



UNIVERSITÀ DI PISA

## **Corso di Laurea in Informatica Umanistica**

RELAZIONE

### **Verso un'Applicazione Web per la costruzione di una base di conoscenza per documenti digitali, basata su Linked Data**

**Candidato:** *Francesco Sacchini*

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

**Correlatore:** *prof. Alberto Ambrosini*

Anno Accademico 2015-2016

# Indice

<b>1 Introduzione</b>	<b>3</b>
<b>2 Il Web Semantico</b>	<b>6</b>
2.1 I Metadati	6
2.2 I Vocabolari e le Ontologie	7
2.3 I Linked Data	8
<b>3 Stato dell'Arte</b>	<b>11</b>
3.1 Strumenti per la gestione di basi di conoscenza	11
3.2 Strumenti per la catalogazione	13
<b>4 WeME - Il Web Metadata Editor</b>	<b>15</b>
4.1 Approccio	15
4.2 Layout	18
4.2.1 Homepage	18
4.2.2 Pagina di visualizzazione delle risorse	19
4.2.3 Editor delle risorse	20
4.2.4 Pagina delle collezioni	21
4.2.5 Pagina del Database	22
4.2.6 Pagina dell'importazione	23
4.3 Tipologie di utenti	24
<b>5 Test di usabilità</b>	<b>25</b>
5.1 Creazione dell'account	27
5.2 Gestione di una collezione	27
5.3 Gestione di nuove risorse	28
5.3.1 Ricerca e inserimento	28
5.3.1.1 Ricerca e inserimento di persone	29
5.3.1.2 Ricerca e inserimento di luoghi	29
5.3.1.3 Ricerca e inserimento di CHO	30
5.3.2 Visualizzazione	30
5.4 Giudizio generale	30
<b>6 Conclusione e Prossimi Lavori</b>	<b>33</b>
<b>Bibliografia</b>	<b>34</b>
<b>Sitografia</b>	<b>35</b>

# 1 Introduzione

Negli ultimi anni, nel campo delle *Digital Humanities* si è compiuto un grande sforzo per digitalizzare documenti e collezioni in diversi formati, come PDF, XML, TXT e immagini. Tutti questi documenti sono spesso memorizzati, sia in librerie digitali, sia in grandi *repositories*, o archivi digitali, in forma di libri e cataloghi (come ad esempio la *Oxford Digital Library*<sup>1</sup>, la *Library of Congress*<sup>2</sup> e la *Perseus Digital Library*<sup>3</sup>). Le librerie digitali possono essere definite come «organizzazioni che forniscono le risorse per selezionare, organizzare, dare l'accesso intellettuale, interpretare, distribuire, preservare l'integrità e assicurare la persistenza nel tempo delle collezioni digitali, così che queste possano essere accessibili prontamente ed economicamente (Tammaro, 2005)». Si tratta quindi di librerie “immateriali”, in cui vengono conservati e resi disponibili documenti digitali, siano essi nativi digitali o convertiti dagli originali, gestiti e catalogati elettronicamente. La digitalizzazione, tuttavia, non si limita a documenti di testo, ma comprende varie forme e tipologie che hanno portato alla definizione più estesa di *digital object*, o oggetto digitale, ovvero un manufatto (ad esempio un'immagine, una registrazione audio, un documento testuale, ecc.) che è stato codificato in modo digitale e integrato con metadati (v. 2.1) tali da supportarne l'individuazione, l'uso e l'immagazzinamento (California Digital Library, 2001).

Il processo di catalogazione richiede la creazione di una *knowledge base*, o base di conoscenza, un tipo speciale di Database per la gestione della conoscenza, per scopi aziendali, culturali o didattici. Questa deve contenere risorse contestuali associate ai documenti della collezione, come gli autori dei documenti e i luoghi in cui sono stati scritti, e dunque costituisce un ambiente volto a facilitare la raccolta, l'organizzazione e la distribuzione della conoscenza. Le informazioni contenute nella *knowledge base* possono essere usate per arricchire le informazioni a corredo dei documenti (ad esempio, i metadati associati ed essi).

La maggior parte degli strumenti esistenti per la creazione di cataloghi digitali permette di costruire manualmente la propria base di conoscenza, ma il processo manuale è spesso macchinoso e ripetitivo: l'utente è costretto a inserire informazioni ben note su un documento (ad esempio l'autore, la data e il luogo in cui è stato scritto, ecc.), e dato che molti

---

<sup>1</sup> <http://www.odl.ox.ac.uk>

<sup>2</sup> <https://www.loc.gov>

<sup>3</sup> <http://www.perseus.tufts.edu>

documenti possono avere informazioni comuni (ad esempio stesso autore, stesso luogo, ecc.), gli stessi dati devono essere inseriti più volte. In generale, l'inserimento manuale porta a tre grandi svantaggi:

- la probabilità di introdurre errori aumenta;
- l'intero processo è rallentato perchè non è automatico;
- le informazioni inserite sono isolate, non connesse al resto del Web.

