google Maps e si può riscontrare ad esempio nella città di Torino, nella quale rimane un cluster contenente 16 strutture. Per risolvere il problema, è necessario modificare nel database le coordinate delle strutture che presentano questo difetto.

- Nella mappa delle pagine di comuni, province e luoghi, prima dell'implementazione degli open data di OpenStreetMap, erano stati disabilitati i popup al click sui marker delle strutture. In precedenza, utilizzando le mappe di Google Maps, in città o province nelle quali erano presenti molte strutture, il caricamento della pagina non andava a buon fine. Il problema era dovuto al fatto che Drupal non era in grado gestire l'enorme quantità di codice generato per costruire la pagina web, in particolare la creazione dei popup contenenti le informazioni della struttura selezionata. Di conseguenza, durante l'implementazione delle mappe di OpenStreetMap, si è cercato di generare meno codice possibile, per permettere il corretto funzionamento dei popup anche nelle grandi città come Roma.
- Nel pannello *Distanze* delle pagine dei bed and breakfast, per recuperare le coordinate del centro cittadino, inizialmente si sono usate le Overpass API. Si richiedevano le coordinate del comune nel quale si trovava la struttura e per una risposta più precisa si includeva nell'interrogazione anche il nome della regione, determinando prima quello della provincia. Tali dati venivano recuperati dal database MySQL con apposite query. Il problema delle Overpass API era che i nomi dei luoghi dovevano corrispondere a quelli presenti nel database di OpenStreetMap. Quindi si sono risolti i problemi grammaticali più comuni, passando alla query per le Overpass API i nomi corretti, altrimenti la richiesta non sarebbe andata a buon fine. Nonostante le correzioni, per alcuni luoghi il problema permaneva. Ad esempio, i comuni dell'Alto Adige su OpenStreetMap contengono anche la traduzione in tedesco. Per risolvere definitivamente il problema, si sono usate le Nominatim API. Con queste API, eseguendo richieste contenenti solo il nome del comune e senza alcun tipo di correzione, si ricevono in risposta le coordinate del centro cittadino. Nominatim API funziona quindi in modo più intelligente e non è necessario che i nomi dei luoghi cercati corrispondano esattamente a quelli presenti su OpenStreetMap.

- Un problema rimasto irrisolto riguarda la posizione dei popup al click sui marker delle strutture, nelle mappe delle pagine di comuni, province e luoghi. Premendo sui marker delle strutture con prezzo, il popup appare spostato in basso. Il problema avviene normalmente solo al primo click su questo tipo di marker, mentre premendo sui marker senza prezzo non si riscontrano problemi. L'errore riguarda l'errata posizione iniziale di comparsa del popup, poiché nel calcolo viene incluso anche lo spazio occupato dalla label fissa contenente il prezzo.
- Altri problemi rimasti irrisolti riguardano la comparsa in console degli errori *Cannot read property '0' of null*, *Cannot read property 'off' of null* e *Cannot read property 'trigger' of null* nelle mappe delle pagine di comuni, province e luoghi e nel pannello *Nelle vicinanze*. Il primo errore avviene quando l'utente esegue uno zoom-in su una porzione di mappa contenente marker delle strutture con prezzo. Gli altri due errori quando l'utente esce al mouseover dall'area del marker e il popup viene distrutto. Tali problemi riguardano le librerie OpenLayers e Bootstrap e sono causati dal sistema che tenta di leggere delle proprietà dei marker e dei popup settati al valore *null*. Nonostante gli errori descritti, le mappe interattive, la gestione dei cluster e la comparsa/scomparsa dei popup al mouseover funzionano correttamente.
- Un problema riscontrato saltuariamente riguarda il caricamento di tutte le *tiles* delle mappe OpenStreetMap. Si tratta di un problema di sovraccarico dei server di OpenStreetMap che non riescono a soddisfare tutte le richieste, di conseguenza compare l'errore in console e sulla mappa rimangono riquadri grigi, cioè *tiles* non caricate. Per risolvere il problema di norma è sufficiente ricaricare la pagina.
- Un altro problema riscontrato saltuariamente riguarda il caricamento di tutti i punti di interesse selezionati nel pannello *Nelle vicinanze*. Se l'utente seleziona velocemente molti tipi di marker da mostrare (specialmente nelle grandi città), Overpass API potrebbe non supportare una tale richiesta di dati in breve tempo. In questo caso compare l'errore in console e le query non vanno a buon fine. Per arginare il problema si è introdotta la comparsa della rotellina di caricamento dei dati e l'apertura di un solo *collapse* alla volta, limitando l'utente nell'esecuzione di un numero eccessivo di richieste in pochi secondi.



