

# Realizzazione di una piattaforma per l'erogazione di open data

Relatore:  
**Dott.**  
**Andrea Maurino**

Relazione della prova finale di:  
**Luca Racchetti**  
**Matricola 703311**

Co-relatore:  
**Dott.**  
**Gianluigi Viscusi**

Anno Accademico 2012-2013

# Indice

<b>1</b>	<b>Introduzione</b>	<b>3</b>
1.1	Obiettivo della tesi . . . . .	3
1.2	Struttura . . . . .	4
<b>2</b>	<b>Cosa si intende per Open Data</b>	<b>5</b>
2.1	Movimento Open Data . . . . .	5
2.1.1	Scala 5 stelle . . . . .	8
2.1.2	Open Data in ambito governativo . . . . .	9
2.1.3	Open Data in Italia . . . . .	9
2.2	Impianto Normativo . . . . .	10
2.2.1	Leggi Regionali . . . . .	14
<b>3</b>	<b>Piattaforme Tecnologiche</b>	<b>20</b>
3.1	Socrata (Open Source) . . . . .	20
3.2	CKAN . . . . .	23
<b>4</b>	<b>Progettazione portale Open Data a Vigevano</b>	<b>24</b>
4.1	Descrizione del Comune . . . . .	24
4.1.1	Servizio Informatico Comunale . . . . .	25
4.2	Progettazione piattaforma . . . . .	26
4.3	Integrazione con Regione Lombardia . . . . .	31
<b>5</b>	<b>Sviluppo della soluzione</b>	<b>32</b>
5.1	Preview e Web API . . . . .	33
5.2	Archiviazione dei file . . . . .	34
5.3	Automazione analisi file caricati . . . . .	35
5.4	Catalogo dati . . . . .	35
5.5	Dizionario dati . . . . .	39
<b>6</b>	<b>Utilizzo della piattaforma</b>	<b>40</b>

**INDICE** **2**

---

<b>7 Conclusioni e sviluppi futuri</b>	<b>57</b>
7.1 Conclusioni . . . . .	57
7.2 Sviluppi futuri . . . . .	59
<b>Bibliografia</b>	<b>60</b>
<b>Ringraziamenti</b>	<b>63</b>

# **Capitolo 1**

## **Introduzione**

Lo stage è stato svolto presso il Laboratorio MiSS (Management of Information Systems and Services) del Dipartimento di Informatica Sistemistica e Comunicazione (DISCo). L'attività ha visto l'interlocuzione con il Servizio Informatico Comunale (SIC) del Comune di Vigevano. Il progetto prevedeva la creazione di una demo di un portale Open Data seguendo le direttive richieste dal Comune stesso.

### **1.1 Obiettivo della tesi**

L'obiettivo di questa tesi è di:

- descrivere la filosofia e la pratica che permeano gli Open Data
- analizzare il processo che ha portato all'attuale situazione legislativa, in particolare la variegata situazione regionale
- confrontare le piattaforme tecnologiche al momento esistenti
- spiegare le esigenze implementative del Comune di Vigevano
- spiegare come questa piattaforma è stata realizzata

Infine ha lo scopo di effettuare un'analisi riguardante la reale sostenibilità da parte del Comune della scelta proposta.

## 1.2 Struttura

Il secondo capitolo parlerà di cosa sono gli Open Data, cercando di darne una definizione che includa le varie sfaccettature di questa nuova pratica e filosofia, del loro ruolo nella società e della loro importanza in ambito governativo. Si affronterà l'analisi di chi sono gli attori principali intorno a questo argomento e le loro esigenze. Successivamente si affronterà l'aspetto legale riguardante la materia Open Data, dalla spinta iniziale Americana, fino alle leggi emanate negli ultimi anni nella nostra Penisola. Si farà inoltre un'analisi dell'attuale situazione Italiana per comprendere la differente situazione nelle varie regioni.

Nel terzo capitolo si approfondiranno alcune soluzioni tecnologiche, cercando di comprendere la loro architettura, analizzando le caratteristiche offerte e spiegando l'esperienza dell'utente finale.

All'interno del quarto capitolo si esaminerà la situazione del Comune di Vigevano, studiando un'adeguata soluzione rispetto all'esigenza di pubblicare le informazioni presenti nelle loro banche dati e alla volontà di integrare la piattaforma che verrà realizzata con quella in uso da Regione Lombardia.

Il quinto capitolo descriverà la realizzazione della piattaforma, partendo dall'installazione della piattaforma scelta e proseguendo con la personalizzazione e la creazione delle funzionalità richieste in fase di progettazione.

Nel sesto capitolo sarà fatta una presentazione dell'utilizzo della piattaforma realizzata, cercando di mostrare tutte le sue potenzialità.

Infine nel settimo capitolo sarà fatta un'analisi del lavoro svolto e saranno tratte le dovute conclusioni e ipotizzati gli sviluppi futuri del progetto.

# Capitolo 2

## Cosa si intende per Open Data

### 2.1 Movimento Open Data

Negli anni la pratica e l'ideologia che caratterizza i dati aperti si sono ben consolidate, ma solo recentemente è nata una nuova accezione legata maggiormente al mondo di Internet, che diventa il canale principale di diffusione dei dati stessi, e con essa si identifica il termine “Open Data”. Secondo wikipedia [1] per Open Data si intendono *“alcune tipologie di dati liberamente accessibili a tutti, privi di brevetti o altre forme di controllo che ne limitino la riproduzione e le cui restrizioni di copyright eventualmente si limitano ad obbligare di citare la fonte o al rilascio delle modifiche allo stesso modo”*.

Partecipazione, Democrazia, Comunità, Diritti, sono solo alcune delle parole che vengono associate agli OpenData. In figura 2.1 si vede come Open Data sia uno dei pilastri dell’Open Government, ovvero l’idea in base alla quale la Pubblica Amministrazione deve essere aperta ai cittadini in termini di trasparenza, di collaborazione e di partecipazione diretta al processo decisionale.

La partecipazione diretta può avvenire attraverso il ricorso alle nuove tecnologie dell’informazione e della comunicazione, sulla base delle esperienze positive già sperimentate in altri movimenti e comunità “open”, quali l’*open source*, l’*open access* e l’*open content*.

È bene dunque evidenziare che Open Data di per sé non vuol dire *e-participation* (partecipazione in processi di governo e di governance) o *e-democracy*, ma insieme ad essi costituisce i pilastri dell’Open Government. Ciò comporta che i dati siano messi a disposizione dalla Pubblica Amministrazione in sola lettura da parte degli utenti, i quali li utilizzano in base alle loro

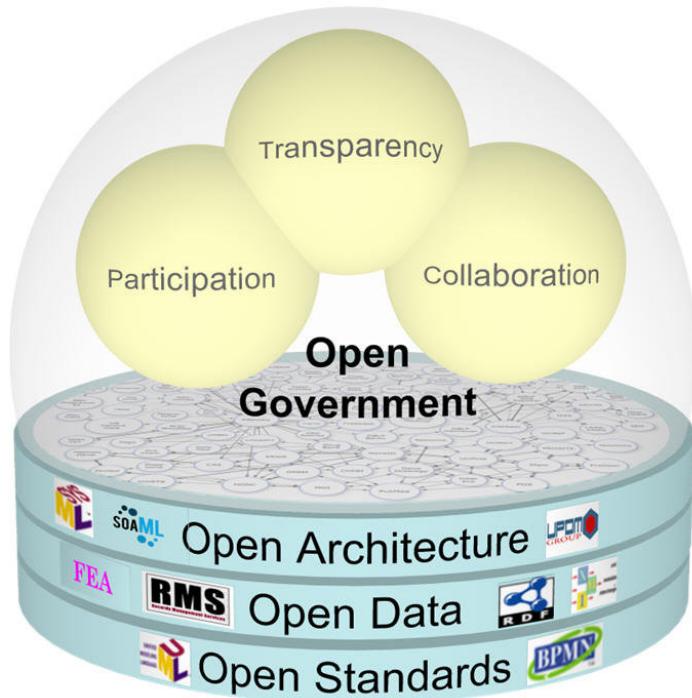


Figura 2.1: Pilastri dell'OpenGovernment

esigenze:

- **Cittadini:** per misurare le performance e la trasparenza della Pubblica Amministrazione (es. dati di bilancio), o per informazione (es. orari mezzi pubblici, sportelli sanitari, ...).
- **Civil Hackers:** per spirito civico sviluppano app basate sugli Open Data per intenti civili, trovando il meccanismo per mettere facilmente a disposizione dei cittadini i dati pubblici.
- **Giornalisti:** per cercare informazioni legati a interessi particolari (es. luoghi della città che presentano statisticamente un maggior tasso di criminalità).
- **Imprese:** principalmente per analisi di mercato (es. esercizi sul territorio interessati a un prodotto) o per produzione di software che sfrutta l'informazione presente nei dati.
- **Pubblica Amministrazione:** per migliorare le scelte decisionali (es. un tasso di malattia anomalo può essere correlato alla presenza di sostanze tossiche).

I dati del governo sono considerati aperti se sono resi pubblici in modo da rispettare gli **8 principi dell'Open Government Data** definiti nel 2007 a Sebastopol, California [2] :

1. **Completi:** Tutti i dati pubblici sono resi disponibili. Sono dati che non sono soggetti a limitazioni di privacy, di sicurezza o limitazioni di privilegio.
2. **Primari:** I dati vengono raccolti alla fonte, con il più alto livello di “gnularità” possibile, non in forme aggregate o modificate.
3. **Tempestivi:** I dati sono resi disponibili nei tempi necessari a preservarne il valore e devono essere mantenuti aggiornati.
4. **Accessibili:** I dati vengono resi disponibili al maggior numero di utenti possibile per i più vari scopi.
5. **Processabili:** I dati sono adeguatamente strutturati per consentire l’elaborazione automatizzata.
6. **Non discriminatori:** I dati sono disponibili a chiunque, senza necessità di registrazione.
7. **Non proprietari:** I dati sono disponibili in un formato su cui nessun ente ha l’esclusivo controllo.
8. **Liberi da licenza:** I dati non sono soggetti a diritti d’autore, brevetti, marchi o segreto commerciale. Possono essere ammesse ragionevoli restrizioni di privacy, sicurezza e privilegio.

In aggiunta agli 8 principi è inoltre importante rendere i dati:

- **Riutilizzabili:** Gli utenti devono essere messi in condizione di riutilizzare e integrare i dati, in modo da creare nuove risorse, applicazioni, programmi e servizi di pubblica utilità.
- **Ricercabili:** Facilmente identificabili in rete e facilmente indicizzabili dai motori di ricerca.
- **Permanenti:** Le caratteristiche sopra descritte devono caratterizzare i dati nel corso della loro intera vita sul Web.

Siccome le necessità di utilizzo dei dati sono le più svariate, anche i file che sono stati resi disponibili sono di diverso formato e possono riguardare statistiche o censimenti, dati sensibili accorpati, dati georeferenziati o materiale multimediale.

### 2.1.1 Scala 5 stelle

Tim Berners Lee in un suo testo del 2006, successivamente integrato nel 2010 [3], per incoraggiare i governi ed altri enti pubblici a rilasciare i dati in formato aperto, introdusse il concetto di **Dati a 5 stelle**. I concetti trattati nell'articolo sono il *Semantic Web* e i *Linked Open Data*, ossia particolari tipi di dati strutturati semanticamente che vengono pubblicati in rete con licenze di consultazione e d'uso poco restrittive (quali le licenze Creative Commons o le licenze nazionali) e che viene classificato secondo la scala a 5 stelle.

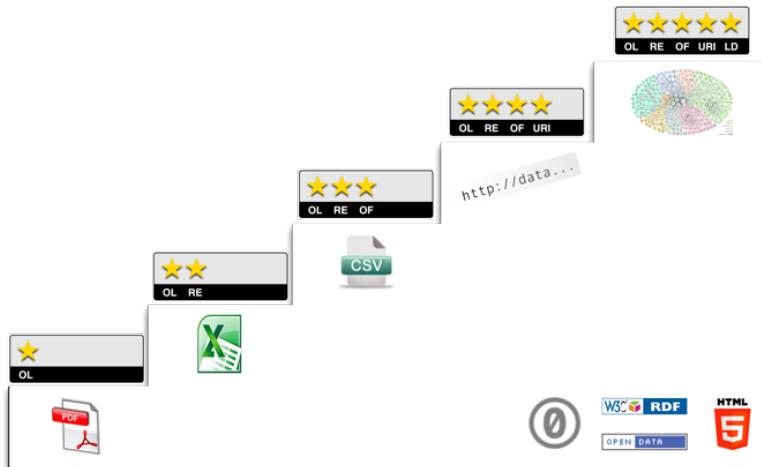


Figura 2.2: Scala a 5 stelle

La scala a 5 stelle (figura 2.2) vuole misurare quanto i dati siano aperti e accessibili, assegnando le stelle secondo i seguenti criteri:

- no star    **Web data** (in qualsiasi formato) senza nessuna licenza open
- ★    Il dato è disponibile sul Web (in qualsiasi formato) con una **licenza aperta**
- ★ ★    Il dato è disponibile in un **formato strutturato** (ad esempio un Excel invece della scansione di una tabella) in modo che possa essere riutilizzato
- ★ ★ ★    Il dato è in un **formato non proprietario** (ad esempio un CSV invece di un Excel)
- ★ ★ ★ ★    Il dato utilizza gli **URI** per identificare gli oggetti, in modo che le persone possano far riferimento alle sue risorse e fare uso di RDF
- ★ ★ ★ ★ ★    Il dato è **collegato** ad altri dati per fornire un contesto

### 2.1.2 Open Data in ambito governativo

Una grossa spinta all'affermazione del movimento Open Data in ambito governativo è stata data dal presidente degli Stati Uniti d'America **Barack Obama** con la promulgazione di una Direttiva sull'Open Government nel dicembre 2009 [4]:

*Fin dove possibile e sottostando alle sole restrizioni valide, le agenzie devono pubblicare le informazioni on line utilizzando un formato aperto (open) 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.*

Alla direttiva è stato dato seguito con la pubblicazione sul sito web [Data.gov](http://Data.gov) di molti dati in formato aperto. Il portale ha infatti lo scopo di raccogliere tutte le informazioni rese disponibili dagli enti statunitensi. Nel maggio 2013 Obama ha firmato un nuovo ordine esecutivo [5] che obbliga tutte le agenzie governative ad adottare dati in formato aperto compatibili anche con le future infrastrutture informatiche che verranno adottate negli USA. Le pubbliche amministrazioni dovranno assicurarsi, dove è legalmente consentito, che i dati vengano rilasciati in modo tale da risultare facilmente accessibili e utilizzabili.

### 2.1.3 Open Data in Italia

La prima esperienza **in Italia** è stata realizzata tra il 2007 e il 2010 dalla volontà di alcune amministrazioni locali che han voluto partecipare al progetto **OpenStreetMap**, il quale crea e fornisce dati cartografici (ad esempio mappe stradali) liberi e gratuiti a chiunque ne abbia bisogno. I comuni coinvolti, insieme a dei volontari, hanno condiviso i dati dei propri stradari sotto licenza aperta.

Nel giugno 2010 Renato Brunetta, l'allora Ministro per la pubblica amministrazione e l'innovazione, dichiarò in un'intervista l'intenzione di realizzare un portale italiano dell'Open Data seguendo il modello dei datagov anglosassoni entro la fine dell'anno. Il 18 Ottobre 2011 il portale [dati.gov.it](http://dati.gov.it) è stato messo on line.

Nel maggio 2010 la regione Piemonte ha realizzato il proprio portale regionale [dati.piemonte.it](http://dati.piemonte.it). Il sito è ancor oggi una delle esperienze nazionali con

la miglior riuscita. Nel 2011 la regione Emilia-Romagna ha seguito l'esempio piemontese.

A partire da settembre 2011 il Comune di Firenze, grazie anche alla spinta del suo Sindaco Matteo Renzi, ha realizzato il proprio portale [opendata.comune.fi.it](#) rilasciando un dataset al giorno nella fase iniziale. Inoltre Firenze è stato il primo comune a testare OpenBilancio, una visualizzazione dei dati che vuole essere comprensibile da tutti.

Nel marzo 2012 è stata rilasciata la seconda versione della licenza progettata per i dati delle Pubbliche Amministrazioni: Italian Open Data License, indicata come IODL v2.0 [6], priva del vincolo condividi-allo-stesso-modo e con la sola richiesta di attribuzione della fonte per il riutilizzo dei dati. [1]

## 2.2 Impianto Normativo

Il percorso che ha portato all'attuale normativa in ambito Open Data è lungo e intreccia le sue radici con i concetti di *libertà di informazione* e *trasparenza della pubblica amministrazione* ed ha interessato i governi di differenti nazioni.

- 4 luglio 1966, viene emanato negli USA il **Freedom of Information Act** (FOIA) [7], “atto per la libertà di informazione”, una legge sulla libertà di informazione, che impone alle pubbliche amministrazioni una serie di regole per permettere a chiunque di sapere come opera il Governo federale. Il FOIA ha aperto a giornalisti e studiosi l'accesso agli archivi di Stato statunitensi, a molti documenti riservati e coperti da segreto di Stato, di carattere storico o di attualità. Il provvedimento è stato un punto importante per garantire la trasparenza della pubblica amministrazione nei confronti del cittadino, il diritto di cronaca e la libertà di stampa dei giornalisti.
- 24 novembre 2000, vengono pubblicate nella Gazzetta Ufficiale le **Disposizioni per la delegificazione di norme e per la semplificazione di procedimenti amministrativi** (legge 340/2000) [8]

*Le pubbliche amministrazioni di cui all'articolo 1, comma 2, del decreto legislativo n. 29 del 1993 hanno accesso gratuito ai dati contenuti in pubblici registri, elenchi, atti o documenti da chiunque conoscibili. [8]*

- 17 novembre 2003, viene approvata dal Parlamento Europeo e dal Consiglio la **Direttiva relativa al riutilizzo dell'informazione del**

**settore pubblico** (Direttiva PSI) [9] che incoraggia gli enti pubblici degli stati membri a massimizzare il potenziale dell'informazione rendendo disponibili e favorendo il riuso dei documenti, attraverso indici on line e licenze standard.