In questo lavoro di tesi, è stata implementata un'Applicazione Web intitolata *Web Metadata Editor* (WeME), in modo da fornire agli utenti un'interfaccia *user-friendly* per costruire una base di conoscenza associata a una collezione di dati. WeME mitiga i tre svantaggi descritti sopra, provocati dal processo manuale, estraendo metadati da alcuni nodi Linked Data (v. 2.3) o da altre tradizionali fonti Web. I vantaggi derivanti dall'uso di WeME per la ricerca semi-automatica di metadati sono essenzialmente tre:

- la probabilità di introdurre errori è ridotta;
- il processo di costruzione di una base di conoscenza è facilitato e velocizzato;
- l'applicazione stabilisce nuove relazioni sia tra documenti appartenenti allo stesso catalogo, sia tra documenti che appartengono ad altre fonti Web, grazie all'utilizzo dei Linked Data.

L'applicazione è stata sviluppata per gradi, partendo dall'interfaccia di base fino all'ottimizzazione e alla rifinitura dei dettagli. Per prima cosa, è stato implementato l'editor per l'inserimento dei dati, perfezionato successivamente con le funzionalità di ricerca automatica. Allo stesso tempo, sono state definite le funzioni per la creazione dell'account e per l'accesso a WeME: partendo da un login automatico che sfruttasse i *social media accounts* dell'utente, è stata poi aggiunta una modalità di creazione e accesso manuale. Una volta definita la procedura d'inserimento, si è passati alla creazione di una funzione di ricerca all'interno del Database, per la visualizzazione delle risorse. In seguito è stato possibile lavorare sulle funzioni per la gestione delle collezioni. In ognuno di questi casi, sono state testate diverse possibilità per quando riguarda il *layout*, ovvero il modo in cui presentare i dati, sia per la visualizzazione delle singole risorse, sia per la presentazione delle collezioni, fino a ottenere un'impostazione che privilegiasse la leggibilità, ma senza prescindere dalla cura per il design. Infine, sono state definite le funzioni per l'esportazione e l'importazione dei dati.

Per testare l'usabilità di WeME è stato predisposto un questionario, in modo da comprendere il livello di interesse verso il progetto e il grado di apprezzamento dell'applicazione. Su un campione di 26 intervistati, 5 l'hanno trovata eccellente (giudizio 5/5), 13 l'hanno giudicata molto utile (giudizio 4/5), 6 l'hanno definita un buono strumento (giudizio 3/5) e solo 1 l'ha trovata inutile (giudizio 1/5). Gli intervistati erano per lo più esperti nel settore dell'informatica, ricercatori nel campo delle *Digital Humanities* o utenti con competenze archivistiche.

Lo scopo di questo lavoro di tesi è presentare esaurientemente il progetto, comunicando le motivazioni che ne stanno alla base e le future potenzialità, partendo dall'inserimento dell'applicazione nel relativo contesto, fino alla descrizione più approfondita di tutte le funzionalità. Per questo, le varie sezioni sono organizzate nel modo seguente:

- il *capitolo 2* introduce il concetto di Web Semantico, di Metadati e di Linked Data, in modo che il lettore familiarizzi con il contesto e acquisisca alcune informazioni di base per comprendere il lavoro nell'insieme;
- il *capitolo 3* contiene un'analisi dei *tools* già esistenti nel campo della catalogazione e della gestione di basi di conoscenza, i quali vengono messi a confronto con le caratteristiche di WeME;
- il *capitolo 4* contiene una descrizione esauriente della piattaforma, specificando l'approccio su cui l'applicazione si basa, mostrando le varie sezioni e il layout utilizzato, e analizzando le possibili categorie di utenza;
- il *capitolo 5* contiene i risultati e le statistiche ricavate dal test di usabilità, insieme a tutti i suggerimenti che gli utenti hanno comunicato;
- il *capitolo 6* contiene delle considerazioni finali e la descrizione di possibili sviluppi dell'applicazione.

## 2 Il Web Semantico

Con il termine “Web Semantico”, coniato dal suo ideatore Tim Berners-Lee, si intende la trasformazione del *World Wide Web* dal Web dei documenti al Web dei dati. Mentre la versione precedente della rete, detta anche Web 2.0, si basava essenzialmente su documenti collegati tra di loro attraverso dei link, il Web 3.0, o Web Semantico, cerca di collegare direttamente i dati. In questo modo si possono generare nuovi collegamenti, associando ai documenti pubblicati (pagine HTML, file, immagini, ecc.) delle informazioni che ne specificano il contesto semantico in un formato adatto all’elaborazione automatica. Con l’interpretazione del contenuto dei documenti che il Web Semantico impone, sono possibili ricerche molto più evolute, basate sulla presenza nel documento di parole chiave, e altre operazioni specialistiche, come la costruzione di reti di relazioni e connessioni tra documenti secondo logiche più elaborate del semplice collegamento ipertestuale (Wikipedia, voce *Web Semantico*). Con il Web Semantico, quindi, si sposta l’attenzione dai documenti ai loro contenuti (i dati), i quali devono essere associati a meta-informazioni (i metadati) che ne descrivano il contenuto semantico.