## Conclusioni ed estensioni future

Da quanto discusso in questa tesi si è compresa la sempre crescente importanza che hanno i dati nella società attuale, in particolare gli open data, liberamente utilizzabili da chiunque e per qualsiasi scopo, rappresentano un generatore di nuova conoscenza. Sono state approfondite le loro caratteristiche principali, relative al recupero e utilizzo. Implementando gli open data in progetti software come siti web e app è possibile sfruttarli per creare servizi innovativi e arricchire quelli già esistenti, oltre a poter essere utilizzati per diffondere conoscenza e aiutare la ricerca.

Nel lavoro descritto nel presente elaborato, dopo una prima fase progettuale e di analisi su come usare al meglio gli open data disponibili, essi sono stati implementati all'interno del portale web turistico OkBed&Breakfast in modo da creare un valore aggiunto per chi naviga tra le pagine del sito web. Grazie agli open data l'utente ha a disposizione molte più informazioni relative al luogo nel quale è presente il bed and breakfast selezionato, come i punti di interesse di vario tipo disponibili nelle vicinanze. Di conseguenza, tramite la nuova conoscenza acquisita, esso sarà in grado di prendere una decisione migliore relativamente alla struttura più adatta alle sue esigenze.

Essendo la tematica degli open data un argomento in espansione, non è stato semplice inizialmente trovare tutte le informazioni necessarie per comprenderne appieno i vari aspetti. Di conseguenza è stato necessario eseguire ricerche approfondite, specialmente nelle fonti ufficiali, prima di passare alla fase progettuale e successivamente a quella implementativa. Per consentire in futuro un utilizzo più comodo degli open data durante l'implementazione all'interno di progetti software, è necessario che vengano adottati

sempre di più standard comuni tra le varie fonti, preferendo la pubblicazione di dati con cinque stelle di apertura, cioè i linked open data. È importante che gli open data disponibili siano sempre più di qualità, completi e aggiornati, oltre a doverne promuovere pubblicazione e utilizzo da parte di governi, aziende e cittadini.

Di seguito sono raccolte le possibili estensioni future relative al progetto realizzato:

- Un'estensione futura, riguarda la possibilità implementare all'interno del portale OkBed&Breakfast altre tipologie di open data. Ad esempio dati ambientali forniti da Arpa, oppure dati sui luoghi forniti da DBpedia. Tali open data potrebbero essere aggiunti in tutte le pagine di comuni, province, regioni, luoghi e strutture o solo nelle aree geografiche dove sono disponibili i dati stessi.
- Altra possibile estensione futura, riguarda la possibilità di caricare nelle diverse mappe interattive di OpenStreetMap solo i punti di interesse o i marker delle strutture visualizzati all'interno della *viewport* (cioè la porzione della mappa mostrata sullo schermo in un preciso momento). In questo modo, quando l'utente si sposta all'interno della mappa, si caricherebbero dinamicamente i punti rimanenti, senza quindi richiederli tutti inizialmente.
- Altra possibile estensione futura, riguarda la possibilità di realizzare un modulo di Drupal per il back-end di OkBed&Breakfast. Tale modulo consentirebbe di modificare i punti di interesse, le categorie e le distanze che l'utente può visualizzare nelle pagine dei singoli bed and breakfast. Grazie ad un modulo di Drupal tali modifiche potrebbero essere fatte da un amministratore del sito web, senza la necessità di modificare il codice presente nei file.
- L'unico riferimento a Google Maps, che è stato sostituito da OpenStreetMap, è rimasto nella pagina back-end di creazione e modifica dei dati di un bed and breakfast. Quindi una possibile implementazione futura è di rimuovere anche in questa pagina la mappa di Google Maps, sostituendola con quella di OpenStreetMap. In particolare, il modulo rimasto delle mappe Google serve a impostare o modificare le coordinate geografiche della struttura tramite le Geocoding API.