- 7 marzo 2005, viene approvato il decreto legislativo **Codice dell'amministrazione digitale** (C.A.D.) [10].
- 24 gennaio 2006, l'attuazione italiana della direttiva europea avviene con il **Decreto legislativo 24 gennaio 2006 relativo al riutilizzo di documenti nel settore pubblico** [11]. Il provvedimento è stato predisposto dal Ministro per le Politiche comunitarie e da quello per l'innovazione e le tecnologie, in accordo con i dicasteri degli Affari Esteri, Giustizia, Economia e Finanze, Funzione pubblica. Esso definisce che il titolare del dato predisponga le licenze standard per il riutilizzo e le renda disponibili, ove possibile in forma elettronica, sui propri siti istituzionali. Gli Enti pubblici possono richiedere un compenso in denaro: in questo caso hanno l'obbligo di fissare e pubblicare le tariffe e le relative modalità di versamento da corrispondere a fronte delle attività, individuandole sulla base dei costi effettivi sostenuti dalle Amministrazioni e aggiornato ogni due anni; questo compenso comprende i costi di raccolta, di produzione, di riproduzione e diffusione dei dati, maggiorati nel caso di riutilizzo per fini commerciali. È fatto divieto di esclusività sull'utilizzo dei dati.
- 12 giugno 2008, il Garante per la Protezione dei Dati Personalini, in corrispondenza dell'avvio dell'Operazione Trasparenza, ha sottolineato la necessità di predisporre *“idonei accorgimenti volti a consentire forme proporzionate di consultabilità dei dati, prevedendo in particolare elenchi relativi a singole amministrazioni, consultabili distintamente (anche sulla base di eventuali link a siti web delle amministrazioni medesime), senza che per gli utenti sia possibile modificarli agevolmente o reperire direttamente dati mediante motori di ricerca”* [12].
- 31 ottobre 2009, viene pubblicato in Gazzetta Ufficiale il Decreto legislativo n.150/2009, Attuazione della legge 4 marzo 2009, n. 15, in materia di **ottimizzazione della produttività del lavoro pubblico e di efficienza e trasparenza delle pubbliche amministrazioni** (legge Brunetta) [13]. Il Decreto indica le linee guida rispetto alle quali articolare il riordino della pubblica amministrazione nella direzione della produttività, dell'efficienza e della trasparenza. Le diretrici individuate sono in particolare: il merito, le nuove norme sull'ordinamento del lavoro

pubblico, il sistema sanzionatorio e disciplinare e, per quanto riguarda le performance, il ciclo di gestione, la trasparenza e rendicontazione, la misurazione e valutazione.

- 20 ottobre 2011, viene pubblicato dal Ministero per la Pubblica Amministrazione e l'Innovazione il **Vademecum sull'OpenData**[14]
- 22 giugno 2012, viene approvato il Decreto Legge **Misure urgenti per la crescita del Paese** (Decreto Sviluppo) [15].

*Art. 18*

1. *La concessione delle sovvenzioni, contributi, sussidi ed ausili finanziari alle imprese e l'attribuzione dei corrispettivi e dei compensi a persone, professionisti, imprese ed enti privati e comunque di vantaggi economici di qualunque genere [...] ad enti pubblici e privati, sono soggetti alla pubblicità sulla rete internet, ai sensi del presente articolo e secondo il principio di accessibilità totale [...]*
2. *Nei casi di cui al comma 1 ed in deroga ad ogni diversa disposizione di legge o regolamento, nel sito internet dell'ente obbligato sono indicati:*
  - (a) *il nome dell'impresa o altro soggetto beneficiario ed i suoi dati fiscali;*
  - (b) *l'importo;*
  - (c) *la norma o il titolo a base dell'attribuzione;*
  - (d) *l'ufficio e il funzionario o dirigente responsabile del relativo procedimento amministrativo;*
  - (e) *la modalità seguita per l'individuazione del beneficiario;*
  - (f) *il link al progetto selezionato, al curriculum del soggetto incaricato, nonché al contratto e capitolato della prestazione, fornitura o servizio.*
3. *Le informazioni [...] sono riportate, con link **ben visibile** nella homepage del sito, nell'ambito dei dati della sezione «Trasparenza, valutazione e merito» [...], che devono essere resi di facile consultazione, accessibili ai motori di ricerca ed in formato tabellare aperto che ne consente l'esportazione, il trattamento e il riuso [...]*

5. *A decorrere dal 1° gennaio 2013, per le concessioni di vantaggi economici successivi all'entrata in vigore del presente decreto-legge, la pubblicazione ai sensi del presente articolo costituisce condizione legale di efficacia del titolo legittimante delle concessioni ed attribuzioni di importo complessivo superiore a mille euro nel corso dell'anno solare previste dal comma 1, e la sua eventuale omissione o incompletezza è rilevata d'ufficio dagli organi dirigenziali e di controllo, sotto la propria diretta responsabilità amministrativa, patrimoniale e contabile per l'indebita concessione o attribuzione del beneficio economico. La mancata, incompleta o ritardata pubblicazione è altresì rilevabile dal destinatario della prevista concessione o attribuzione e da chiunque altro abbia interesse, anche ai fini del risarcimento del danno da ritardo da parte dell'amministrazione, [...]*
7. *Dall'attuazione del presente articolo non derivano nuovi o maggiori oneri a carico della finanza pubblica e alle attività previste si farà fronte con le risorse umane, finanziarie e strumentali disponibili a legislazione vigente. [15]*

Con l'entrata in vigore di questo Decreto Legge inizia a prendere vita “Dentro il bilancio”, un progetto che amplia quello che precedentemente era *OpenFatture*.

- 18 ottobre 2012, vengono approvate le “Ulteriori misure urgenti per la crescita del Paese” [16] che modificano l'articolo 52 del **Codice dell'Amministrazione Digitale** (C.A.D.) [10]. In forza di questa modifica, gli Open Data sono diventati un “obbligo di legge”, nel rispetto della privacy e delle normative a tutela del segreto statistico, brevetti, diritto d'autore, ...

*L'accesso telematico a dati, documenti e procedimenti e il riutilizzo dei dati e documenti e' disciplinato dai soggetti di cui all'articolo 2, comma 2, secondo le disposizioni del presente codice e nel rispetto della normativa vigente. Entro 120 giorni dall'entrata in vigore del presente decreto-legge, le pubbliche amministrazioni pubblicano nel proprio sito web, all'interno della sezione <Trasparenza, valutazione e*

*merito» i regolamenti che disciplinano l'esercizio della facoltà di accesso telematico e il riutilizzo, compreso il catalogo dei dati e dei metadati in loro possesso. [16]*

Le Pubbliche Amministrazioni devono promuovere “progetti di elaborazione e di diffusione dei dati pubblici di cui sono titolari”, nonché assicurarne la pubblicazione “in formati aperti”, al fine di “valorizzare e rendere fruibili” i dati stessi. Inoltre viene specificato che a partire dal 19 marzo 2013 tutti i dati e documenti che le pubbliche amministrazioni pubblicano con qualsiasi modalità, senza l'espressa adozione di una licenza d'uso, si intendono rilasciati come dati aperti (Open Data by default).

- 6 novembre 2012, vengono approvate le **Disposizioni per la prevenzione e la repressione della corruzione e dell'illegalità nella pubblica amministrazione** (legge anticorruzione) [17].
- 14 marzo 2013, entra in vigore il Decreto Legislativo **Riordino della disciplina riguardante gli obblighi di pubblicità, trasparenza e diffusione di informazioni da parte delle pubbliche amministrazioni** [18] con il quale la l'Amministrazione comunale provvede all'adeguamento delle pagine dei propri siti web.
- 26 giugno 2013, è stata pubblica la nuova direttiva del parlamento e del consiglio europeo relativa al riutilizzo dell'informazione del settore pubblico [19]: l'Open Data si estende a biblioteche, archivi e musei. Apertura diventa in pratica sinonimo di riusabilità.

### 2.2.1 Leggi Regionali

Diverse amministrazioni regionali hanno compreso il valore e l'importanza dell'Open Data e hanno quindi deciso di dotarsi di una Legge in materia [20]. Al momento tale processo è ancora in fase evolutiva e la situazione è piuttosto variegata nella nostra penisola (figura 2.3).

- **Lombardia:** Nell'Agenda Digitale di Lombardia per il periodo 2012-2015 (DGR 2585/2011) è inclusa l'Area Prioritaria d'Intervento “Patrimonio Informativo Pubblico”, (paragrafo 2.4). Con la delibera 2094/2012 si approvano i “Criteri generali per l'Open Data”. Con La LR N. 7/2012 - art. 52 Accessibilità e valorizzazione del patrimonio informativo pubblico - la Giunta regionale, adotta determinazioni in ordine alla definizione delle basi di dati regionali da rendere disponibili a

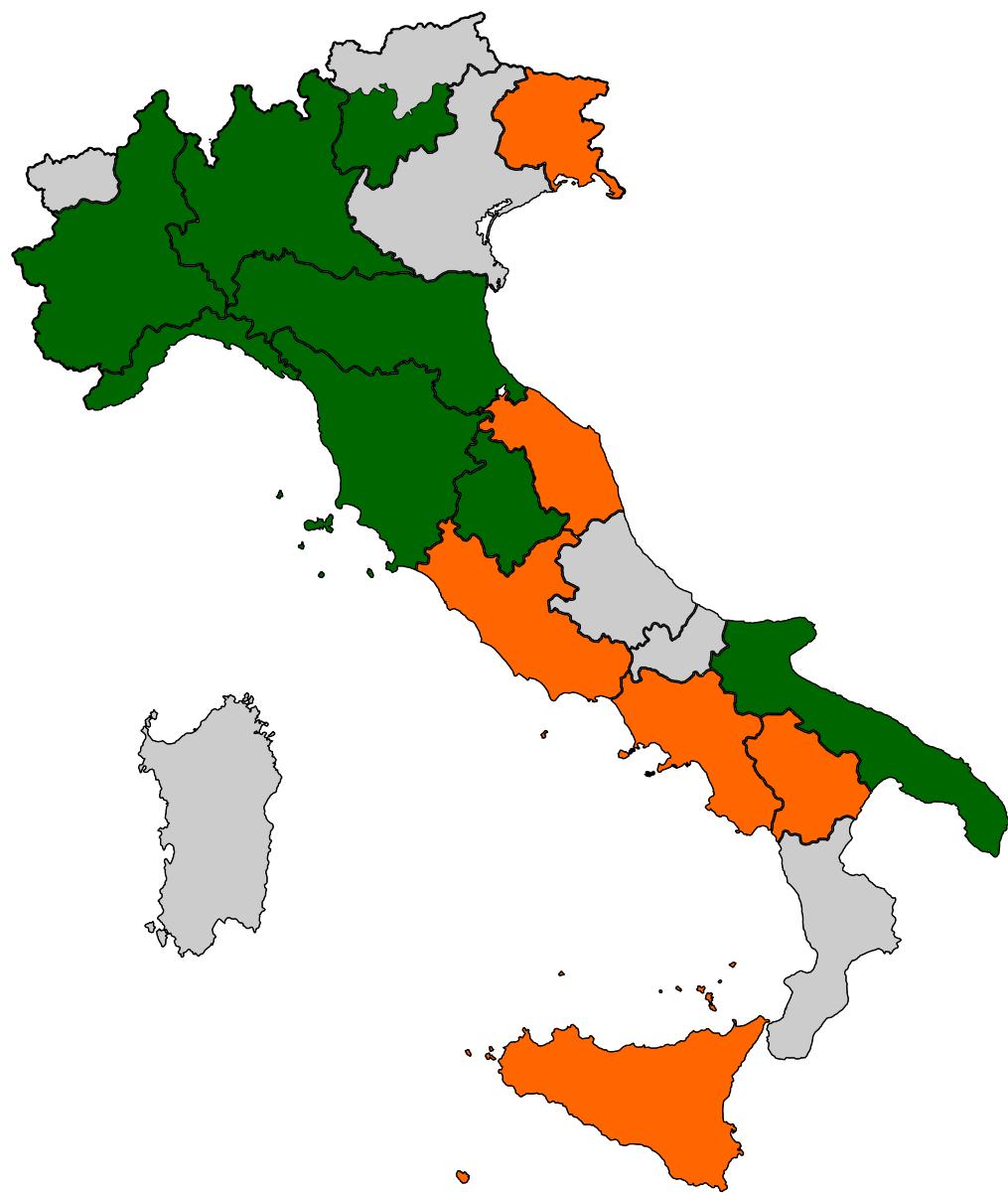


Figura 2.3: Situazione normativa nelle Regioni italiane su Open Data:  
**Verde:** legge approvata.  
**Arancione:** disegno o proposta di legge in discussione.

cittadini ed imprese in formato aperto, nonché le modalità di fornitura, senza oneri per finanza regionale, dei dati rilevati nell'esercizio delle attività da parte di concessionari di servizi pubblici. Il DG 6115/2012 definisce il Modello di Governance, stabilendo ruoli e modalità d'intervento per rendere disponibili i dati del patrimonio informativo regionale come Open Data ed il metodo con cui sono pianificati gli interventi, e il DGR n. 4324/2012 costituisce le linee guida per gli EELL che ripropongono, principalmente ai Comuni Lombardia, il modello organizzativo e le scelte operative adottate da Regione per la gestione degli Open Data.

- **Umbria:** Nell'Agenda Digitale dell'Umbria (DGR 397/2012) è inclusa tra le linee prioritarie d'azione quella dedicata agli Open Data. Nella legge regionale n.8/2011 sulla semplificazione e sullo sviluppo dell'Amministrazione digitale l'art.15 prevede la promozione dei dati pubblici aperti. Con D.G.R. 1551/2012 è stato approvato il Disciplinare dei criteri generali per l'Open Data in Umbria, vincolante per l'Amministrazione regionale e di riferimento per gli enti del territorio (che sono invitati a pubblicare i loro dati nel repertorio regionale).
- **Toscana:** Il Programma regionale per la promozione e lo sviluppo dell'amministrazione elettronica e della società dell'informazione e della conoscenza (DGR 104/2012), si pone come agenda digitale regionale prevede un'azione specifica relativa all'apertura dei dati. La successiva evoluzione normativa in materia di amministrazione digitale consta di tre Leggi Regionali (LR 1/2004, LR 40/2009 e LR 54/2009). La LR 40/2009 (e s.m.i.) in riferimento alla trasparenza e all'attività amministrativa, ambito più ampio e precedente rispetto allo specifico degli Open Data, tratta il diritto di accesso. La LR 54/2009 pone riferimenti importanti alla politica di Open Data, tra questi il principio per cui lo sviluppo dell'amministrazione digitale deve avvenire con il coinvolgimento degli enti del territorio. L'attuazione degli Open Data è definita con DGR 23/2013, con cui l'apertura dei dati si pone come necessario passaggio culturale ed elemento per arrivare anche all'apertura dei servizi.
- **Emilia Romagna:** Con DGR 52/2011 sono state approvate le nuove Linee Guida al Piano Telematico regionale 2011-2013 (PiTER), documento che rappresenta l'Agenda Digitale regionale. La Giunta della Regione Emilia-Romagna ha successivamente approvato con delibera 1587/201 il Programma Operativo 2011 del PiTER con il quale è stata data concretezza agli obiettivi strategici enunciati nelle Linee Guida, consolidando le azioni intraprese e future in un progetto operativo sottoposto a monitoraggio e valutazione. Con successiva DGR 2080/2012,

Emilia approva le “Linee guida relative al riutilizzo e messa a disposizione in Open Data dei dati pubblici dell’Amministrazione Regionale con i quali condivide il valore e la funzione attribuiti dall’Unione Europea alle informazioni pubbliche, e la consapevolezza delle conseguenze positive a livello di trasparenza e di partecipazione attiva dei cittadini alle attività e alla vita delle Pubbliche Amministrazioni, a seguito della diffusione di tali informazioni.

- **Piemonte:** È un esempio significativo con il processo che parte con l’emissione di linee guida nel 2009 (DGR 11679/2009) che definiscono i processi di riuso, e un aggiornamento delle stesse nel 2010 (DGR 1109/2010). Segue l’adozione della Legge Regionale in materia (LR 24/2011) e relativo regolamento attuativo con la quale Regione si vincola ad assicurare la disponibilità, la gestione, l’accesso, la trasmissione, la conservazione e la fruibilità dei dati in modalità digitale. Detto regolamento si distingue rispetto ad altri casi perché stabilisce la possibilità per l’utente esterno, di qualunque tipologia (privato cittadino, ente privato etc.), di presentare formale richiesta per il rilascio in modalità open di un determinato dataset di suo interesse, e delinea il processo interno dell’Ente per l’adempimento della richiesta. L’utente che non vedesse accolta la propria richiesta senza giustificato motivo ha inoltre la possibilità di sporgere formale reclamo.
- **Puglia:** La LR 20/2012 “Norme sul software libero, accessibilità di dati e documenti e hardware documentato” tratta i temi dell’Open Data specificatamente all’art. 6 “Riutilizzo dei documenti e dei dati pubblici”. Le prime Linee di indirizzo in materia di OD sono state oggetto della DGR 2183/2012.
- **Provincia Autonoma di Trento:** Il percorso normativo ha inizio con l’emissione delle DGR 1510/2010 e DGR 1510/2011, delibere inerenti alla definizione di una Strategia di Legislatura in materia di innovazione e servizi ICT. Con la prima si introducono gli Open Government Data tra le infrastrutture abilitanti l’innovazione quale condizione necessaria per lo sviluppo dell’innovazione nei servizi; nella seconda si esplicita il l’obiettivo di valorizzare il patrimonio informativo della PA al fine di aumentare la trasparenza, semplificare e rendere più efficiente il sistema pubblico e fornire risorse informative ai settori produttivi. Segue l’adozione della LP 16/2012 sulla promozione della società dell’informazione e amministrazione digitale. Con DGR 2858/2012 sono approvate le linee guida per la diffusione e riutilizzo di dei dati pubblici.