### 2.1 I Metadati

Un metadato (dal greco μετά "oltre, dopo, per mezzo" e dal latino datum "informazione"), letteralmente "(dato) per mezzo di un (altro) dato", è un'informazione strutturata che descrive un altro dato (Wikipedia, voce *Metadato*). Per fare un esempio concreto, i metadati riguardanti un documento di testo potrebbero contenere informazioni sulla lunghezza del documento, sull’autore, sul luogo e il momento in cui è stato scritto, ecc. Nei progetti di digitalizzazione e nelle attività di gestione degli archivi di oggetti digitali, i metadati rivestono un'importanza crescente, tanto da venire considerati parte costituente della definizione di oggetto digitale. In particolare, le funzioni di individuazione, uso e immagazzinamento sono rese possibili dai metadati di tipo descrittivo, sebbene ne esistano altri di tipo gestionale e strutturale che rendono possibili altre funzionalità.

*Resource Description Framework (RDF)*<sup>4</sup> e *SPARQL Protocol and RDF Query Language (SPARQL)*<sup>5</sup> sono i due modelli su cui il Web Semantico si basa:

- il modello RDF è lo strumento base per la codifica, lo scambio e il riutilizzo di metadati strutturati, e consente l'interoperabilità tra applicazioni che si scambiano sul Web informazioni *machine-understandable*. La codifica avviene nella forma di *statements*, o asserzioni, formate da tre parti: soggetto, predicato e oggetto; per questo, le asserzioni sono dette *triple*, e sono rappresentate sotto forma di grafi orientati;
- SPARQL è un linguaggio di interrogazione per dati rappresentati tramite RDF, che permette di estrarre informazioni dalle basi di conoscenza distribuite sul Web; le *queries* si basano su un costrutto, il *triple pattern*, che ricalca la configurazione a triple delle asserzioni RDF fornendo un modello ottimale per la ricerca.

Grazie a questi due modelli è stato possibile definire uno standard di comunicazione e di scambio per garantire l'interoperabilità dei dati, la quale viene resa possibile anche grazie all'utilizzo di Linked Data (v. 2.3).

## 2.2 I Vocabolari e le Ontologie

Altri due elementi portanti del Web Semantico sono i vocabolari e le ontologie, i quali poggiano sul modello RDF: si tratta di schemi di metadati usati per definire concetti e relazioni appartenenti a una certa area di interesse. Un'ontologia, ad esempio, può definire le classi “libri” e “persone”, le quali avranno un certo numero di proprietà (ad esempio titolo, autore, nome, cognome, ecc.); la stessa ontologia, inoltre, può stabilire che tutte le risorse collegate alla classe “libri” tramite la proprietà “autore”, debbano necessariamente appartenere alla classe “persona”, e così via. Alcuni esempi di ontologie utilizzate sono FOAF<sup>6</sup> (*Friend of a Friend*, utilizzata per descrivere persone, attività e relazioni con altre persone), *Dublin Core*<sup>7</sup> (un sistema di metadati per definire attributi di risorse culturali, come titolo, autore, data, soggetto, ecc.), ecc.

---

<sup>4</sup> <https://www.w3.org/TR/2014/REC-rdf-schema-20140225>

<sup>5</sup> <https://www.w3.org/TR/2013/REC-sparql11-protocol-20130321>

<sup>6</sup> <http://xmlns.com/foaf/spec>

<sup>7</sup> <http://dublincore.org/documents/dces>

## 2.3 I Linked Data

Il termine “Linked Data” si riferisce a un insieme di *best practices*, o linee guida, per la pubblicazione di dati strutturati sul Web, che siano atti ad essere collegati fra loro e quindi utilizzabili attraverso interrogazioni semantiche (Heath, 2011). Questo tipo di dati si basa su quattro principi definiti da Tim Berners-Lee:

- utilizzare un *Uniform Resource Identifier (URI)*<sup>8</sup> per identificare ogni oggetto;
- più specificatamente, utilizzare URIs HTTP, in modo che questi oggetti possano essere ricercati tramite il protocollo HTTP da persone e *user agents*;
- quando qualcuno arriva a un URI, indicare informazioni utili sull'oggetto usando formati standard come RDF e SPARQL;
- includere link ad altri URI relativi ai dati esposti per migliorare la ricerca di altre informazioni correlate nel Web.