- Altra possibile estensione futura, riguarda la realizzazione dell'applicazione di Ok-Bed&Breakfast per dispositivi mobili iOS e Android. Con l'app si potrebbero far interagire le mappe interattive create con la posizione dell'utente, localizzato tramite GPS, rete cellulare e Wi-Fi. Ad esempio, suggerendo i bed and breakfast e i punti di interesse disponibili in base alla sua posizione. Inoltre, con le notifiche push si potrebbe contattare l'utente in modo più mirato ed efficace, invitandolo ad interagire con il portale, ad esempio per lasciare una recensione.

Si è compreso che il valore dei dati e delle informazioni oggi è nettamente superiore a qualsiasi altra forma di contributo. Di conseguenza, all'aumentare dello sviluppo e della diffusione degli open data, aumenteranno i benefici per l'intera società, che potrà sfruttare e diffondere più velocemente la conoscenza. Utilizzando i moderni mezzi tecnologici, informatici e di comunicazione, gli open data sono parte integrante di questo circolo virtuoso.



# Elenco delle figure

1.1	Tim Berners-Lee . . . . .	16
1.2	Icone dei principali browser web . . . . .	17
1.3	Logo di Google . . . . .	19
1.4	Rappresentazione astratta della rete internet mondiale . . . . .	21
2.1	Logo di Open Knowledge International . . . . .	29
2.2	Le 5 stelle degli open data . . . . .	31
2.3	Loghi di alcune licenze aperte . . . . .	37
2.4	Principali tipologie di open data . . . . .	40
2.5	Logo di CKAN . . . . .	43
2.6	Logo di OpenStreetMap . . . . .	45
2.7	Loghi di I.Stat e Linked Open Data . . . . .	49
2.8	Logo di Dati.gov.it . . . . .	51
2.9	Logo di DatiOpen.it . . . . .	53
2.10	Logo di Open Data Comune di Bologna . . . . .	55
2.11	Logo di Arpae Emilia-Romagna . . . . .	56
2.12	Logo di DBpedia . . . . .	57
2.13	Loghi di alcune fonti di open data . . . . .	59
2.14	Numero di dataset pubblicati in Italia tra il 2012 e il 2015 . . . . .	60
2.15	Classifica degli Stati europei in ambito open data . . . . .	63
2.16	Il Linked Open Data Cloud ad agosto 2018 . . . . .	63
2.17	<i>L'Open Data Button</i> indica i progetti che utilizzano gli open data . . . . .	66
3.1	Logo di OkBed&Breakfast . . . . .	72