- **Liguria:** La Regione ha fatto proprie le politiche di pubblicazione dei dati pubblici regionali, in linea con il principio di trasparenza, in considerazione del grande valore attribuito all'economia immateriale e alla luce delle prescrizioni della legge regionale 42/06 “Istituzione del Sistema informativo integrato regionale per lo sviluppo della Società dell'informazione in Liguria.
- **Regione Lazio** Legge Regionale n. 7/2012 - Disposizioni in materia di dati aperti e riutilizzo di informazioni e dati pubblici e iniziative connesse”.  
Proposta di legge n. 200/IX - Norme in materia di pubblicazione e riutilizzo dei dati e delle informazioni delle pubbliche amministrazioni regionali
- **Regione Basilicata** Proposta di legge - Disposizioni in materia di accesso, pubblicazione e riutilizzo dei documenti e dei dati pubblici dell'Amministrazione regionale in formato aperto tramite la rete internet
- **Regione Sicilia** Disegno di legge - Norme in materia di pubblicazione tramite la rete internet e di riutilizzo dei documenti e dei dati della pubblica amministrazione regionale e locale
- **Regione Friuli Venezia Giulia** Legge Regionale n. 9/2011 - Disciplina del sistema informativo integrato regionale del Friuli Venezia Giulia
- **Regione Campania** Proposta di legge - Legge sulla trasparenza amministrativa e l'accesso e il riutilizzo dei dati di titolarità regionale
- **Regione Marche** Proposta di legge n. 256/12 - “Disposizioni in materia di pubblicazioni e riutilizzo dei dati e delle informazioni dell'amministrazione regionale”

### Regione Lombardia

La Regione Lombardia ha approvato, con la Delibera di Giunta Regionale 2904 del 11/01/2012 i Criteri generali per l'Open Data [21] nei quali delibera di realizzare la piattaforma [dati.lombardia.it](http://dati.lombardia.it) sulla quale caricare, in una prima fase, i dati già pubblicati sui portali della Regione e sul sistema SIREG. Delibera inoltre di pubblicare dati aggregati di cui la Regione è proprietaria, il tutto con una licenza che concede all'utente la possibilità di riprodurre, distribuire, trasmettere e adattare liberamente i dati, anche a scopi commerciali, a condizione che venga citata la fonte, la licenza che è stata scelta è la Italian Open Data License v.2.0 (IODL 2.0).

### Comune di Vigevano

Il Comune di Vigevano il 23 maggio 2013 ha deliberato l'**Adesione alle “Linee guida Open Data per gli enti locali”, di Regione Lombardia e sperimentazione congiunta in materia** [22]:

1. *di aderire alle “Linee Guida per gli Enti Locali” facendo propri i relativi allegati e usufruendo per la diffusione dei dati individuati come riutilizzabili del portale dati.lombardia.it;*
2. *d'impegnarsi, coerentemente con quanto previsto dalle Linee Guida, a licenziare i dati – quale regola generale - con licenza IODL 2.0, optando a favore di altre licenze solo ove ricorrano giustificati motivi; la licenza scelta dovrà comunque consentire il riutilizzo dei dati pubblicati anche per fini di lucro e commerciali, come richiesto dai “Criteri generali per l'Open Data”, Allegato 1 alla DGR 2904 dell'11/1/2012;*
3. *di avviare una collaborazione con Regione Lombardia / Lombardia Informatica sui temi della interoperabilità dei Open Data pubblicati in portali diversi tramite l'uso principale di soluzioni Open Source con l'obiettivo, fra gli altri, di consentire un più semplice accesso ai dati del futuro portale Open Data del Comune di Vigevano anche attraverso l'accesso dal portale della Regione Lombardia* [22]

# Capitolo 3

## Piattaforme Tecnologiche

### 3.1 Socrata (Open Source)

Socrata è una piattaforma sviluppata dall'omonima società con sede negli Stati Uniti d'America. Socrata è un servizio Software as a service. Dal 10 aprile 2013 è disponibile anche in una versione Community Edition rilasciata come Open Source al fine di promuovere uno standard intorno ai dati aperti e far crescere la comunità.

Il Socrata Open Data Server e le componenti che lo compongono sono stati resi disponibili con licenza Apache. L'architettura del sistema può quindi essere suddivisa in una serie di componenti separate che la compongono come mostrato in figura 3.1.

SODA 2 fornisce quindi l'accesso ai dati puri attraverso SoQL (SODA Query Language) e delle semplici API per aggiornarli. La scrittura avviene quindi attraverso un percorso che prevede il seguente percorso:

- ① Un'applicazione o un processo inizia un'operazione sul SODA Server.
- ② La richiesta viene trasformata in una serie di operazioni più elementari chiamate "mutations" e vengono passate al Data Coordinator in modo da essere eseguite sul database (Postgres).
- ③ Dopo che il Data Coordinator ha finito, il Secondary Watcher si sveglia e guarda se ci sono cambiamenti nel database che non sono ancora stati sincronizzati.
- ④ L'adattatore per l'archivio secondario (in questo caso Elastic Search) importa i dati dal primario. Ciò permette in caso di guasti di re-sincronizzare il database.

Mentre la lettura avviene attraverso:

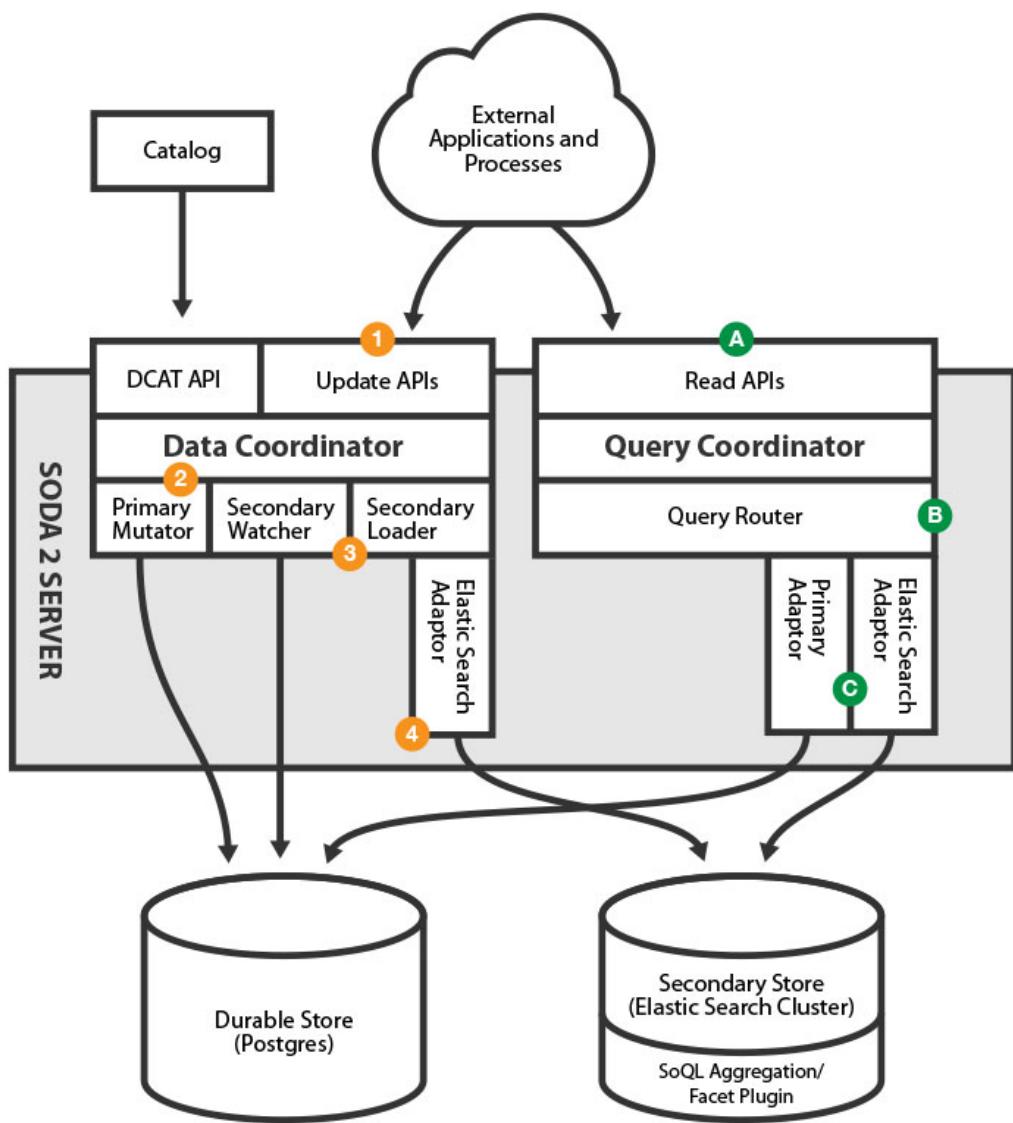


Figura 3.1: Architettura di Socrata

- Ⓐ Un'applicazione invia una richiesta al server SoQL. Questa viene analizzata come descritto serctitto in `soql-reference`.
- Ⓑ Il Query Coordinator determina a quale database inviare la richiesta (in futuro ci portano essere diversi archivi secondari).
- Ⓒ Il Query Coordinator passa fuori la query all'adattatore appropriato che esegue la query sull'archivio corretto e restituisce l'appropriato payload C-JSON.

Ad oggi sono state rese disponibili con licenza Open Source solo alcune parti:

- **soql-reference**, Implementazione di riferimento del linguaggio di interrogazione SoQL.
- **socrata-http**, Toolkit per la creazione di servizi HTTP.
- **soql-es-adapter**, ElasticSearch Secondary Store per SoQL Data Service.
- **socrata-csv**, Un sottile involucro Scalaish intorno ad opencsv.
- **socrata-utils**, Classi-Utility utilizzate in tutto il Socrata Open Data Server.
- **data-coordinator**, Coordina la distribuzione degli aggiornamenti tra archivi di dati primari e secondari.

La Community Edition condivide lo stesso core della versione proprietaria, ma al momento manca completamente il front-end per gestire comodamente i dataset che è invece presente nella versione SaaS. Secondo la Roadmap intorno all'Agosto 2013 era previsto il rilascio della versione beta finale, compresa della documentazione e degli strumenti che avrebbero permesso di installare e rendere funzionante l'intero Open Data Server.

Socrata offre sia numerose API per gestire i dataset, che strumenti di visualizzazione che mostrano i dati caricati attraverso un sistema di preview tabellare dotato di filtri avanzati, oppure su diverse tipologie di mappe in caso di dati geolocalizzati. Permette inoltre di esportare i dati in vari formati (CSV, XLS, XLSX, XML, JSON).

In Italia Socrata è stata usata con successo dalla Regione Lombardia con un centinaio di dataset pubblicati.

## 3.2 CKAN

CKAN (Comprehensive Knowledge Archive Network) è un potente sistema DMS (Data Management System, ovvero un sistema di gestione dei dati) open source. Esso ha lo scopo di rendere i dati accessibili permettendo la loro archiviazione e distribuzione attraverso una comoda interfaccia web che ne semplifica la pubblicazione, la condivisione, la ricerca e l'utilizzo.

Il funzionamento dell'architettura del sistema è mostrato in figura 3.2.

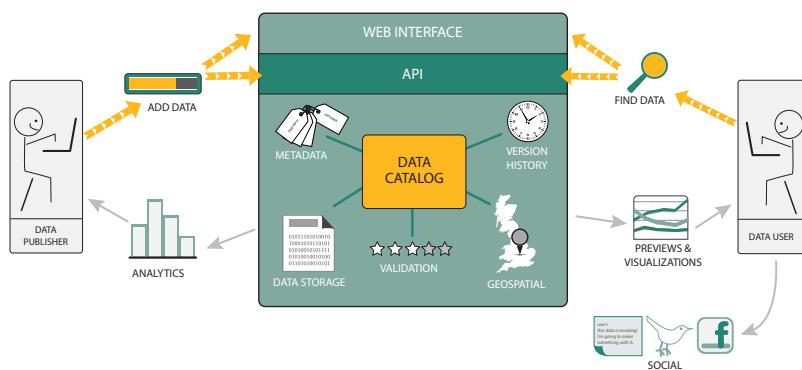


Figura 3.2: Funzionamento Architettura CKAN

CKAN offre sia numerose API per gestire i dataset, che strumenti di visualizzazione che ne permettono una rapida anteprima. La piattaforma supporta molto bene i dati in formato tabellare, siano essi semplici CSV o XLS, e la loro visualizzazione può essere fatta in semplice tabella, come grafico a punti o su una mappa in caso di dati geolocalizzati.

CKAN è rivolto ai data publishers, come governi, imprese o organizzazioni nazionali/regionali, che vogliono rendere disponibili e facilmente fruibili i propri dati. A ciò va aggiunta la possibilità di integrare facilmente i cataloghi di altri portali sviluppati anch'essi in CKAN.

Il suo codice è mantenuto dalla Open Knowledge Foundation ed è utilizzato come piattaforma per gestire cataloghi pubblici da siti come [datahub.io](http://datahub.io), [catalog.data.gov](http://catalog.data.gov), [data.gov.uk](http://data.gov.uk) e [PublicData.eu](http://PublicData.eu).

CKAN può essere installato gratuitamente su un server di proprietà dello stesso soggetto che intende pubblicare i dati, oppure è possibile acquistare il servizio di hosting dedicato gestito dallo stesso produttore di software (quindi in modalità SaaS).

CKAN ha iniziato il suo sviluppo nel 2007, ma solo nel maggio 2010 è stata pubblicata la sua prima versione (1.0), alla quale sono seguiti aggiornamenti minori fino alla versione 1.8, che nel maggio 2013 è stata superata dalla nuova versione 2.0. Attualmente è in sviluppo la versione 2.2.

## Capitolo 4

# Progettazione portale Open Data a Vigevano

### 4.1 Descrizione del Comune

Vigevano è un comune italiano di circa 60.000 abitanti della provincia di Pavia in Lombardia; dista circa 38 km da Pavia e 36 km dal centro di Milano. Il comune è il secondo della provincia per numero di abitanti dopo il capoluogo e primo per superficie. La città si trova nei pressi del Parco naturale lombardo della Valle del Ticino e la fitta vegetazione e i canali che attraversano la città offrono al visitatore magnifici scorci e oasi naturali.

La città ha la sua origine come luogo fortificato longobardo dove ora sorge il cortile del castello, successivo è lo sviluppo del borgo esterno, oggi occupato dalla Piazza Ducale.

A partire dal 1198 si trasforma in un libero Comune, mentre nel 1277 la storia delle potenti famiglie milanesi dei Visconti e degli Sforza si lega a quella di Vigevano. Grazie all'opera di Luchino Visconti e di Ludovico Sforza, detto il Moro, tra il XIV e il XV secolo, il borgo di Vigevano inizia la sua trasformazione in residenza estiva: il castello viene adibito a dimora di prestigio, grazie all'opera di artisti come Bramante, e la Piazza Ducale, liberata da case ed edifici, diviene il regale atrio d'ingresso al castello.

Nel 1530 Vigevano ottiene il titolo di città con una propria sede vescovile.

#### 4.1.1 Servizio Informatico Comunale

Durante la progettazione del portale Open Data per il Comune di Vigevano, il Dipartimento di Informatica Sistemistica e Comunicazione (DISCo) dell'Università di Milano Bicocca si è relazionato con il responsabile del Servizio Informatico Comunale (SIC) **Oscar Baracchi**.

Il SIC è la struttura organizzativa cui compete la gestione del sistema informativo dell'Ente (il Comune di Vigevano), attraverso alcune decine di server con supporto ad oltre 350 utenti utilizzatori di apparecchiature informatiche, appositamente configurate per l'accesso alle informazioni presenti nelle banche dati dell'Ente ed alle procedure per l'erogazione di servizi ai cittadini ed alle imprese.

Il Servizio cura la pianificazione, lo sviluppo, il mantenimento, il coordinamento ed il controllo di tutte le iniziative ed attività che afferiscono i sistemi informativi comunali, le infrastrutture informatiche, la rete trasmissione dati, la conduzione di progetti nel campo dell'ICT di notevole complessità tecnologica ed organizzativa.