Per renderli completamente accessibili al mondo del Web, i dati stessi devono essere pubblicati sotto condizioni d'uso aperte (o libere), che ne consentano consultazione, navigazione e aggregazione. L'*Open Data*, quindi, è l'infrastruttura di cui il Linked Data ha bisogno per poter creare la rete di inferenze tra i vari dati sparsi nel Web.

I Linked Data vengono pubblicati sul Web in diversi *datasets*, o archivi di dati, ovvero collezioni di risorse, tra loro collegate, che hanno una caratteristica comune: coprono lo stesso ambito disciplinare (dataset geografici, governativi, bibliografici, ecc.) e/o si originano dalla stessa fonte di dati (ad esempio, DBpedia è un dataset originato dai dati di Wikipedia). L'interoperabilità è uno dei vantaggi più importanti del modello Linked Data: il valore dei dati aumenta sensibilmente quando differenti *datasets*, prodotti e pubblicati in modo indipendente da diversi soggetti, possono essere incrociati o aggregati liberamente dal fruitore. Per consentire l'interoperabilità è necessario stabilire un *link*, o collegamento, diretto tra dati (possibilmente provenienti da diverse sorgenti) che si riferiscono a oggetti identici o comunque relazionati tra loro. Tale collegamento diretto si manifesta come la possibilità di

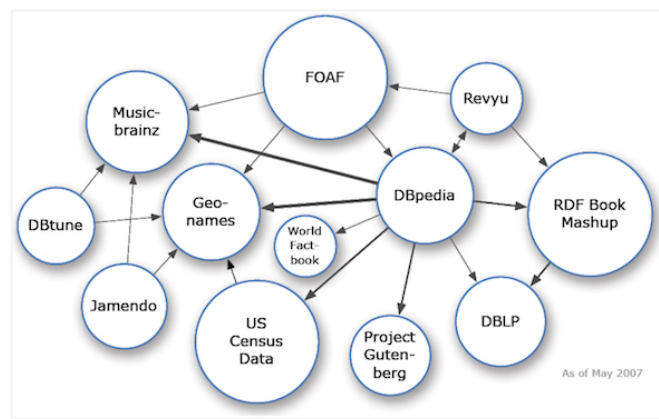
---

<sup>8</sup> La locuzione Uniform Resource Identifier (in acronimo URI) in informatica, si riferisce a una stringa che identifica univocamente una risorsa generica che può essere un indirizzo Web, un documento, un'immagine, un file, un servizio, un indirizzo di posta elettronica, ecc. (Wikipedia, voce *Uniform Resource Identifier*) Nell'ambito dei Linked Data, le URI sono utilizzate per identificare anche oggetti del mondo reale e concetti astratti: persone, luoghi, cose o relazioni tra i vari elementi (ad esempio, URI<sub>1</sub>: Dante, URI<sub>2</sub>: è autore di, URI<sub>3</sub>: Divina Commedia).

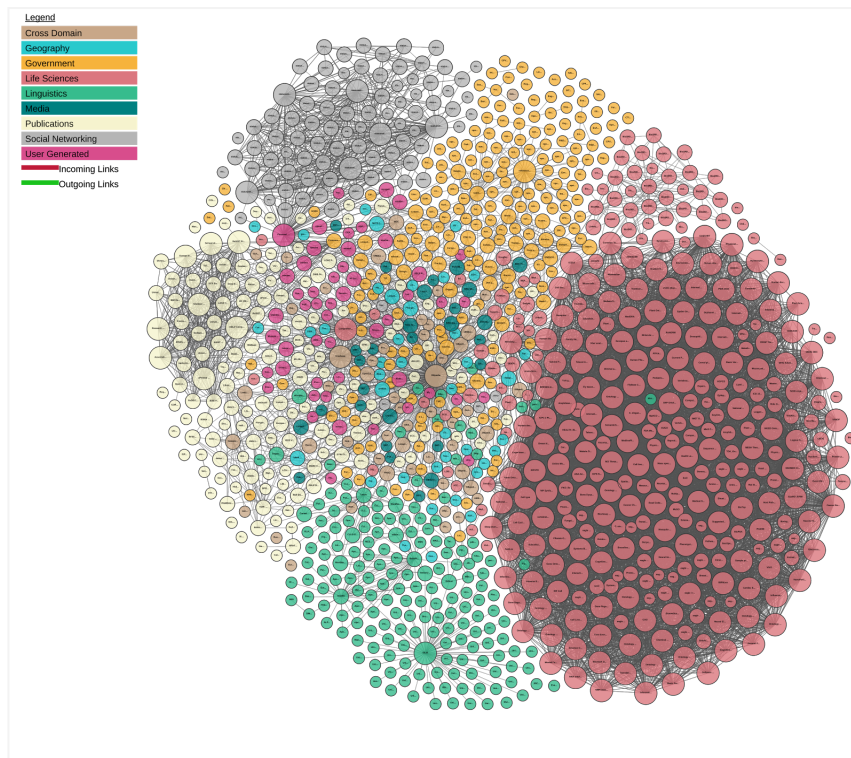


“saltare” da un *dataset* all’altro, ad esempio quando si vuole accedere a informazioni (come i dettagli su una particolare entità) che non si posseggono direttamente.