---

3.2	Home page di OkBed&Breakfast su computer, tablet e smartphone . . .	73
3.3	Architettura del progetto . . . . .	75
3.4	Pagine dei bed and breakfast, pannello <i>Luogo</i> . . . . .	86
3.5	Pagine dei bed and breakfast, pannello <i>Nelle vicinanze</i> . . . . .	87
3.6	Pagine dei bed and breakfast, pannello <i>Nelle vicinanze</i> e punti selezionati	87
3.7	Pagine dei bed and breakfast, pannello <i>Nelle vicinanze</i> e punti selezionabili	88
3.8	Pagine dei bed and breakfast, pannello <i>Distanze</i> . . . . .	88
3.9	Pagine dei comuni con marker e cluster sulla mappa . . . . .	89
3.10	Pagine dei comuni con marker selezionati e popup . . . . .	89
3.11	Pagine delle province con marker e cluster sulla mappa . . . . .	90
3.12	Pagine dei luoghi con marker e cluster sulla mappa . . . . .	90
3.13	Pagine dei comuni con gli open data di Istat . . . . .	91
3.14	Pagine dei comuni con gli open data di Istat calcolati . . . . .	91
4.1	Logo di Drupal . . . . .	94
4.2	Loghi di HTML e CSS . . . . .	95
4.3	Loghi di JavaScript e jQuery . . . . .	97
4.4	Loghi di PHP e MySQL . . . . .	99
4.5	Logo W3C del semantic web . . . . .	100
4.6	Loghi di Bootstrap e OpenLayers . . . . .	101
4.7	Logo di Overpass API . . . . .	103
4.8	Logo di Nominatim API . . . . .	105





## Elenco delle tabelle

2.1	Licenze aperte raccomandate dall'Open Knowledge International . . . . .	39
2.2	Numero di dataset disponibili nelle principali fonti di open data . . . . .	61
2.3	Top 10 dataset scaricati dallo European Union Open Data Portal . . . . .	62



# Bibliografia e sitografia

- [1] Sito web "A Scuola di OpenCoesione".  
[<http://www.ascuoladiopencoessione.it/portfolio-articoli/lezione-approfondire-1718/>]
  
- [2] Sito web "IETF Tools".  
[<https://tools.ietf.org/html/rfc2616>]
  
- [3] Andrea D'Alessandro, "Una storia dell'ipertesto". Politecnico di Torino.  
[<https://areeweb.polito.it/didattica/polymath/ICT/Htmls/Argomenti/Appunti/StoriaIpertesto/Img/Una%20Storia%20dell%27Ipertesto.pdf>]
  
- [4] Sito web "StatCounter".  
[<http://gs.statcounter.com>]
  
- [5] Sito web "Search Engine History".  
[<http://www.searchenginehistory.com>]
  
- [6] Victoria Shannon, "A 'more revolutionary' Web". The New York Times, 2006.  
[<https://www.nytimes.com/2006/05/23/technology/23iht-web.html>]
  
- [7] Sito web "Linking Open Data Project".  
[<https://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>]
  
- [8] Keshab Nath, Sourish Dhar e Subhash Basishtha, "Web 1.0 to Web 3.0 - Evolution of the Web and its various challenges". IEEE, 2014.  
[<https://ieeexplore.ieee.org/document/6798297/>]

- [9] Christian Bizer, Tom Heath e Tim Berners-Lee, "Linked Data - The Story So Far". The International Journal on Semantic Web and Information Systems, 2009.  
[<https://www.igi-global.com/article/linked-data-story-far/37496>]
- [10] Sareh Aghaei, Mohammad Ali Nematbakhsh e Hadi Khosravi Farsani, "Evolution of the World Wide Web : From Web 1.0 to Web 4.0". International Journal of Web & Semantic Technology, 2012.  
[<http://airccse.org/journal/ijwest/papers/3112ijwest01.pdf>]
- [11] Felix Wortmann e Kristina Flüchter, "Internet of Things - Technology and Value Added". Business & Information Systems Engineering, 2015.  
[<https://aisel.aisnet.org/bise/vol57/iss3/8/>]
- [12] Michael Armbrust, Armando Fox, Rean Griffith, Anthony D. Joseph, Randy Katz, Andy Konwinski, Gunho Lee, David Patterson, Ariel Rabkin, Ion Stoica e Matei Zaharia, "A View of Cloud Computing". Communications of the ACM, 2010.  
[<https://dl.acm.org/citation.cfm?id=1721672>]
- [13] Report "Global Digital 2018" e "Italia Digital 2018". We Are Social, 2018.  
[<https://wearesocial.com/it/blog/2018/01/global-digital-report-2018>]
- [14] Sito web "Open Innovation Regione Lombardia".  
[<http://www.openinnovation.regione.lombardia.it>]
- [15] Sito web "Open Definition".  
[<https://opendefinition.org>]
- [16] Giuseppe Rizzo, Federico Morando e Juan De Martin, "Open Data: la piattaforma di dati aperti per il Linked Data". Informatica e diritto, 2011.  
[<http://www.ittig.cnr.it/EditoriaServizi/AttivitaEditoriale/InformaticaEDiritto/Rizzo.Ied.1-2-2011.html>]
- [17] Sito web "Open data Handbook". Open Knowledge International, 2010.  
[<http://opendatahandbook.org/>]