In particolare il Servizio ha la responsabilità:

- della pianificazione strategica e tattica per tutti gli aspetti relativi all'utilizzo dell'ICT (Information & Communication Technology) nell'Amministrazione Comunale;
- dello sviluppo di nuove iniziative ed attività per il miglioramento del grado di efficienza ed efficacia dell'azione amministrativa tramite l'utilizzo di opportuni sistemi informativi, infrastrutture informatiche e telematiche;
- del mantenimento in efficienza dei sistemi informativi comunali, delle infrastrutture ed apparecchiature informatiche, di rete trasmissione dati utilizzate nell'Amministrazione comunale;
- di centro acquisitore per quanto riguarda le forniture di beni e servizi che rientrino nell'ambito dell'ICT;
- del supporto nella gestione di progetti ed attività in cui sia presente una componente informatica o telematica;
- della consulenza e supporto alle unità organizzative dell'Amministrazione comunale su aspetti che attengono in qualche misura all'ICT;
- del supporto nella gestione di progetti ed attività di e-government e sovracomunali;

- del supporto nei rapporti con i fornitori di servizi, tecnologie e soluzioni;

Inoltre sono stati sviluppati e manutenuti in proprio, già da anni, numerosi applicativi con tecnologia web (Java-Jsp) su database MySql ed application server Tomcat/Apache. [23]

Il SIC lavora da tempo adottando soluzioni Open Source per le piattaforme e i servizi offerti dal Comune, inoltre ha manifestato un forte interesse alla pubblicazione di diverse informazioni in formato Open Data e alla realizzazione di una piattaforma che ne permetta la divulgazione.

## 4.2 Progettazione piattaforma

Con delibera 14 del 19/01/2012 il Comune di Vigevano, nell’ambito del Progetto InnoVi, cofinanziato dalla regione, ha adottato, in riuso dal Comune di Milano, la piattaforma **GIT** [22].

La piattaforma GIT, di proprietà della Pubblica Amministrazione italiana e realizzato attraverso un progetto (GIT “Gestione Intersetoriale del Territorio”) del programma ELISA del Dipartimento Affari Regionali realizzato da Comuni, è stata implementata al fine di integrare, relazionare e rendere interoperabili i dati in possesso di un ente per il territorio di competenza. GIT costituisce la dorsale sulla quale i dati possono transitare al fine di acquisire un maggior livello di accuratezza, attendibilità e completezza.

Pur preservando la peculiarità di poter rendere disponibili i dati grezzi, rispettando l’indicazione “ROW DATA NOW”, il processo di pubblicazione può essere realizzato in modo progressivo attraverso il trattamento dell’informazione per renderla più chiara, relazionata e dunque economicamente più vantaggiosa per il fruttore finale.

Gli Elementi conoscitivi del territorio sono stati catalogati per la piattaforma in tematismi:

- Soggetti che si muovono sul territorio e che agiscono con interazione diretta o intermediata.
- Oggetti presenti sul territorio che costituiscono strumenti dell’azione dei Soggetti, Patrimonio economico rappresentativo del valore stesso dei luoghi e contenitori in grado di generare lavoro per la loro conservazione e manutenzione.
- Ambiente che caratterizza il territorio amministrato e costituisce un importante riferimento per le regole del vivere civile per la valorizzazione del patrimonio naturale e di manufatti da gestire, valorizzare e preservare.

- toponomastica e geolocalizzazione dei luoghi degli oggetti, e delle persone che rappresenta la mappatura e il riferimento per la ricerca, la misurazione e la relazione delle informazioni territoriali di interesse.

Attraverso queste “chiavi” di conoscenza primarie del territorio, i dati delle singole fonti vengono assunti come elementi descrittori, analizzati, diagnosticati e resi disponibili in modo integrato per la consultazione o per l'utilizzo attraverso canali consultazione puntuale, di estrazione e pubblicazione massiva.

La piattaforma GIT è costituita da componenti software realizzati per assicurare determinate funzioni di trattamento delle informazioni, oltre che da servizi di navigazione trasversale fra le fonti e da servizi verticali specialistici, anche cartografici.

Ma sono le funzioni di trattamento delle fonti dati e di orchestrazione, implementate dal componente Controller a poter dare un valore aggiunto al processo di apertura dei dati, si descrivono ora brevemente le peculiarità e le funzioni principali:

- sistema di reperimento e acquisizione dell'informazione attraverso varie modalità e protocolli. I sistemi operazionali e/o gli utenti possono depositare su repository relazionali / ftp / file system i loro dati e la piattaforma si preoccuperà di importare tali dati con periodicità stabilità. Oppure possono essere implementate e configurate delle sonde applicative che controllano lo stato di variazione aggiornamento di una determinata fonte dati (relazionale e non) al fine di importare le variazioni nel sistema GIT.
- Catalogo configurabile e programmabile tramite API di diagnostiche sulle informazioni oggetto di importazione, Le diagnostiche sono uno strumento da asservire al processo di bonifica di un dato per valutarne il grado di qualità e definire dunque eventuali step successivi. Le diagnostiche sono di due tipi: controllo e confronto; il sistema valuta la qualità del dato in termini di obbligatorietà di attributi informativi, formati e codifiche (controllo) e ne valuta inoltre le anomalie rispetto ad altre fonti informative (confronto). Un esempio di questa seconda tipologia di diagnostica è il “controllo del numero di famiglie in un immobile rispetto al numero di unità abitative e al numero di contratti di fornitura della rete elettrica o idrica”. Le diagnostiche di confronto, utilizzate in GIT al fine di indagine e ricerca di “anomalie”, può essere sfruttata nella generazione di relazioni interne di tipo appunto semantico.
- Sistema di correlazione delle fonti dati basato sui tematismi Soggetto, Via, Civico, Oggetto Territoriale e Fabbricato. Attraverso un sistema

di metadati vengono caratterizzati gli attributi di un archivio in modo da poter attribuire ad ognuno una categoria e un tipo. Per esempio vengono classificati tutti gli archivi che si riferiscono ad un immobile e all'interno di ogni archivio vengono identificati gli attributi territoriali (via, civico, coordinate catastali ecc.). Attraverso algoritmi standard di ricerca sql, attraverso l'utilizzo di algoritmi dinamici per la ricerca delle similarità e di algoritmy fuzzy ogni record di una fonte informativa viene correlato ad altri di altre fonti con un determinato grado di attendibilità. Queste correlazioni sono alla base di alcuni algoritmi diagnostici per la ricerca dell'evasione fiscale e alla base del sistema di navigazione dinamico GIT.

- Trattamento di dati cartografici, anche all'interno della correlazione delle informazioni, oltre che negli algoritmi diagnostici al fine di collegare il maggior numero di informazioni alfanumeriche ad un riferimento geografico. La presenza in GIT del catasto terreni / fabbricati, della toponomastica e di altre fonti dati (concessioni, tributi) contenenti riferimenti geografici certi ha permesso di implementare algoritmi deduttivi che attribuiscono gli attributi di localizzazione a fonti dati altrimenti sprovviste di questa informazione. Questo permette potenzialmente di visualizzare graficamente all'interno di un WebGIS quasi tutte le fonti dati.
- Presenza di un catalogo dei dati cartografici dinamico in grado di dare la possibilità di ingessare e tipizzare qualsiasi fonte dati spaziale.

Il sistema GIT è quindi un ricco contenitore di informazioni alfanumeriche e grafiche correlate, periodicamente aggiornate, e che, per questo, è un candidato naturale ad alimentare dei dataset di dati da rendere disponibili con licenze open, opportunamente filtrati ed aggregati allo scopo di tutelare la privacy.

Di conseguenza si vuole realizzare una piattaforma che permetta di integrare le informazioni presenti sulla piattaforma GIT con il portale Open Data che verrà realizzato. Per fare ciò sono percorribili (come mostrato in figura 4.1) tre strade distinte:

1. Permettere alla piattaforma di accedere direttamente ai dati di GIT.
2. Effettuare periodicamente delle esportazioni dei dati da GIT per caricarle sulla piattaforma.
3. Effettuare l'aggiornamento dei dati sulla piattaforma ogni volta che vengono aggiornati su GIT in automatico.

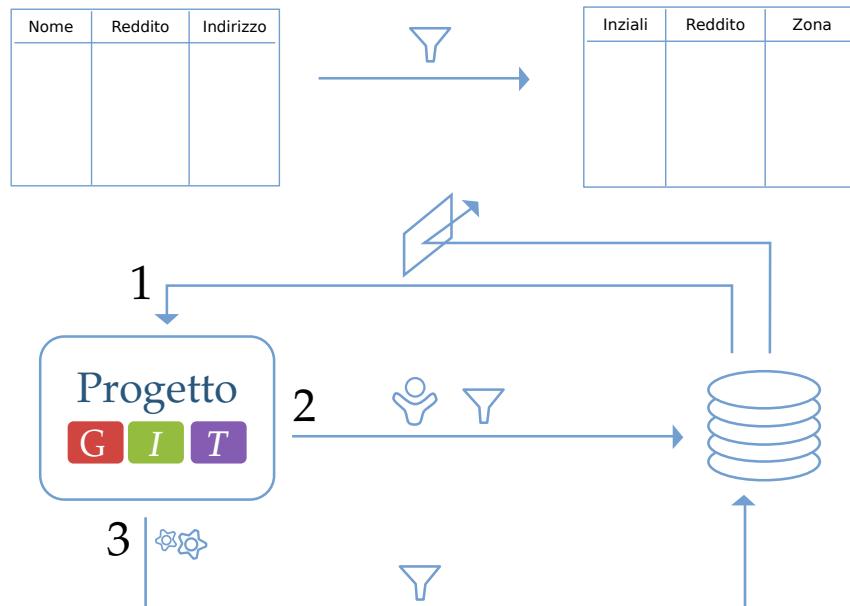


Figura 4.1: Possibili soluzioni di integrazione tra GIT e il portale

La prima soluzione presenta dei problemi di sicurezza, perché per realizzarla bisogna modificare GIT creando un’interfaccia che permetta l’interrogazione ai dati dall’esterno. Ciò può diventare un problema nel momento in cui si creano delle API di interrogazioni generiche e quindi in alcune occasioni si finisce a violare la privacy degli utenti.

La seconda soluzione può essere applicata nei casi in cui le informazioni hanno una frequenza di aggiornamento rada, ma diventa impraticabile nel momento in cui i dati vengono aggiornati con una frequenza rilevante, in quanto bisogna fare manualmente un aggiornamento ogni volta.

La soluzione vincente risulta quindi essere la terza. Questa comporta un controllo degli eventi sulla piattaforma GIT e l’esistenza di un sistema di aggiornamento automatizzato (senza intervento umano) sulla piattaforma Open Data. Nell’ottica del riutilizzo del software, questa componente deve essere in grado di controllare eventuali aggiornamenti di GIT, di annullare i dati e di inserirli nella piattaforma attraverso API.

Al fine di agevolare l’integrazione tra la piattaforma e applicazioni di terze parti si vuole rendere disponibile un **catalogo** contenente tutti i dataset e i relativi metadati. Con Metadati si intendono le informazioni riguardanti il singolo dataset quali il suo nome, la sua descrizione, il formato del file, la sua data di ultima modifica, il link a cui raggiungere la risorsa, ...

Si vuole inoltre rendere disponibile per ogni dataset il dizionario dei dati in esso contenuto. Ciò vuol dire che nel catalogo sarà presente anche l'informazione della tabella per ogni singolo dataset.

In questo modo è possibile interrogare il dataset sapendo a priori il tipo di dato che è contenuto in esso.

Si è scelto di utilizzare una piattaforma open source in accordo con il Comune di Vigevano e di conseguenza CKAN è stata la piattaforma sulla quale si è scelto di implementare i requisiti. Al fine di rendere compatibile le modifiche apportate alla piattaforma CKAN tra le varie versioni della stessa si è deciso di analizzarla come una “scatola nera”, senza mettere mano al codice sorgente di CKAN.

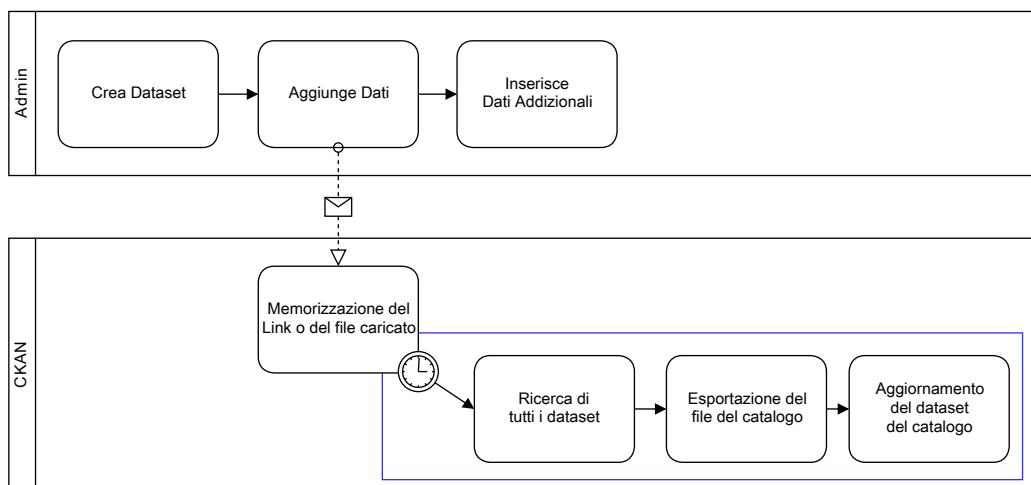


Figura 4.2: Workflow dell'aggiornamento del catalogo al momento dell'inserimento di un nuovo dataset

La progettazione del catalogo è quindi stata fatta come mostrato in figura 4.2 inserendo un trigger (una procedura che viene eseguita in maniera automatica in coincidenza di un determinato evento) nella base di dati, il quale si attiva ad ogni aggiornamento nel database della tabella contenente i dataset, eseguendo una query che restituisce il catalogo (facendo un join tra diverse tabelle) e la esporta in formato CSV. Il file contenente il catalogo è esso stesso parte integrante della piattaforma essendo a sua volta un dataset.

## 4.3 Integrazione con Regione Lombardia

La piattaforma `dati.lombardia.it` sviluppata da Regione Lombardia vuole porsi l’obiettivo di raccogliere i dataset contenenti i dati individuati come riutilizzabili dalle Pubbliche Amministrazioni interessate. È quindi nostro interesse che nella piattaforma regionale sia presente un elenco dei dataset pubblicati nel portale che vogliamo realizzare. Questo elenco deve rispecchiare l’attuale situazione e quindi deve essere aggiornato ad ogni modifica dei dataset presenti sulla nostra piattaforma oppure con una frequenza decisa arbitrariamente.

Si è quindi pensato di eseguire un importazione dei dati presenti nel catalogo della piattaforma CKAN realizzata per il Comune di Vigevano sull’installazione di Socrata utilizzata da Regione Lombardia. Al momento non esiste un automatismo che permetta il dialogo tra questa due piattaforme, ma esiste la possibilità di realizzarla sfruttando le informazioni riportate nella documentata sul sito web di Socrata dedicato agli sviluppatori alla pagina <http://dev.socrata.com/publishers/importing>.

L’importazione dei dati viene eseguita come un processo in due fasi. In primo luogo, il file di dati in sé viene caricato tramite l’API, e viene poi analizzato per determinare il tipo di dati che esso contiene. I risultati di tale analisi vengono restituiti, dopodiché è possibile apportare modifiche al modo in cui l’importatore interpreta i dati. A questo punto si invia il progetto di importazione risultante, e a questo punto inizia il processo di importazione vero e proprio.

Questa operazione andrà eseguita attraverso una procedura automatizzata esterna dal server che ospita la piattaforma CKAN.

# Capitolo 5

## Sviluppo della soluzione

Per prima cosa è stata installato su macchina virtuale il sistema operativo **Ubuntu 12.04 a 64 bit**, necessario per poter installare CKAN precompilato.

Da terminale sono state installate le dipendenze richieste e il pacchetto di CKAN versione 2.0:

```
~$ sudo apt-get update
~$ sudo apt-get install -y nginx apache2 libapache2-mod-wsgi libpq5
~$ wget http://packaging.ckan.org/python-ckan-2.0_amd64.deb
~$ sudo dpkg -i python-ckan-2.0_amd64.deb
```

Successivamente sono stati installati anche PostgreSQL (database relazionale ad oggetti) e Solr (piattaforma di ricerca):

```
~$ sudo apt-get install -y postgresql solr-jetty
```

Si è scelto di utilizzare un'unica istanza di CKAN e quindi un solo file di configurazione dello schema Solr. Di conseguenza si è installato il Java Development Kit e modificato il file di configurazione Jetty:

```
~$ sudo apt-get install solr-jetty openjdk-6-jdk
~$ sudo gedit /etc/default/jetty
 4: NO_START=0
16: JETTY_HOST=127.0.0.1
19: JETTY_PORT=8983
31: JAVA_HOME=/usr/lib/jvm/java-6-openjdk-amd64
~$ sudo gedit /etc/default/jetty
~$ sudo mv /etc/solr/conf/schema.xml /etc/solr/conf/schema.xml.bak
```

---

```
~$ sudo gedit /etc/default/jetty
~$ sudo ln -s /usr/lib/ckan/src/ckan/config/solr/schema-2.0.xml /etc
/solr/conf/schema.xml
```

A questo punto è stato avviato il server Jetty:

```
~$ sudo service jetty start
```

Si è poi configurato il database PostgreSQL creando l'utente `ckanuser` a cui è stata data la proprietà del database `ckan_default`. Successivamente si è modificato il valore di `sqlalchemy.url` nel file di configurazione `/etc/ckan/default/production.ini`. Infine si è inizializzato il database:

```
~$ sudo -u postgres createuser -S -D -R -P ckanuser
~$ sudo -u postgres createdb -O ckanuser ckan_default -E utf-8
~$ sudo gedit /etc/ckan/default/production.ini
46: sqlalchemy.url = postgresql://ckanuser:pass@localhost/ckan_default
~$ sudo ckan db init
```

## 5.1 Preview e Web API

Si è proseguiti con l'impostazione del **DataStore**, un componente che fornisce un database per l'archiviazione strutturata di dati insieme a delle potenti API accessibili dal WEB. Tutto ciò è perfettamente integrato nell'interfaccia web e nella gestione dei permessi di CKAN.