Tutti i *datasets* connessi tra loro entrano a far parte della cosiddetta *Linked Data Cloud*, a patto che rispettino diversi requisiti: devono basarsi su RDF e HTTP, devono contenere almeno 1000 triple, devono contenere almeno 50 links ad altri *datasets* all’interno della nuvola, ecc. Nel corso di un decennio, la grandezza del diagramma si è ampliata enormemente (v. fig. 1 e 2), così come il numero di ambiti disciplinari, mettendo in mostra il crescente interesse nella pubblicazione di dati sul Web in forma di Linked Data (Abele, McCrae, 2007-2017).



**Figura 1.** La Linked Data Cloud nel 2007



**Figura 2.** La Linked Data Cloud nel 2017

Una manifestazione concreta delle potenzialità del Web Semantico e dei Linked Data è *Google Knowledge Graph*<sup>9</sup> (v. fig. 3), o grafo della conoscenza, una funzione di ricerca semantica grazie alla quale il motore di ricerca associa un elemento alle parole cercate, mettendolo in relazione con altri per ottenere un risultato più completo. La ricerca semantica è un tipo di ricerca che ha lo scopo di migliorare l'accuratezza dei motori di ricerca Web, tentando di comprendere l'intento del ricercatore e il significato contestuale dei termini adoperati al fine di generare risultati più rilevanti. La ricerca semantica si basa su più fattori: il contesto di ricerca, l'ubicazione del ricercante, l'intento, la variazione delle parole, sinonimi, interrogazioni generalizzate e specializzate, riscontro di concetti, ecc. (Wikipedia, voce *Ricerca Semantica*)