- [18] Documento "Ten Open Data Guidelines". Transparency International Georgia, 2011. [<https://www.transparency.ge/en/ten-open-data-guidelines>]
- [19] Sito web "5-star Open Data". [<https://5stardata.info/en/>]
- [20] Barack Obama, "Transparency and Open Government". The White House, 2009. [<https://obamawhitehouse.archives.gov/the-press-office/transparency-and-open-government>]
- [21] Alessandra Cornero, Ernesto Belisario e Stefano Epifani, "Vademecum Open Data: come rendere aperti i dati delle pubbliche amministrazioni". Formez PA, 2011. [<http://trasparenza.formez.it/content/vademecum-open-data-come-rendere-aperti-dati-pubbliche-amministrazioni>]
- [22] Sito web "Open Government Partnership Italia". [<http://open.gov.it>]
- [23] Ernesto Belisario, "Licenze per il riutilizzo dei dati pubblici e Open Data". Formez PA, 2014. [[http://egov.formez.it/sites/all/files/open\\_data\\_-\\_licenze\\_per\\_il\\_riutilizzo\\_dei\\_dati\\_pubblici\\_e\\_open\\_data.pdf](http://egov.formez.it/sites/all/files/open_data_-_licenze_per_il_riutilizzo_dei_dati_pubblici_e_open_data.pdf)]
- [24] Sito web "Open Definition". [<https://opendefinition.org/licenses/>]
- [25] Luca Delucchi, Maurizio Napolitano, Alessio Zanol, "Introduzione a OpenStreetMap". Comunità italiana di OpenStreetMap, 2012. [[https://svn.openstreetmap.org/misc/pr\\_material/italy\\_miniguide/tutorial.pdf](https://svn.openstreetmap.org/misc/pr_material/italy_miniguide/tutorial.pdf)]
- [26] Sito web "OpenStreetMap". [<https://www.openstreetmap.org>]
- [27] Sito web "OpenStreetMap Wiki". [<https://wiki.openstreetmap.org>]
- [28] Sito web "Linked Open Data Istat". [<http://datiopen.istat.it>]

- [29] Sito web "I.Stat".  
[<http://dati.istat.it>]
- [30] Sito web "Dati.gov.it".  
[<https://www.dati.gov.it>]
- [31] Sito web "DatiOpen.it".  
[<http://www.datiopen.it>]
- [32] Sito web "Open Data Comune di Bologna".  
[<http://dati.comune.bologna.it>]
- [33] Sito web "Linked Open Data Comune di Bologna".  
[<http://linkeddata.comune.bologna.it>]
- [34] Sito web "Open Data Arpae Emilia-Romagna".  
[<https://dati.arpae.it>]
- [35] Sito web "DBpedia".  
[<https://wiki.dbpedia.org>]
- [36] Sito web "Open Data Hub Italia".  
[<https://www.sciamlab.com/opendatahub/it/>]
- [37] Sito web "Open Data Emilia-Romagna".  
[<http://dati.emilia-romagna.it>]
- [38] Sito web "European Data Portal".  
[<https://www.europeandataportal.eu>]
- [39] Sito web "European Union Open Data Portal".  
[<http://data.europa.eu/euodp/en/home>]
- [40] Sito web "Data.gov".  
[<https://www.data.gov>]