Per fare ciò si è abilitata l'estensione, separatamente, nel file di configurazione:

```
~$ sudo gedit /etc/ckan/default/production.ini
84: 84: ckan.plugins = datastore stats json_preview recline_preview
```

Dopodiché si è creato l'utente `ckanreaduser` con sola lettura al database `datastore_default` e si sono assegnati i relativi permessi. Inoltre si è modificato il file di configurazione con gli opportuni valori:

```
~$ sudo -u postgres createuser -S -D -R -l ckanreaduser
~$ sudo -u postgres createdb -O ckanuser datastore_default -E utf-8
~$ sudo gedit /etc/ckan/default/production.ini
50: ckan.datastore.write_url = postgresql://ckanuser:pass@localhost
/datastore_default
```

```
51:   sqlalchemy.url = ckan.datastore.read_url = postgresql://ckanreaduser
:pass@localhost/datastore_default
~$ . /usr/lib/ckan/default/bin/activate
(default) ~$ cd /usr/lib/ckan/default/src/ckan
(default) /usr/lib/ckan/default/src/ckan$ paster datastore set-permissions
postgres -c /etc/ckan/default/production.ini
```

## 5.2 Archiviazione dei file

Successivamente si sono impostati anche i componenti aggiuntivi **File-Store** e **File Uploads**, che contrariamente al FileStore, che fornisce un’archiviazione “blob” di interi file senza alcun accesso o interrogazione a parti di essi, fa in modo che CKAN si comporti come un database in cui i singoli elementi del dato sono accessibili e interrogabili. Se il file è memorizzato nel DataStore, si è in grado di accedere alle singole righe del dataset attraverso delle semplici web API, oltre ad essere in grado di effettuare interrogazioni sul suo intero contenuto.

Si è creata la directory per il salvataggio in locale e si è modificato il file di configurazione per impostarne il path. Successivamente si sono assegnati i permessi a CKAN sulla directory e infine si è impostato l'url e riavviato il server web:

```
~$ sudo mkdir -p /var/lib/ckan/default
~$ sudo gedit /etc/ckan/default/production.ini
122:   ofs.impl = pairtree
123:   ofs.storage_dir = /var/lib/ckan/default
~$ sudo chown www-data /var/lib/ckan/default
~$ sudo chmod u+rwx /var/lib/ckan/default
~$ sudo gedit /etc/ckan/default/production.ini
55:   ckan.site_url = http://localhost
~$ sudo service apache2 reload
```

## 5.3 Automazione analisi file caricati

Infine si è installato il **DataStorer**, che permette di aggiunge automaticamente al DataStore i dati che vengono inseriti in CKAN (linkati o caricati nel FileStore). Ciò viene fatto attraverso un'attività di analisi automatica e successiva aggiunta nel DataStore da un processo che viene eseguito in modo asincrono.

Si è quindi installato git, si sono poi recuperati i sorgenti del pacchetto e installati i requisiti:

```
~$ sudo apt-get install git-core
~$ sudo -s
~# . /usr/lib/ckan/default/bin/activate
(default) ~# pip install -e git+git://github.com/okfn/ckanext-datastorer
.egg=ckanext-datastorer
(default) ~# pip install -r /usr/lib/ckan/default/src/ckanext-datastorer
/pip-requirements.txt
(default) ~# exit
```

Infine per abilitare la preview e le API sui singoli file bisogna processarli ad ogni loro inserimento, e quindi si è configurato un Cron Jobs ogni 2 minuti:

```
~$ /usr/lib/ckan/default/bin/paster --plugin=ckanext_datastorer datastore
upload -c /etc/ckan/default/production.ini
~$ crontab -e
*/2 * * * * /usr/lib/ckan/default/bin/paster --plugin=ckanext_datastorer
datastore_upload -c /etc/ckan/default/production.ini >> /tmp/update_datastore
2>&1
```

## 5.4 Catalogo dati

Una volta che CKAN è stato popolato di dataset si è realizzato il catalogo dei dataset presenti. Per fare ciò è stato necessario recuperare le informazioni contenute nel database e quindi le tabelle prese in esame sono state **package**, **resource** e **resource\_group**. La loro struttura è mostrata in figura 5.1:

Le informazioni sono presenti in tabelle differenti, ed è quindi stato necessario fare un JOIN tra le 3 tabelle e successivamente selezionare solo i campi di nostro interesse. Inoltre si sono tradotte le intestazioni delle colonne

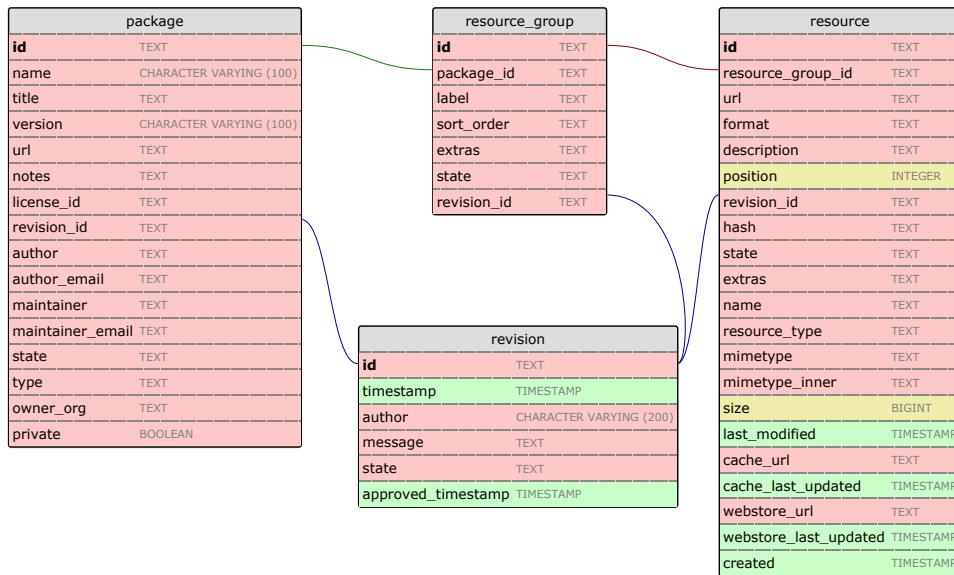


Figura 5.1: Design DataBase

in italiano e si è creato un nuovo campo (link) concatenando l'url base della piattaforma con le informazioni presenti nel database.

Ciò è stato realizzato attraverso una query, il cui risultato è stato esportato in un file in formato CSV (comma-separated values). Si è voluto salvare il file in una posizione che fosse poi recuperabile dal web server, di conseguenza si è scelto di salvarlo nella sua root. Ciò è possibile utilizzando il comando *Copy* dopo aver acceduto al database PostgreSQL:

```
~$ sudo psql -h localhost -U ckanuser ckan_default
\Copy (
    ...QUERY...
) To '/usr/lib/ckan/default/src/ckan/ckan/public/datacatalog.csv'
With CSV HEADER
```

Successivamente, per automatizzare il processo di creazione del catalogo ad ogni nuovo inserimento di un dataset in CKAN, sono stati creati una funzione ed un trigger.

Per fare ciò si accede a postgreSQL e si assegnano i permessi di SuperUser all'utente utilizzato per creare il database:

```
~$ sudo -u postgres psql
=# alter role ckanuser SUPERUSER;
```

Ciò è necessario affinché la funzione che andiamo a creare possa esportare il database all'esterno dell'ambiente. Accediamo quindi al database e procediamo con la creazione di funzione e trigger:

```
~$ sudo psql -h localhost -U ckanuser ckan_default
```

La funzione non fa altro che eseguire l'esportazione della query in CSV nella posizione prestabilita:

```
CREATE FUNCTION create_datacatalog ()
RETURNS trigger
AS $create_datacatalog$
BEGIN EXECUTE '
Copy (SELECT "name" AS "nome", "description" AS "descrizione", "url" AS
"url", "format" AS "formato", "last_modified" AS "ultima modifica",
"author" AS "autore", "author_email" AS "email autore", "maintainer"
AS "manutentore", "maintainer_email" AS "email manutentore" FROM (
SELECT "name", "description", "url", "format", "last_modified",
"resource_group_id", "package_id", resource.id AS "url_id"
FROM "resource_group" JOIN "resource"
ON resource_group.id=resource.resource_group_id) AS "resource"
JOIN (SELECT "id", "author", "author_email", "maintainer",
"maintainer_email", "name" AS "url_name" FROM "package") AS "package"
ON resource.package_id=package.id ORDER BY "last_modified" DESC)
To ''/usr/lib/ckan/default/src/ckan/ckan/public/datacatalog.csv''
With CSV HEADER;
';
RETURN NEW;
END;
$create_datacatalog$
LANGUAGE plpgsql;
```

Mentre invece il trigger si attiva ad ogni modifica della tabella `resource` e invoca la funzione:

```
CREATE TRIGGER datacatalog_update
AFTER insert OR update
ON resource
FOR EACH ROW
EXECUTE PROCEDURE create_datacatalog();
```

Infine per rendere possibile la modifica del file `datacatalog.csv` da parte del web server si ‘e assegnato al file i permessi di scrittura all’utente `postgresql`:

```
~$ sudo touch /usr/lib/ckan/default/src/ckan/ckan/public/datacatalog.csv
~$ sudo chmod 644 /usr/lib/ckan/default/src/ckan/ckan/public/datacatalog.csv
~$ sudo chown postgres:postgres /usr/lib/ckan/default/src/ckan/ckan
/public/datacatalog.csv
```

## 5.5 Dizionario dati

Nelle intenzioni dello sviluppo della piattaforma si è cercato di aggiungere nel catalogo anche le informazioni sul tipo di dato contenuto in ogni singolo dataset. In pratica si voleva inserire un’ulteriore colonna nel catalogo contenente le intestazioni della tabella del dataset. Purtroppo non si è trovata una soluzione per realizzare questa feature in quanto non si è trovato il modo di recuperare le informazioni presenti nei due differenti database che sono relazionati come mostrato in figura 5.2:

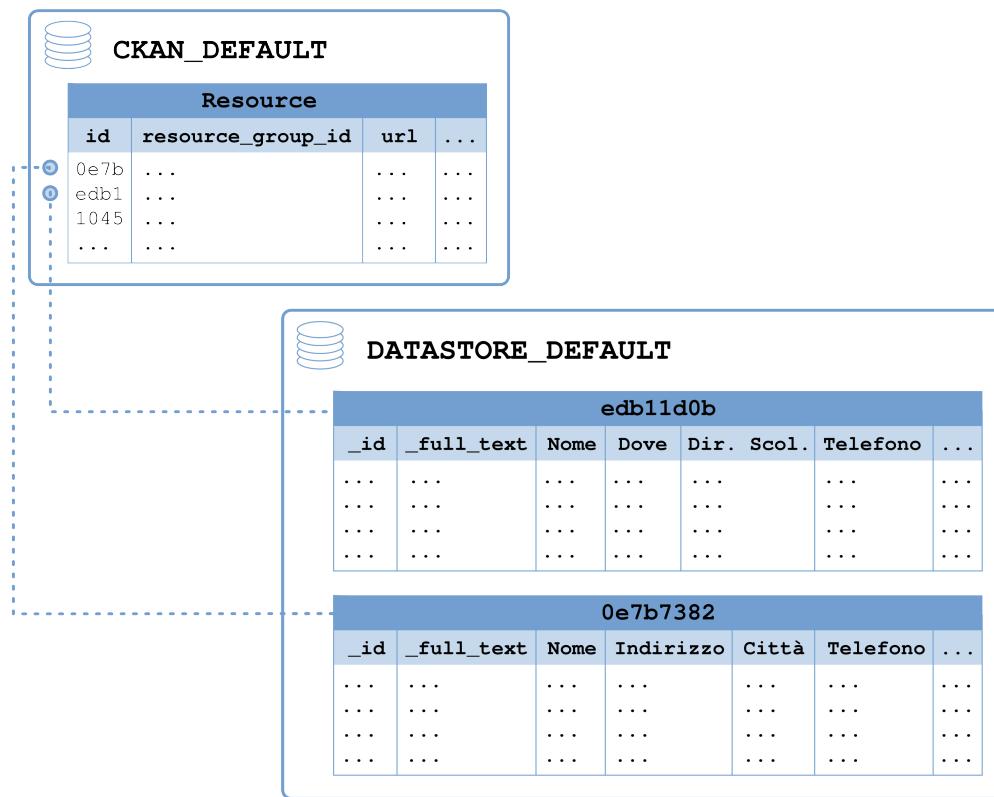


Figura 5.2: relazione tra le tabelle dei DataBase utilizzati da CKAN

L'impossibilità di recuperare l'intestazione delle tabelle del database `datastore_default` una volta conosciuto il valore nel campo `id` dei singoli elementi presenti nella tabella `resource` del database `ckan_default` non ha quindi al momento permesso la creazione del dizionario dati.

# Capitolo 6

## Utilizzo della piattaforma

Dopo la prima installazione la piattaforma si presenta, come mostrato in figura 6.1, con un'interfaccia sobria e pronta a contenere i dataset che varranno caricati.

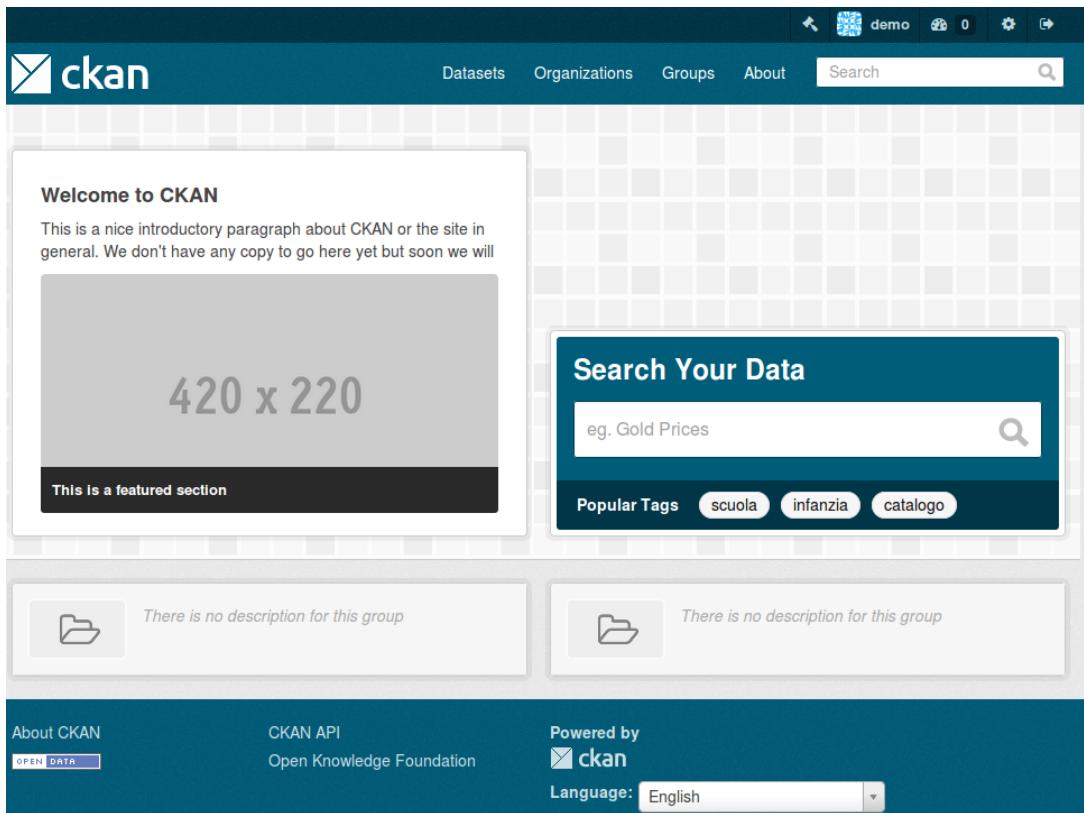


Figura 6.1: Aspetto di CKAN appena installato

Navigando alla voce di menù “Dataset” è possibile vedere l’elenco dei dataset disponibili, ma anche crearne uno nuovo attraverso il pulsante presente in alto a destra (figura 6.2).