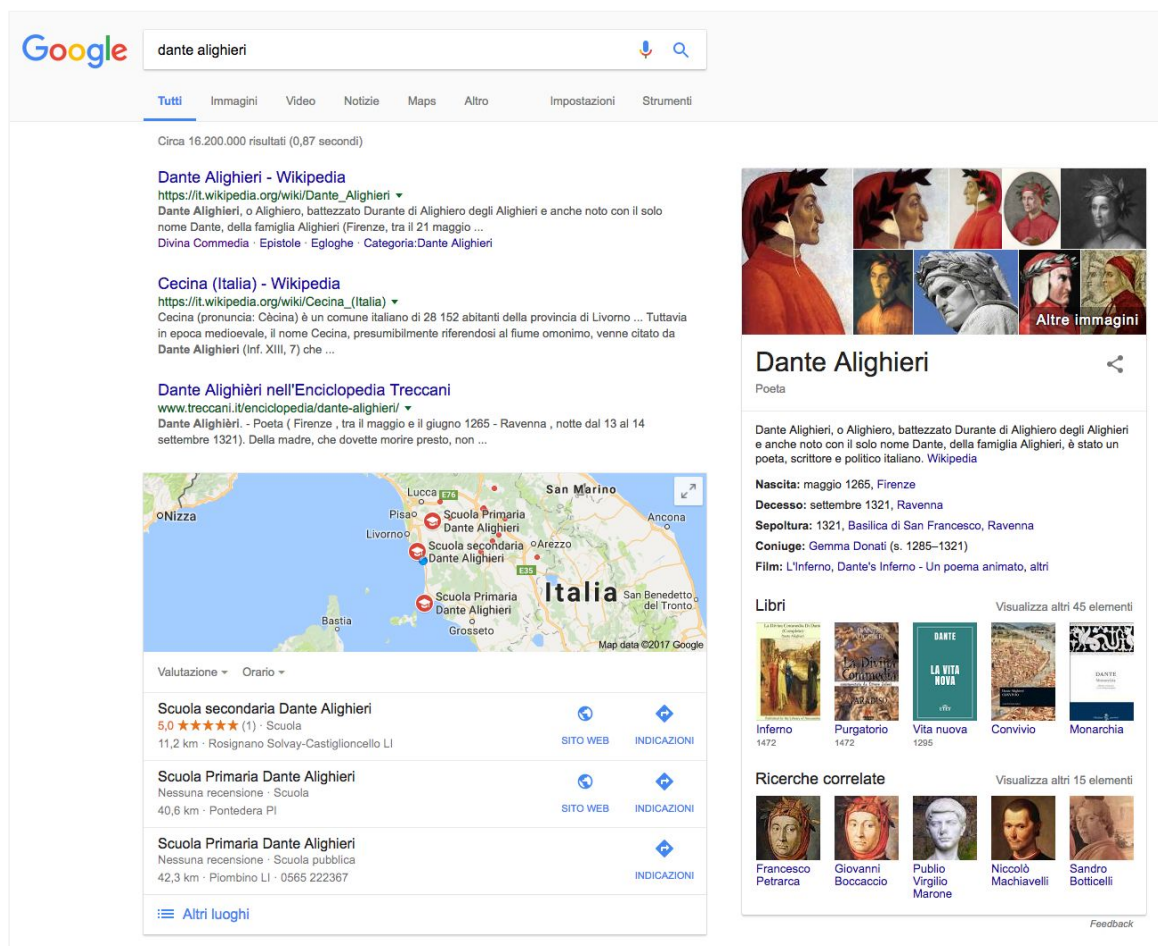


Figura 3. Google Knowledge Graph

<sup>9</sup> <https://www.google.it/intl/it/insidesearch/features/search/knowledge.html>

## 3 Stato dell'Arte

Esistono vari strumenti e progetti che sfruttano Linked Data per costruire basi di conoscenza per diversi scopi, che vanno dalla realizzazione di mostre digitali, o *digital exhibitions*, alla catalogazione di documenti e alla gestione di beni digitali. In questo capitolo, sono stati analizzati differenti *tools*, confrontandoli con WeME sulla base delle seguenti caratteristiche:

- *Open-source*: indica se lo strumento è utilizzabile gratuitamente oppure no;
- *Automatismo*: indica se lo strumento implementa una qualche forma di procedura automatica per velocizzare e rendere più efficiente il lavoro dell'utente;
- *Versione online*: indica se esiste una versione dello strumento utilizzabile online;
- *Versione locale*: indica se esiste una versione dello strumento da scaricare e installare sul proprio PC.

### 3.1 Strumenti per la gestione di basi di conoscenza

Dacura è un sistema di amministrazione e conservazione di basi di conoscenza che permette di creare *datasets* strutturati di alta qualità per un'eventuale pubblicazione sul web (Feeney, 2014). Offrendo un set di diversi strumenti, aiuta l'utente a gestire i *datasets*, dalla raccolta iniziale fino alla pubblicazione sotto forma di Linked Data. La piattaforma non è pensata per librerie e archivi digitali, ma copre diversi aspetti importanti nel contesto dell'Informatica Umanistica, come la provenienza dei dati, la loro qualità, ecc.

Un'altra importante iniziativa è il progetto CULTURA, che sviluppa un ambiente di personalizzazione basato sui metadati, per la navigazione delle collezioni (Hampson, 2012). L'utilizzo di vari algoritmi rende possibile estrarre automaticamente entità e relazioni dai documenti (ad esempio persone, luoghi, date, eventi), per poi elaborarli, contestualizzarli e visualizzarli. In più, supporta differenti categorie di utenti, da ricercatori professionisti fino a semplici utenti interessati.

Un'iniziativa più recente è il progetto Freme<sup>10</sup>, sviluppato dal gruppo che sta dietro a DBpedia (v. 4.1): si tratta di un editor interattivo per identificare e annotare entità nei testi, come ad esempio persone, luoghi ed eventi. Lo scopo del progetto è fornire un insieme di interfacce per l'arricchimento di contenuti testuali digitali, sia in termini di multilinguismo,

---

<sup>10</sup> <http://www.freme-project.eu>

permettendo di inserire informazioni aggiuntive nella lingua originale o in lingue differenti, sia in termini di arricchimento semantico del contenuto, ovvero dell'aggiunta di informazioni in grado di trasformare un contenuto non strutturato nella sua rappresentazione strutturata.

Cold Start KBP Track<sup>11</sup> è un sistema in grado di costruire una base di conoscenza a partire da una collezione di documenti, basandosi su uno schema predefinito di entità e relazioni (attualmente PER, ORG, GPE, LOC e FAC collegate da relazioni binarie). Il sistema individua le varie entità nei documenti e le collega ai nodi corretti nella *knowledge base*.

La Watson Web interface<sup>12</sup> fornisce un motore di ricerca per documenti semantici: inserendo delle *keywords*, o parole chiave, il sistema restituisce una lista di URIs associati a documenti in cui le *keywords* appaiono. Le opzioni di ricerca permettono all'utente di restringere i risultati a particolari tipi di entità (classi, proprietà o individui) o a particolari elementi all'interno dell'entità (*local name, label, comment, ecc.*). Una volta ottenuto il risultato, l'utente può cliccare sul link per visitare la pagina di descrizione dell'entità o del documento.