- [41] Sito web "DataHub".  
[<https://datahub.io>]  
[<https://old.datahub.io>]
- [42] Report "Open Data Maturity in Europe 2017". European Data Portal, 2017.  
[[https://www.europeandataportal.eu/sites/default/files/edp\\_landscaping\\_insight\\_report\\_n3\\_2017.pdf](https://www.europeandataportal.eu/sites/default/files/edp_landscaping_insight_report_n3_2017.pdf)]
- [43] Sito web "The Linked Open Data Cloud".  
[<https://lod-cloud.net>]
- [44] Sito web "Magic".  
[<https://www.magicnet.it/portfolio/okbb-domanda-offerta-alloggio>]
- [45] Sito web "Open Government Data: The Book". Joshua Tauberer, 2014.  
[<https://opengovdata.io/2014/bulk-data-an-api/>]
- [46] Sito web "Drupal".  
[<https://www.drupal.org>]
- [47] Sito web "HTML".  
[<https://www.w3.org/html/>]
- [48] Sito web "CSS".  
[<https://www.w3.org/Style/CSS/>]
- [49] Sito web "JavaScript".  
[<https://developer.mozilla.org/it/docs/Web/JavaScript>]
- [50] Sito web "jQuery".  
[<https://jquery.com>]
- [51] Sito web "AJAX".  
[<https://developer.mozilla.org/it/docs/Web/Guide/AJAX/Iniziare>]
- [52] Sito web "PHP".  
[<http://php.net>]

- [53] Sito web "MySQL".  
[<https://www.mysql.com/it/>]
- [54] Sito web "SPARQL".  
[<https://www.w3.org/TR/sparql11-overview/>]
- [55] Sito web "Bootstrap".  
[<https://getbootstrap.com>]
- [56] Sito web "OpenLayers".  
[<https://openlayers.org>]
- [57] Sito web "AnimatedCluster".  
[<https://github.com/Viglino/OL3-AnimatedCluster>]
- [58] Sito web "Overpass API".  
[<http://overpass-api.de>]
- [59] Sito web "Nominatim API".  
[<https://nominatim.openstreetmap.org>]





# Ringraziamenti

*Vorrei spendere alcune righe di questa mia tesi di laurea per ringraziare tutti coloro che mi hanno sostenuto e aiutato durante questo percorso universitario, un percorso impegnativo e ricco di soddisfazioni. Desidero innanzitutto ringraziare il Prof. Marco Di Felice, relatore di questa tesi, per il tempo che mi ha dedicato, la sua professionalità e competenza. Un ringraziamento particolare va al mio correlatore Massimo Selva, per avermi permesso di svolgere il tirocinio per tesi nella sua azienda, la Magic, oltre ad avermi supportato nella realizzazione del progetto. Questo ringraziamento è rivolto anche a Electra Nadalini, anch'essa titolare dell'azienda, e a Piero Nanni, uno dei dipendenti che mi ha seguito nello sviluppo della parte più tecnica del lavoro.*

*Il ringraziamento più grande va a tutta la mia famiglia, che mi ha sempre supportato moralmente ed economicamente in questo percorso, grazie infinite mamma e papà per tutto quello che fate per me. Desidero ringraziare mia sorella Sonia, alla quale voglio molto bene e che mi è sempre stata accanto. Un ringraziamento di cuore va ai miei zii, Clara e Adelchi, da sempre disponibili e presenti nella mia vita. Grazie anche al nostro amato labrador Kira, un cane molto speciale. Desidero inoltre ringraziare tutti i miei amici, in particolare Riccardo, grazie a tutti voi per i momenti di gioia trascorsi insieme. Ringrazio anche i miei colleghi e amici di corso, in particolare Cristian, Mirko, Simona e Anna, grazie davvero per esserci stati, rendendo più piacevoli questi anni di università. Infine un grazie a me stesso, per essere riuscito con tanto impegno a raggiungere questo importante traguardo, punto di arrivo e di inizio nella mia vita.*

*Lorenzo Biagio Lanzarone*

*Bologna, 16 ottobre 2018*