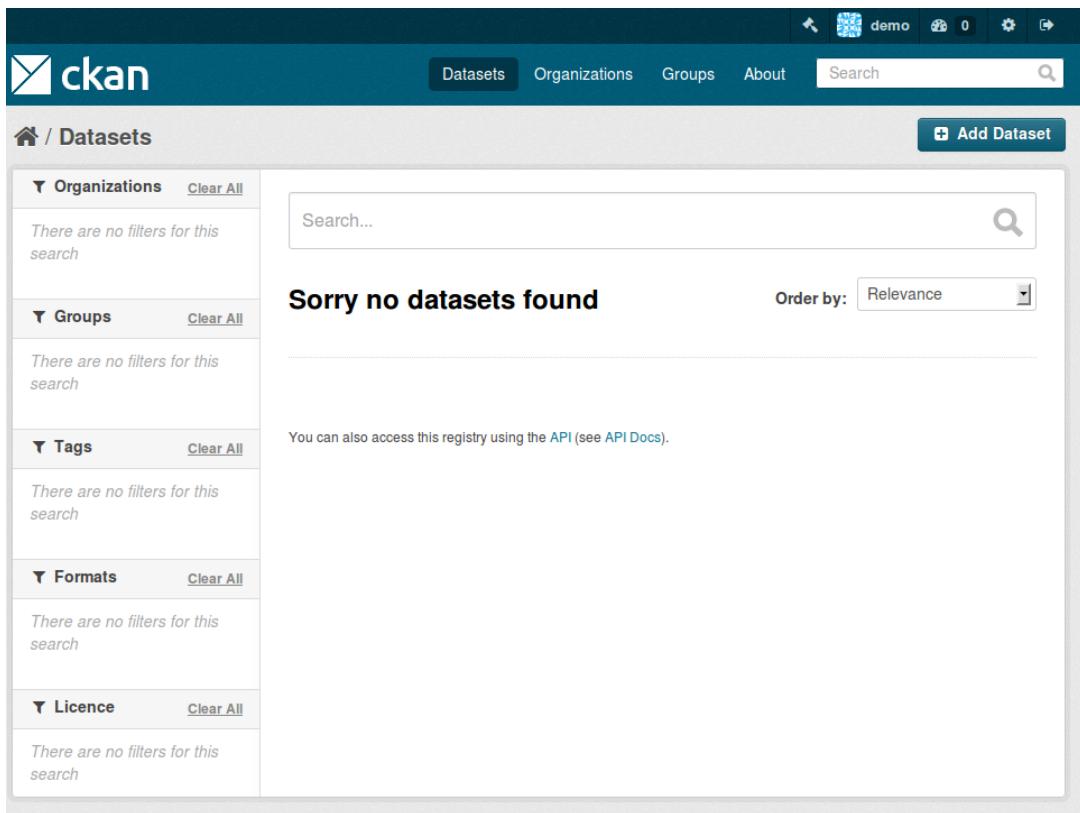


Figura 6.2: Elenco dei Dataset vuoto

L'inserimento di un nuovo Dataset inizia con la richiesta di inserimento delle meta-information necessarie alla sua identificazione (figura 6.3), ovvero il titolo e la descrizione; l'utente può anche decidere di inserire opzionalmente alcuni tag per facilitarne la ricerca e la licenza da un elenco predefinito.

The screenshot shows the CKAN interface for creating a new dataset. The top navigation bar includes links for Datasets, Organizations, Groups, About, and a search bar. The main title is 'Datasets / Create Dataset'. The process is divided into three steps: 1. What are datasets? (completed), 2. Create dataset (current), and 3. Add data. Step 2 is highlighted with a green arrow. The 'Create dataset' section contains fields for Title (e.g., A descriptive title), URL (localhost/dataset/<dataset>), Description (e.g., Some useful notes about the data), Tags (e.g., economy, mental health, government), and License (License Not Specified). A note at the bottom states: 'Important: By submitting content, you agree to release your contributions under the Open Database License.' Buttons for 'Cancel' and 'Next: Add Data' are visible at the bottom right.

Figura 6.3: Definizione dei metadati durante la creazione di un dataset

Ad esempio possiamo inseriti i metadati necessari alla creazione del catalogo (figura 6.4) per poi passare all'aggiunta dei dati veri e propri.

The screenshot shows the CKAN interface for creating a dataset. The top navigation bar includes links for Datasets, Organizations, Groups, About, and a search bar. The main title is 'Datasets / Create Dataset'. The process is divided into three steps: 1. Create dataset (highlighted in green), 2. Add data, and 3. Additional data. The 'Create dataset' step contains fields for Title ('Catalogo'), URL ('localhost/dataset/catalogo'), Description ('catalogo dei dataset presenti'), Tags ('catalogo'), and License ('License Not Specified'). A note at the bottom states: 'Important: By submitting content, you agree to release your contributions under the Open Database License.' Buttons for 'Cancel' and 'Next: Add Data' are visible.

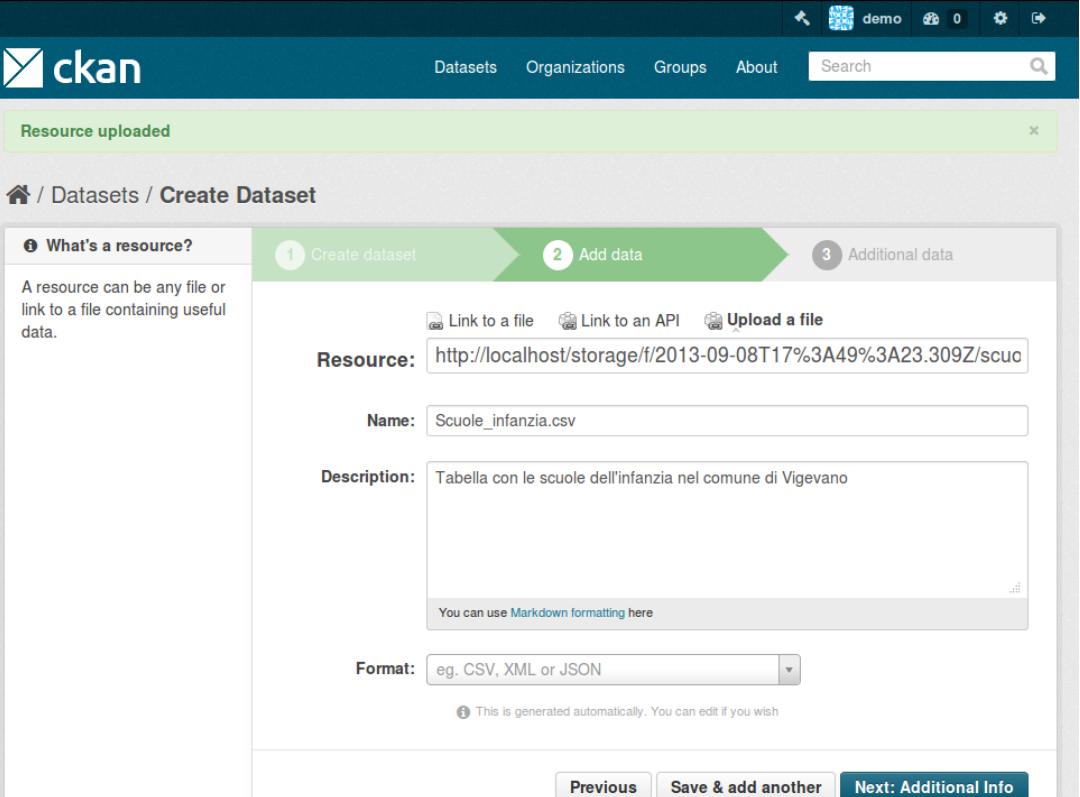
Figura 6.4: Inserimento metadati del catalogo

Nel secondo passaggio è possibile linkare un file o una risorsa esterna e specificare il nome e la descrizione che verranno utilizzati nella piattaforma (figura 6.5). Il formato del file viene identificato automaticamente, in caso ciò non avvenga è possibile inserirlo manualmente.

The screenshot shows the CKAN web interface for creating a dataset. The top navigation bar includes links for Datasets, Organizations, Groups, About, and a search bar. The main title is 'Datasets / Create Dataset'. The process is divided into three steps: 1. What's a resource? (completed), 2. Add data (current), and 3. Additional data (not yet reached). Step 2 is titled 'Add data' and includes three options: 'Link to a file', 'Link to an API', and 'Upload a file'. The 'Link to a file' option is selected, and the URL 'localhost/datacatalog.csv' is entered. Below this, the 'Name' field contains 'datacatalog' and the 'Description' field contains 'Catalogo'. A note indicates that Markdown formatting is supported. The 'Format' dropdown is set to 'eg. CSV, XML or JSON'. A small note below the description area states: 'This is generated automatically. You can edit if you wish'. At the bottom of the form are buttons for 'Previous', 'Save & add another' (disabled), and 'Next: Additional Info'.

Figura 6.5: Inserimento risorsa nel dataset

In alternativa, una volta attivato il File Store è possibile caricare un file dal proprio computer in modo che venga memorizzato nel server (figura 6.6). Un messaggio ci avverterà del successo dell'upload o dell'eventuale fallimento nella parte alta dello schermo.



The screenshot shows the CKAN interface for creating a dataset. At the top, there's a navigation bar with links for Datasets, Organizations, Groups, About, and a search bar. A green banner at the top indicates a successful upload: "Resource uploaded". Below this, the main form is titled "Datasets / Create Dataset". It consists of three steps: 1. What's a resource? (disabled), 2. Create dataset (active), and 3. Additional data (disabled). Step 2 has three options: Link to a file, Link to an API, and Upload a file. The "Upload a file" option is selected, and a file path is shown: "http://localhost/storage/f/2013-09-08T17%3A49%3A23.309Z/scuole\_infanzia.csv". The "Name" field contains "Scuole\_infanzia.csv". The "Description" field contains "Tabella con le scuole dell'infanzia nel comune di Vigevano". A note below the description says "You can use Markdown formatting here". The "Format" dropdown is set to "eg. CSV, XML or JSON". A small note below it says "This is generated automatically. You can edit if you wish". At the bottom of the form are buttons for "Previous", "Save & add another" (disabled), and "Next: Additional Info" (disabled).

Figura 6.6: Upload fine in CKAN

Infine è possibile aggiungere ulteriori informazioni riguardanti l'autore dei dati e il manutentore (colui che si preoccupa di aggiornare il dataset nel tempo). Inoltre è possibile assegnare in dataset ad un gruppo (figura 6.7).

The screenshot shows the CKAN 'Create Dataset' interface. At the top, there is a navigation bar with links for Datasets, Organizations, Groups, About, and a search bar. Below the navigation bar, the page title is 'Datasets / Create Dataset'. The main content area is divided into three steps: '1 Create dataset', '2 Add data', and '3 Additional data'. The '3 Additional data' step is currently active, indicated by a green background and a right-pointing arrow. This step contains fields for 'Author' (set to 'Comune di Vigevano'), 'Author Email' (set to 'protocollo@cert.comune.vigevano.pv.it'), 'Maintainer' (set to 'Joe Bloggs'), and 'Maintainer Email' (set to 'joe@example.com'). There are also three 'Custom Field' input fields, each with 'Key:' and 'Value:' fields. Below these is a dropdown menu labeled 'Add Group'. At the bottom right of the form are 'Previous' and 'Finish' buttons.

Figura 6.7: Inserimento dati addizionali del dataset

A questo punto ci viene mostrato il dataset con tutte le informazioni che abbiamo inserito (figura 6.8). Finchè il DataStorer non entra in funzione la preview non sarà disponibile.

The screenshot shows the CKAN interface for a dataset named 'catalogo.csv'. The top navigation bar includes links for Datasets, Organizations, Groups, About, and a Search bar. Below the header, the URL is shown as [http://localhost/datacatalog.csv](#). The dataset title 'catalogo.csv' is displayed prominently. A note indicates that the resource cannot be previewed at the moment, with a link to more information. There are buttons for Edit, Download, and Data API. On the left, a sidebar titled 'Resources' lists 'Scuole\_infanzia.csv', 'Social' (with links to Google+, Twitter, and Facebook), and 'Additional Information'. The 'Additional Information' section contains a table with the following data:

Field	Value
Last updated	Sep 08, 2013
Created	Sep 08, 2013
Format	CSV
License	License Not Specified

A 'Show more' link is visible at the bottom of the table.

Figura 6.8: Dataset appena caricato, il DataStorer non è ancora entrato in funzione

Una volta che il file sarà processato e i dati contenuti nel dataset saranno salvati nel DataStore e la pagina si presenterà come mostrato in figura 6.9.

The screenshot shows the CKAN Datastore interface. At the top, there is a dark header bar with the CKAN logo, user information (demo, 0 datasets), and navigation links for Datasets, Organizations, Groups, About, and a search bar. Below the header, the URL is shown as <http://localhost/storage/fi/2013-09-08T17%3A49%3A23.309Z/scuole-infanzia.csv>. The main content area displays the dataset title "Scuole\_infanzia.csv". It includes a "Edit" button, a "Download" button (highlighted in green), and a "Data API" button. A subtitle below the title reads "Tabella con le scuole dell'infanzia nel comune di Vigevano". The dataset is presented as a grid table with 4 records. The columns are labeled: \_Id, Nome, Vlia, and Telefono. The data rows are:

_Id	Nome	Vlia	Telefono
1	Scuola de...	Cararola ...	(0381)872...
2	Scuola de...	Beatrice ...	(0381)73326
3	Scuola de...	Sforza n. 5	(0381)83143
4	Scuola de...	Togliatti n...	(0381)325...

Below the table are buttons for "Search data ..." and "Go >". There is also a "Filters" button on the right side of the table.

Figura 6.9: Dataset con le informazioni memorizzato nel DataStore

L'interfaccia grafica è personalizzabile rispetto alle esigenze e nel nostro caso è stata adattata per il Comune di Vigevano (figura 6.10).

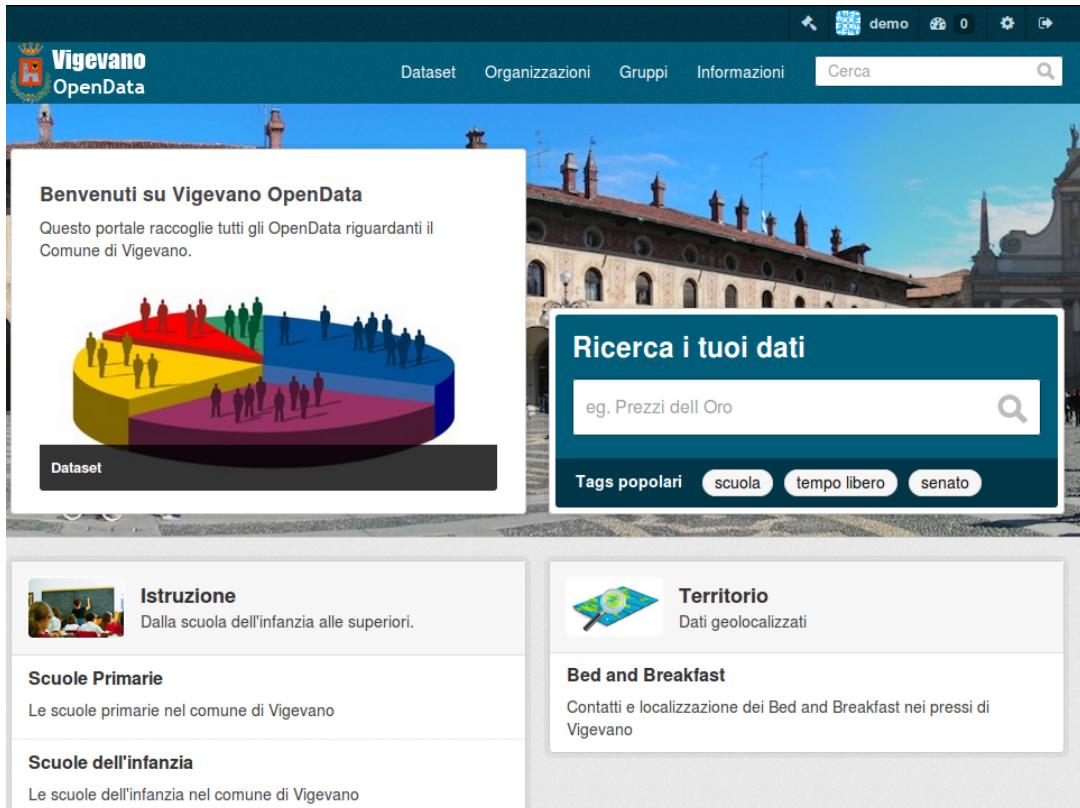


Figura 6.10: Homepage di CKAN realizzato per il Comune di Vigevano

Come mostrato in figura 6.11 è possibile suddividere i dataset per gruppi tematici, una funzione molto apprezzata dagli utenti che ricercano le informazioni di una determinata categoria.

The screenshot shows the Vigevano OpenData website's 'Gruppi' (Groups) page. At the top, there is a navigation bar with links for Dataset, Organizzazioni, Gruppi (which is highlighted in blue), and Informazioni. A search bar is also present. On the left, there is a sidebar with a section titled 'Cosa sono i Gruppi?' containing text about how groups allow users to associate datasets from different sources or topics. Below this, there are three main group categories:

- Elezioni**: Shows a thumbnail of a person holding a ballot and a '1 Dataset' button.
- Istruzione**: Shows a thumbnail of students in a classroom and a '4 Dataset' button.
- Territorio**: Shows a thumbnail of a magnifying glass over a map and a '2 Dataset' button.

At the bottom of the page, there are footer links for 'Informazioni Vigevano OpenData', 'CKAN API', 'Open Knowledge Foundation', and language selection ('Fatto con ckan' and 'Linguaggio: italiano').

Figura 6.11: Elenco dei gruppi di dataset disponibili

Una volta entrati nella pagina di un gruppo viene presentato l'elenco dei dataset disponibili con al relativa descrizione e formato (figura 6.12).

The screenshot shows a web interface for managing datasets. At the top, there's a navigation bar with icons for Home, Groups, and the current group, 'Istruzione'. Below the navigation, there are three tabs: 'Dataset' (selected), 'Flusso di attività', and 'Informazioni'. A search bar with placeholder text 'Cerca...' and a magnifying glass icon is positioned above the dataset list.