Rispetto agli strumenti, alle piattaforme e ai progetti esistenti, WeME fornisce una semplice Applicazione Web che non richiede nessuna specifica abilità, e che può essere usata da tutti i tipi di utenti, come ricercatori, archivisti/bibliotecari, ma anche studenti. WeME è fruibile direttamente sul Web, ma il codice sorgente può essere scaricato come *open-source* dalla piattaforma GitHub<sup>13</sup>: in questo modo, l'utente può installare e avviare l'applicazione mediante l'utilizzo di un *Web Server* (per esempio XAMPP).

Strumento	Open-source	Automatismo	Versione Online	Versione Locale
Dacura	No	No	Sì	n.a.
CULTURA	No	No	Sì	No
FREME	Sì	No	No	Sì
Cold Start KBP	No	Sì	n.a.	Sì
Watson WI	Sì	Sì	Sì	Sì
WeME	Sì	Sì	Sì	Sì

**Tabella 1.** Confronto tra i vari strumenti per la gestione di basi di conoscenza

<sup>11</sup> <https://tac.nist.gov/2016/KBP/ColdStart/index.html>

<sup>12</sup> <http://watson.kmi.open.ac.uk/WatsonWUI>

<sup>13</sup> [https://github.com/alod83/metadata\\_editor](https://github.com/alod83/metadata_editor)

## 3.2 Strumenti per la catalogazione

CONTENTdm<sup>14</sup> è uno strumento proprietario creato da OCLC, ed è uno degli strumenti più utilizzati da librerie e musei per la creazione di mostre digitali e per la gestione di beni digitali. Si tratta di uno strumento potente, con un'interfaccia intuitiva, che permette di caricare, descrivere, gestire e accedere a collezioni digitali. Tuttavia, il costo del servizio è proibitivo per molte organizzazioni no-profit, dato che la licenza *entry-level* parte da 4.300 dollari all'anno.

Omeka<sup>15</sup>, è uno strumento *open-source* che fornisce un sistema di pubblicazione Web per archivi digitali. Sia l'interfaccia Web che il sistema di catalogazione convivono in un'unica applicazione, che permette la memorizzazione di *items*, o risorse, e la creazione di collezioni; i metadati associati alle risorse rispettano lo standard *Dublin Core*, costituito da un nucleo di elementi essenziali (*Title, Author, Subject, Description*, ecc.). Lo strumento permette poi di creare siti Web e mostre digitali basati su *templates* e *layouts* pre-costruiti, senza la necessità che l'utente modifichi il codice. L'applicazione è limitata dal punto di vista della catalogazione e dei metadati, poiché punta maggiormente verso l'usabilità da parte dell'utente. Omeka può essere utilizzato sul Web, oppure scaricato in locale e implementato sullo spazio Web del fruitore.

Collective Access<sup>16</sup> è uno strumento *open-source* per la catalogazione, utile per musei e archivi digitali. Grazie a questa applicazione è possibile creare un robusto *dataset* di risorse, specificando le relazioni tra i diversi *records*. La piattaforma offre vari standard di metadati, tra cui *Dublin Core, VRA*, ecc., con l'eventuale possibilità di modificarli. Dal momento che non esiste una versione accessibile direttamente dal Web, Collective Access deve essere scaricato e gestito nello spazio Web del fruitore: questo processo prevede l'installazione di una piattaforma *front-end* e la modifica di codice PHP per l'adattamento al sito Web.

CollectionSpace<sup>17</sup> è uno strumento *open-source* per la gestione di collezioni, utilizzato da musei, librerie, società storiche e altre organizzazioni. Al di là delle funzionalità di gestione, il software non permette direttamente la creazione di mostre digitali, ma offre all'utente una connessione con altre applicazioni *open-source* che possano provvedere al

---

<sup>14</sup> <http://www.oclc.org/en/contentdm.html>

<sup>15</sup> <http://omeka.org>

<sup>16</sup> <http://collectiveaccess.org>

<sup>17</sup> <http://www.collectionspace.org>

posto suo. CollectionSpace supporta diversi schemi di metadati, tra cui *Dublin Core*, che possono essere modificati a piacimento. L'applicazione non è fruibile sul Web, ma deve essere scaricata e gestita sullo spazio Web del fruitore.

Gli strumenti esistenti forniscono potenti interfacce per aggiungere, modificare o eliminare metadati associati a documenti digitali, ma tutte le informazioni devono essere modificate manualmente dall'utente. WeME, invece, sfrutta diversi *Web Services* (Linked Data e RESTful APIs) per estrarre risorse contestuali in modo automatico.