**4 dataset trovati**

Ordina per: Rilevanza

**Scuole Superiori**  
Le scuole superiori nel comune di Vigevano  
[CSV](#)

**Scuole Secondarie**  
Le scuole secondarie nel comune di Vigevano  
[CSV](#)

**Scuole Primarie**  
Le scuole primarie nel comune di Vigevano  
[CSV](#)

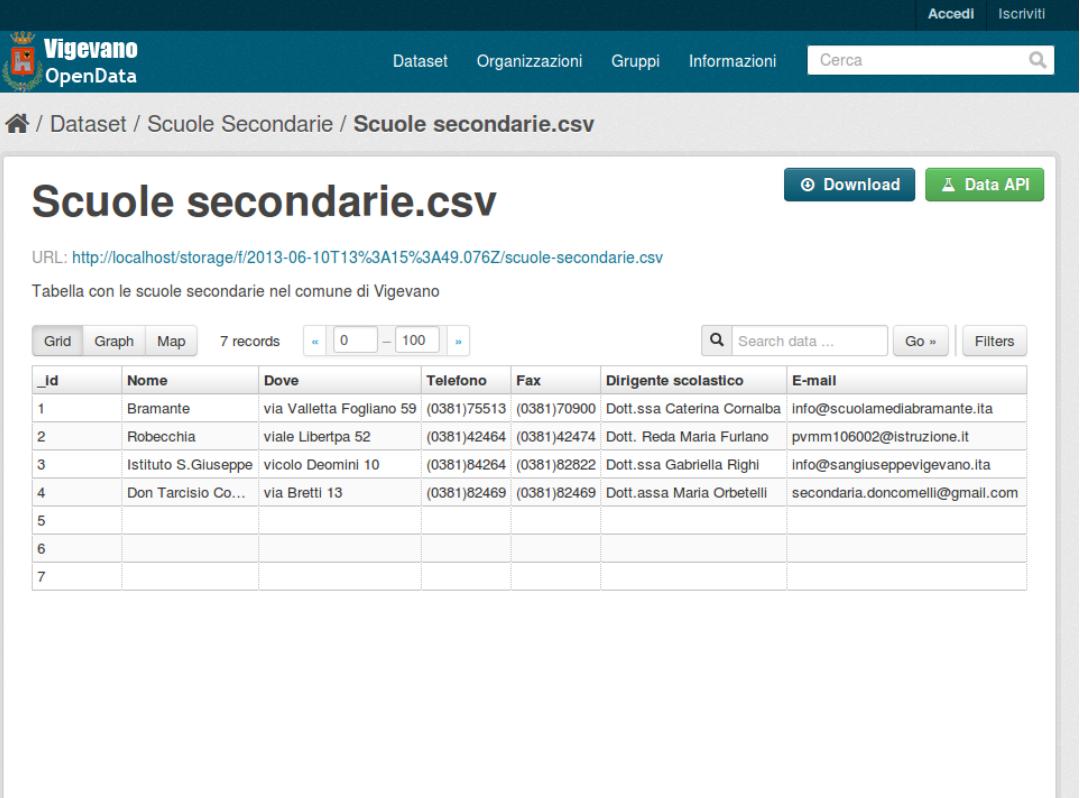
**Scuole dell'Infanzia**  
Le scuole dell'infanzia nel comune di Vigevano  
[CSV](#)

On the left side, there are several filter panels:

- Sostenitori**: 0
- Dataset**: 4
- organization**: Pulisci tutto
- Istruzione (4)**
- Mostra più Gruppi**
- Tag**: Pulisci tutto
- scuola (4)**
- infanzia (1)**
- Mostra più Tag**
- Formati**: Pulisci tutto
- CSV (4)**
- Mostra più Format**
- license\_id**: Pulisci tutto
- Licenza non specificata (4)**
- Mostra più license\_id**

Figura 6.12: Dettaglio dei dataset di un singolo gruppo

I dati contenuti in un dataset possono essere visualizzati in forma tabellare, dove è possibile ricercare una determinata informazione attraverso il campo di ricerca a destra sopra la tabella oppure attraverso i filtri (figura 6.13).



The screenshot shows a web interface for the 'Vigevano OpenData' platform. At the top, there's a navigation bar with links for 'Accedi' (Login), 'Iscriviti' (Register), 'Dataset', 'Organizzazioni', 'Gruppi', 'Informazioni', and a search bar labeled 'Cerca'. Below the header, the URL 'http://localhost/storage/f/2013-06-10T13%3A15%3A49.076Z/scuole-secondarie.csv' is displayed. The main content area is titled 'Scuole secondarie.csv' and includes download buttons for 'Download' and 'Data API'. A sub-header states 'Tabella con le scuole secondarie nel comune di Vigevano'. Below this, there's a table with 7 records, showing data for various secondary schools. The columns are labeled: '\_Id', 'Nome', 'Dove', 'Telefono', 'Fax', 'Dirigente scolastico', and 'E-mail'. The data rows are as follows:

_Id	Nome	Dove	Telefono	Fax	Dirigente scolastico	E-mail
1	Bramante	via Valletta Fogliano 59	(0381)75513	(0381)70900	Dott.ssa Caterina Cornalba	info@scuolamediabramante.it
2	Robecchia	viale Libertà 52	(0381)42464	(0381)42474	Dott. Reda Maria Furlano	pvm106002@istruzione.it
3	Istituto S.Giuseppe	vicolo Deomini 10	(0381)84264	(0381)82822	Dott.ssa Gabriella Righi	info@sangiusseppevigevano.it
4	Don Tarcisio Co...	via Brettì 13	(0381)82469	(0381)82469	Dott.asssa Maria Orbetelli	secondaria.doncomelli@gmail.com
5						
6						
7						

Figura 6.13: Visuale tabellare di un dataset

In caso esista una dipendenza tra due colonne è possibile visualizzare le informazioni attraverso dei grafici di vario tipo, ad esempio come grafico a punti o come grafico a barre (figura 6.14).

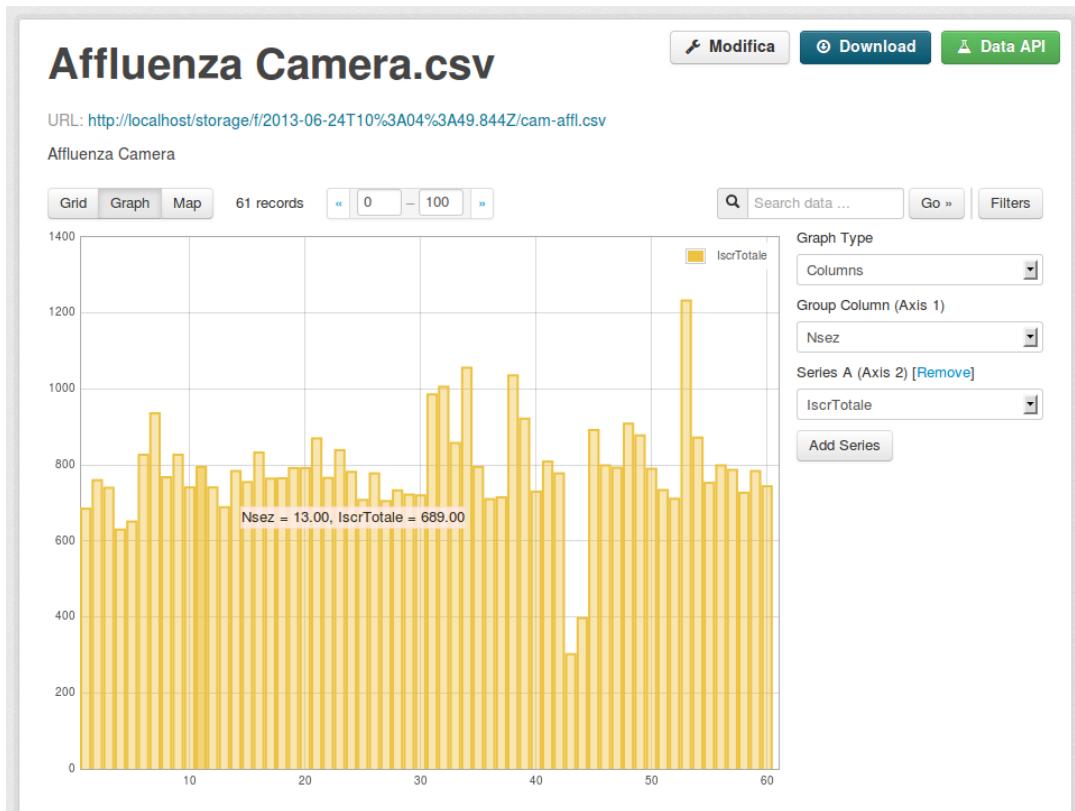


Figura 6.14: Visuale a grafico di un dataset

Infine in caso di dati geolocalizzati è possibile “mapparli” su una cartina (figura 6.15) di MapQuest. Ciò risulta veramente intuitivo per gli utenti finali.

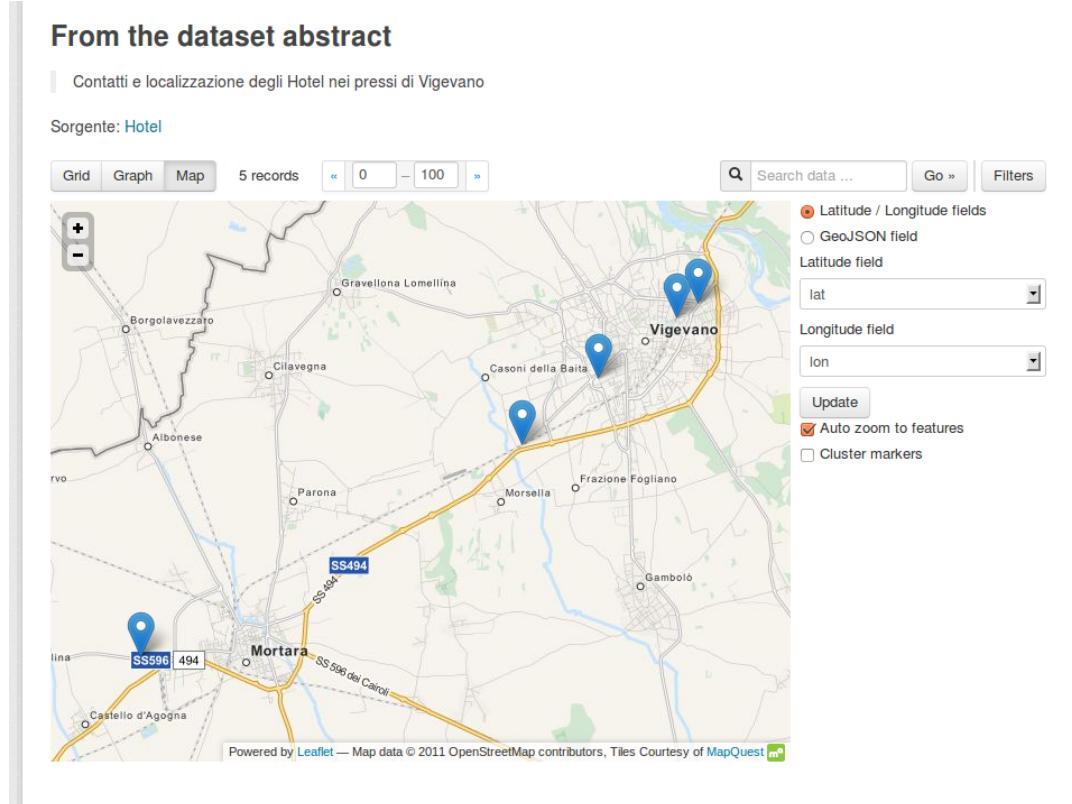


Figura 6.15: Visuale geolocalizzata di un dataset

Inoltre clickando su uno dei marcatori della cartina si visualizzano tutte le informazioni disponibili nel dataset relative al punto selezionato (figura 6.16).

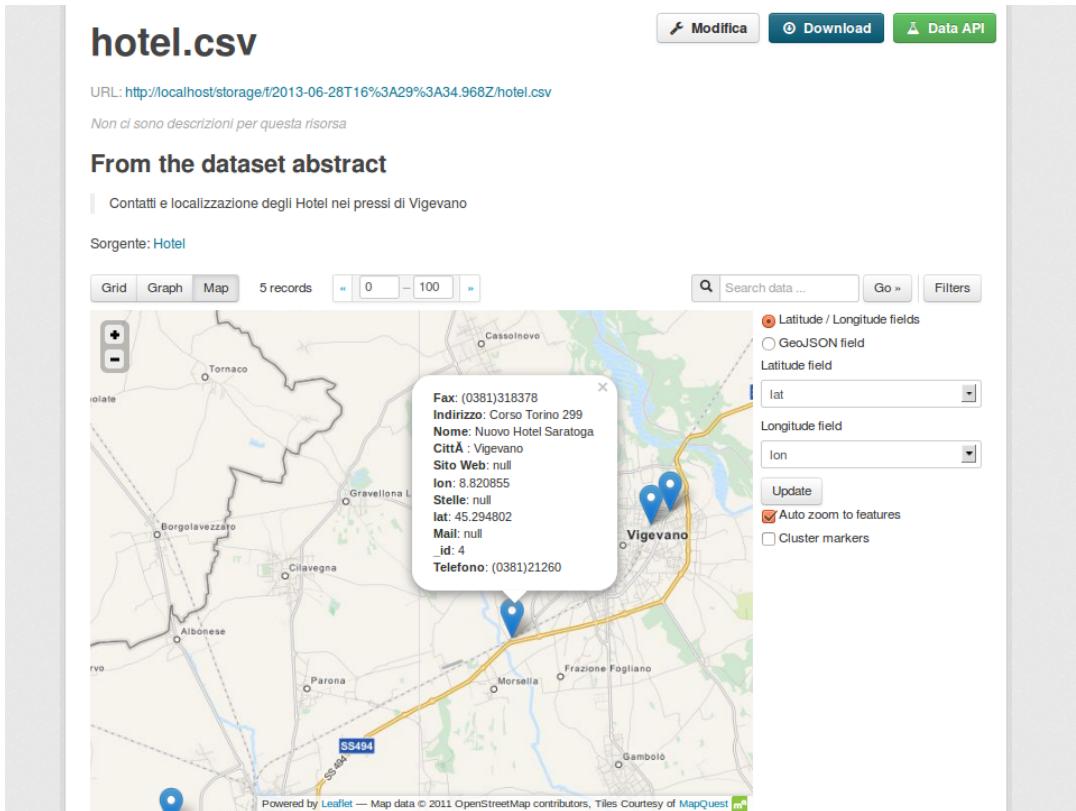
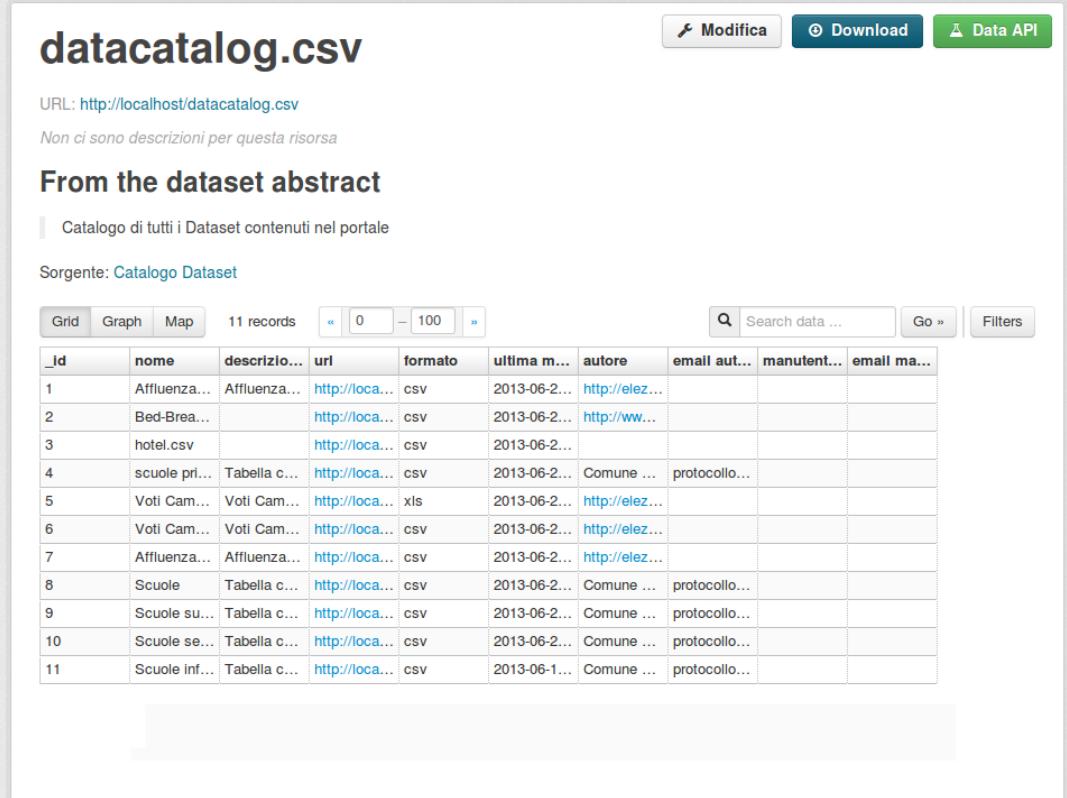


Figura 6.16: Dettaglio del punto selezionato sulla cartina

Man mano che nuovi dataset vengono caricati nella piattaforma il catalogo si aggiorna e si riempie delle informazioni presenti (figura 6.17).



The screenshot shows a web interface for managing datasets. At the top, there's a title "datacatalog.csv" and three buttons: "Modifica" (Edit), "Download", and "Data API". Below the title, it says "URL: <http://localhost/datacatalog.csv>". A note "Non ci sono descrizioni per questa risorsa" (There are no descriptions for this resource) is displayed. The main section is titled "From the dataset abstract" and contains a sub-section "Catalogo di tutti i Dataset contenuti nel portale" (Catalog of all datasets contained in the portal). It says "Sorgente: Catalogo Dataset". Below this, there's a search bar with placeholder "Search data ..." and a "Go »" button, along with a "Filters" button. The central part is a table with 11 records:

_Id	nome	descrizio...	url	formato	ultima m...	autore	email aut...	manutent...	email ma...
1	Affluenza...	Affluenza...	<a href="http://loca...">http://loca...</a>	csv	2013-06-2...	<a href="http://elez...">http://elez...</a>			
2	Bed-Brea...		<a href="http://loca...">http://loca...</a>	csv	2013-06-2...	<a href="http://ww...">http://ww...</a>			
3	hotel.csv		<a href="http://loca...">http://loca...</a>	csv	2013-06-2...				
4	scuole pri...	Tabella c...	<a href="http://loca...">http://loca...</a>	csv	2013-06-2...	Comune ...	protocollo...		
5	Voti Cam...	Voti Cam...	<a href="http://loca...">http://loca...</a>	xls	2013-06-2...	<a href="http://elez...">http://elez...</a>			
6	Voti Cam...	Voti Cam...	<a href="http://loca...">http://loca...</a>	csv	2013-06-2...	<a href="http://elez...">http://elez...</a>			
7	Affluenza...	Affluenza...	<a href="http://loca...">http://loca...</a>	csv	2013-06-2...	<a href="http://elez...">http://elez...</a>			
8	Scuole	Tabella c...	<a href="http://loca...">http://loca...</a>	csv	2013-06-2...	Comune ...	protocollo...		
9	Scuole su...	Tabella c...	<a href="http://loca...">http://loca...</a>	csv	2013-06-2...	Comune ...	protocollo...		
10	Scuole se...	Tabella c...	<a href="http://loca...">http://loca...</a>	csv	2013-06-2...	Comune ...	protocollo...		
11	Scuole inf...	Tabella c...	<a href="http://loca...">http://loca...</a>	csv	2013-06-1...	Comune ...	protocollo...		

Figura 6.17: Catalogo dei dataset inseriti

# Capitolo 7

## Conclusioni e sviluppi futuri

### 7.1 Conclusioni

Lo stage ha raggiunto i suoi obiettivi solo in parte, è infatti stata realizzata la piattaforma secondo i requisiti richiesti e caricati alcuni esempi di dataset, ma non si è riusciti a realizzare l'integrazione a monte con la piattaforma GIT e quella a valle con il portale di Regione Lombardia.

Il processo di **installazione** per il pacchetto base di CKAN risulta essere molto facile nel caso si scelga di utilizzare il sistema operativo consigliato nella documentazione ufficiale (Ubuntu 12.04 a 64 bit), ma si iniziano a incontrare le prime difficoltà già nel momento in cui si vuole aggiungere i componenti DataStore, FileStore File Uploads e DataStorer, i quali però risultano essenziali al fine di realizzare una piattaforma realmente utilizzabile.

L'installazione di CKAN su altri sistemi operativi (ad esempio CentOS) è invece molto complicata in quanto va effettuata utilizzando i file sorgenti presenti su [github](#) per i quali è richiesto un livello di conoscenza dell'architettura della piattaforma maggiore, oltre che delle buone conoscenze dell'ambiente UNIX, del database PostgreSQL, di Python e di virtualenv.

Inoltre la documentazione su CKAN presente online è molto scarsa e la fonte ufficiale [docs.ckan.org](#) è spesso non concorde con la versione distribuita o presenta errori nei comandi shell da eseguire.

Il processo di **manutenzione** di CKAN è risultato abbastanza complicato, all'inizio dello stage la versione disponibile era la 1.8.2, ma a breve è stata resa disponibile la 2.0.1, sulla quale è stato realizzato il progetto. Oggi la versione distribuita è la 2.2. Il passaggio dalla versione 1.8 alla 2.0 è risultato impossibile e si è preferito riconfigurare la macchina virtuale da zero, anche

perché la versione del sistema operativo necessario per configurare CKAN era cambiata (senza che ciò venisse detto nella documentazione). Per le versioni minori il processo di aggiornamento è più semplice, ma è spiegato solo nel caso si utilizzino i pacchetti precompilati.

CKAN mette a disposizione degli utenti un servizio di supporto gratuito attraverso due mailing list, una dedicata agli sviluppatori (`ckan-dev`) e una di carattere generico (`ckan-discuss`). L'impressione avuta è che in queste mailing list non venga però dato supporto dagli sviluppatori, ma sia solo un canale di comunicazione tra i diversi utenti. Inoltre CKAN mette anche a disposizione un servizio di supporto a pagamento dove però non vengono indicati i costi.

**L'utilizzo** della piattaforma da parte dell'amministratore del sistema è semplice, ma richiede che venga effettuato del lavoro a monte, infatti il processo di caricamento dei dataset risulta essere snello, ma poco efficace quando si carica un file in formato CSV non formattato come richiesto dalla piattaforma (contenente caratteri speciali, elementi separati da puntoevirgola o tabulazione al posto della virgola) poiché l'analisi automatica da parte del DataStorer non viene effettuata correttamente e ciò non viene comunicato. Discorso analogo avviene per i formati XLS.

Le API fornite dalla piattaforma permettono di effettuare comodamente le operazioni più comuni e sono ben descritte.

L'interfaccia è pulita e di facile utilizzo da parte dei cittadini. Il download dei dataset può avvenire solo nel formato in cui sono stati caricati, ciò vuol dire che utenti meno esperti troveranno difficile utilizzare i datset poiché abituati ad utilizzare formati differenti da quelli usati dai tecnici.

Per poter utilizzare la piattaforma CKAN serve quindi un team di persone che si occupino dei 3 aspetti precedentemente elencati. È necessario avere delle buone conoscenze del sistema operativo Ubuntu o delle ottime competenze su altre distribuzioni Linux e saper installare CKAN dai sorgenti. In alternativa è possibile richiedere il supporto di CKAN per il deploy, il cui costo però non è noto. Un'altra alternativa può essere di scegliere di utilizzare il servizio di hosting fornito da CKAN, il cui prezzo varia tra i 550\$ e i 750\$ mensili per la versione che esclude le API sui dataset e tra i 1500\$ e i 2000\$ mensili per la versione che supporta anche le API e include 1 giorno di supporto tecnico.

Ad oggi il processo di evoluzione di CKAN sembra essere molto rapido e vengono rilasciate nuove versioni della piattaforma con una rapidità fino a poco tempo fa impensabile. A breve inoltre dovrebbe essere pubblicato la versione di Socrata Open Source definitiva. In quest'occasione sarà utile fare

un nuovo confronto tra le due piattaforme e valutare quale risponde meglio alle esigenze del Comune.

## **7.2 Sviluppi futuri**

Recentemente CKAN ha reso disponibile una guida per installare CKAN 2.0 su macchina virtuale Windows Azure, questa potrebbe essere una soluzione da sperimentare per abbattere i problemi di configurazione iniziali.

L'integrazione con GIT è ancora da sviluppare, occorre fare un'analisi tecnologica di questa piattaforma e capire quali dati possono essere espostati e in che modalità. Da non sottovalutare la necessità di filtrare ed aggregare i dati al fine di proteggere la privacy e la possibile inferenza legata ai casi in cui le informazioni riguardino uno specifico settore.

L'integrazione con Regione Lombardia sembra essere realizzabile, ma andrà sperimentata accuratamente prima di renderla disponibile per non rischiare di avere dei dataset pubblicati sulle due piattaforme non coerenti l'uno con l'altro.

# Bibliografia

- [1] Wikipedia - Dati aperti - 2013/06/29 - [https://it.wikipedia.org/wiki/Dati\\_aperti](https://it.wikipedia.org/wiki/Dati_aperti)
- [2] Open Government Working Group - 8 Principles of Open Government Data - 2007/12/08 - [https://public.resource.org/8\\_principles.html](https://public.resource.org/8_principles.html)
- [3] Tim Berners-Lee - LinkedData - 2009/06/18 - <http://www.w3.org/DesignIssues/LinkedData.html>
- [4] Peter R. Orszag - Open Government Directive - 2009/12/09 - [http://www.whitehouse.gov/sites/default/files/omb/assets/memoranda\\_2010/m10-06.pdf](http://www.whitehouse.gov/sites/default/files/omb/assets/memoranda_2010/m10-06.pdf)
- [5] The White House - Obama Administration Releases Historic Open Data Rules to Enhance Government Efficiency and Fuel Economic Growth - 2013/05/09 - <http://www.whitehouse.gov/the-press-office/2013/05/09/obama-administration-releases-historic-open-data-rules-enhance-government>
- [6] Dati.gov.it - Italian Open Data License v2.0 - <http://www.dati.gov.it/iod1/2.0/>
- [7] 89th United States Congress - Freedom of Information Act - 1966/07/04 - <http://www.gpo.gov/fdsys/pkg/STATUTE-80/pdf/STATUTE-80-Pg250.pdf>
- [8] Parlamento Italiano - LEGGE 24 novembre 2000, n. 340 ((Disposizioni per la delegificazione di norme e per la semplificazione di procedimenti amministrativi - Legge di semplificazione 1999.)) - 2000/11/24 - <http://www.normattiva.it/uri-res/N2Ls?urn:nir:stato:legge:2000-11-24;340>

- [9] Gazzetta ufficiale dell'Unione europea - DIRETTIVA 2003/98/CE DEL PARLAMENTO EUROPEO E DEL CONSIGLIO del 17 novembre 2003 relativa al riutilizzo dell'informazione del settore pubblico - 2003/11/17 - <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2003:345:0090:0096:IT:PDF>
- [10] Senato Italiano - DECRETO LEGISLATIVO 7 marzo 2005, n. 82 Codice dell'amministrazione digitale. - 2005/03/07 <http://www.normattiva.it/uri-res/N2Ls?urn:nir:stato:legge:2005-03-07;82>
- [11] Senato Italiano - DECRETO LEGISLATIVO 24 gennaio 2006, n. 36 - 2006/01/24 <http://www.normattiva.it/uri-res/N2Ls?urn:nir:stato:decreto.legislativo:2006-01-24;36>
- [12] Garante per la protezione dei dati personali - Le attività economiche e i rapporti di lavoro - 2008/06/12 - <http://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/1637984>
- [13] Senato Italiano - DECRETO LEGISLATIVO 27 ottobre 2009, n. 150 Attuazione della legge 4 marzo 2009, n. 15, in materia di ottimizzazione della produttività del lavoro pubblico e di efficienza e trasparenza delle pubbliche amministrazioni. (09G0164) - 2009/10/27 <http://www.normattiva.it/uri-res/N2Ls?urn:nir:stato:decreto.legge:2009-10-27;150>
- [14] Alessandra Cornero (a cura di), Ernesto Belisario, Stefano Epifani - Vademecum Open Data: come rendere aperti i dati delle pubbliche amministrazioni - 2011/10/20 - <http://www.dati.gov.it/sites/default/files/VademecumOpenData.pdf>
- [15] Senato Italiano - DECRETO-LEGGE 22 giugno 2012, n. 83 Misure urgenti per la crescita del Paese. (12G0109) - 2012/06/22 <http://www.normattiva.it/uri-res/N2Ls?urn:nir:stato:decreto.legge:2012-06-22;83>
- [16] Senato Italiano - DECRETO-LEGGE 18 ottobre 2012, n. 179 Ulteriori misure urgenti per la crescita del Paese. (12G0201) - 2012/10/18 <http://www.normattiva.it/uri-res/N2Ls?urn:nir:stato:decreto-legge:2012-10-18;179>
- [17] Senato Italiano - LEGGE 6 novembre 2012, n. 190 Disposizioni per la prevenzione e la repressione della corruzione e dell'illegalità nella pubblica amministrazione. (12G0213) - 2012/11/06 -<http://www.normattiva.it/uri-res/N2Ls?urn:nir:stato:legge:2012-11-06;190>

- [18] Senato Italiano - DECRETO LEGISLATIVO 14 marzo 2013, n. 33 Rior-dino della disciplina riguardante gli obblighi di pubblicità, trasparenza e diffusione di informazioni da parte delle pubbliche amministrazio-ni. (13G00076) - 2013/03/14 - <http://www.normattiva.it/uri-res/N2Ls?urn:nir:stato:decreto.legislativo:2013-03-14;33>
- [19] Gazzetta ufficiale dell'Unione europea - DIRETTIVA 2013/37/UE DEL PARLAMENTO EUROPEO E DEL CONSIGLIO del 26 giugno 2013 che modifica la direttiva 2003/98/CE relativa al riutilizzo dell'informazione del settore pubblico - 2013/06/26 - <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:175:0001:0008:IT:PDF>
- [20] Agenzia per l'Italia Digitale - Regolamenti e direttive locali open data (a cura delle Regioni-CISIS) - 2013/08/01 - [http://www.digitpa.gov.it/sites/default/files/allegati\\_tec/Regolamenti%20e%20direttive%20locali%20open%20data%20regioni.pdf](http://www.digitpa.gov.it/sites/default/files/allegati_tec/Regolamenti%20e%20direttive%20locali%20open%20data%20regioni.pdf)
- [21] Regione Lombardia - CRITERI GENERALI PER L'OPEN DATA - 2012/01/11 - <http://www.agendadigitale.regione.lombardia.it/shared/ccurl/96/714/dgr%202904%20open%20data.pdf>
- [22] Città di Vigevano - Verbale di deliberazione della Giunta Comunale N. 104 - 2013/05/23 - ADESIONE ALLE "LINEE GUI-DA OPEN DATA PER GLI ENTI LOCALI", DI REGIO-NE LOMBARDIA E SPERIMENTAZIONE CONGIUNTA IN MATERIA - <http://www.agendadigitale.regione.lombardia.it/shared/ccurl/430/952/GCn%20104%20del%2023-05-2013.pdf>
- [23] Comune di Vigevano - Servizio Informatico Comunale - SIC - <http://www.comune.vigevano.pv.it/comune/uffici-e-servizi/affari-general/servizio-informatico-comunale>

# Ringraziamenti

Non è facile citare e ringraziare, in poche righe, tutte le persone che hanno contribuito alla nascita e allo sviluppo di questa tesi di laurea. È stato un lungo e piacevole cammino, che ha contribuito alla mia crescita personale.

Tutto è partito con un treno diretto a Bologna per andare ad un raduno che gli organizzatori stessi definivano essere un “*incontro in carne, ossa e tortellini*”. In questa giornata ho avuto il tempo di parlare approfonditamente con il mio compagno di viaggio e l’occasione di conoscerlo meglio, nascendo così in me un sincero sentimento di fiducia ed amicizia. Nella stessa occasione ho scoperto l’esistenza di Spaghetti Opendata, un gruppo di cittadini interessati al rilascio di dati pubblici in formato aperto, ma anche una comunità viva ed accogliente. A loro va il mio primo ringraziamento, perché da quel 18 Gennaio 2013 ho aperto gli occhi su una nuova realtà e gli Open Data hanno iniziato a ronzare nella mia testa.

Un sentito ringraziamento va anche ai miei genitori, che hanno deciso di supportarmi (o forse sopportarmi) in questi ultimi 4 anni, dopo la decisione di punto in bianco di cambiare il Corso di Laurea, abbandonando quello a cui ero iscritto da 3 anni. Non è stato semplice, è stata una nuova avventura, ma che ha dato finalmente i suoi frutti! Sono riusciti a starmi vicino anche in questa difficile scelta, nonostante inizialmente non condividessero a pieno.

In questi anni ho anche avuto l’occasione di entrare in un’altra grande famiglia, sono entrato timido timido con le orecchie tese ad ascoltare, fino ad arrivare oggi a sentirmi parte attiva di essa. Penso che il Consiglio di Coordinamento Didattico di Scienze e Tecnologie Informatiche sia una rara esempio di come le idee dei singoli possano essere portate con serenità su un tavolo di discussione e di confronto. Per questo motivo voglio ringraziare tutti i componenti del CCD e l’intero Dipartimento di Informatica per avermi trattato come un figlio e avermi dato l’opportunità di crescere.

---

Molti ringraziamenti vanno a coloro che hanno condiviso con me anche solo un piccolo pezzo del mio cammino in quest'Università:

*Abbiamo incontrato la gente più strana  
e imbarcato compagni di viaggio  
qualcuno è rimasto  
qualcuno è andato e non s'è più sentito.*

*Buon viaggio hermano querido<sup>1</sup>  
e buon cammino ovunque tu vada  
forse un giorno potremo incontrarci  
di nuovo lungo la strada.*

*Che le stelle ti guidino sempre  
e la strada ti porti lontano!<sup>2</sup>*

Come dimenticare i compagni di bevute serali e i produttori di buoni dosi di spritz, che sono riusciti ad alleviare il mio stress nei momenti in cui nulla sembrava funzionare. Senza di voi ne sarei uscito più pazzo di quanto già lo sia!

Infine un'enorme grazie a tutti coloro che hanno avuto l'ardire di leggere questa tesi, amici e compagni, grazie di cuore per i suggerimenti e il supporto che mi avete dato nella scrittura di questa tesi, di cui mi sento veramente orgoglioso, grazie a chi si è interessato a una tematica che è diventata a me molto cara, a chi ha anche solo provato a capirci qualcosa. Spero che almeno un granello dell'entusiasmo che mi accompagna nel sentire la parola Open Data abbia potuto travolgervi.

---

<sup>1</sup> “caro fratello” in Spagnolo

<sup>2</sup> strofe (alcune riadattate) di “La Strada” dei Modena City Ramblers