<b>Strumento</b>	<b>Open-source</b>	<b>Automatismo</b>	<b>Versione Online</b>	<b>Versione Locale</b>
<b>CONTENTdm</b>	No	No	Sì	n.a.
<b>Omeka</b>	Sì	No	Sì	Sì
<b>Collective Access</b>	Sì	No	No	Sì
<b>CollectionSpace</b>	Sì	Sì	No	Sì
<b>WeME</b>	Sì	Sì	Sì	Sì

**Tabella 2.** Confronto tra i vari strumenti per la catalogazione

## 4 WeME - Il Web Metadata Editor

Il *Web Metadata Editor* (WeME) fornisce un editor, con un'interfaccia *user-friendly*, per la costruzione di una base di conoscenza contenente risorse contestuali associate a documenti digitali, come gli autori, i luoghi, le date, ecc. L'applicazione è prevista per archivisti e bibliotecari, ma in generale può essere usata da studiosi, studenti e altri utenti che vogliono costruire una base di conoscenza e conmetterla al Web. L'architettura dell'applicazione si basa essenzialmente su tre moduli: l'editor, il motore di ricerca e la base di conoscenza.

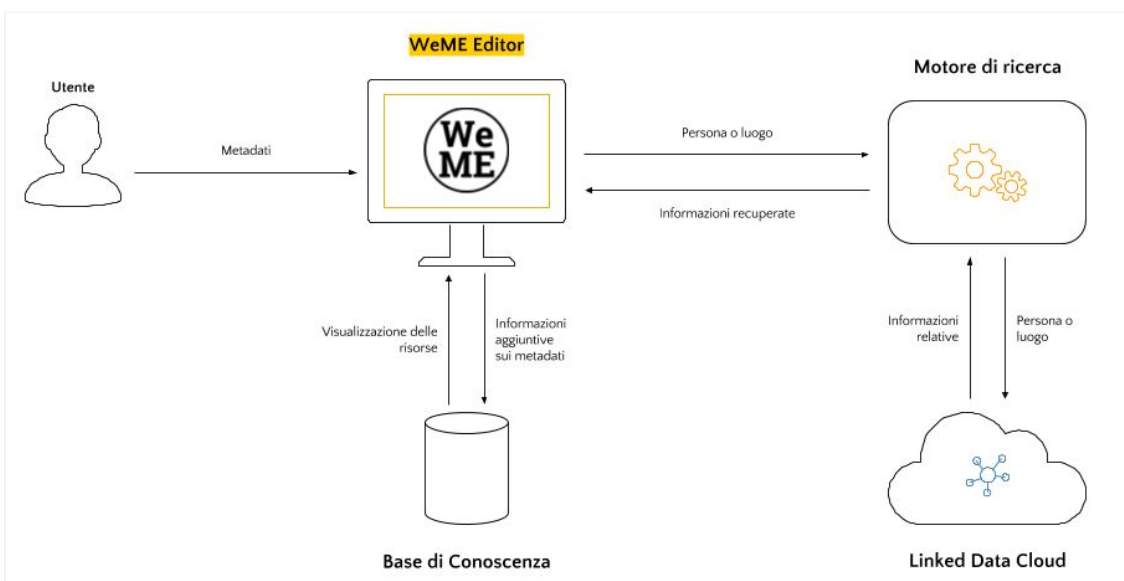


Figura 4. Schema rappresentante l'architettura di WeME

### 4.1 Approccio

Per la rappresentazione delle risorse si è scelto di utilizzare l'*Europeana Data Model* (EDM)<sup>18</sup>, l'ontologia realizzata da Europeana, una biblioteca digitale europea che riunisce contributi già digitalizzati da diverse istituzioni dei 28 paesi membri dell'Unione europea in 30 lingue. In particolare, l'utente potrà gestire risorse appartenenti a tre diverse classi: *person*, *place* e *Cultural Heritage Object (CHO)*. La scelta di EDM è dovuta all'efficacia del modello nel definire relazioni tra le tre classi di risorse: un CHO può essere associato ad una persona, se quest'ultima è l'autore, oppure può avere un legame con un luogo, se il CHO è stato creato in quel luogo specifico.

<sup>18</sup> <http://pro.europeana.eu/page/edm-documentation>

Ogni risorsa può essere costruita attraverso una semplice interfaccia Web, che dà la possibilità di un inserimento manuale o semi-automatico: il processo manuale prevede l'inserimento dei dati da parte dell'utente, che deve compilare i vari campi, alcuni dei quali obbligatori, con le informazioni che possiede; tramite il processo semi-automatico, invece, l'applicazione sfrutta il Web Semantico e quello tradizionale per estrarre informazione da diversi *Web Services* remoti in modo totalmente trasparente per l'utente: nell'interfaccia Web, infatti, è necessario specificare solamente il "nome" della risorsa da cercare, ovvero la coppia nome-cognome per le persone o il nome nella lingua di origine per i luoghi. Il motore di ricerca si basa su *queries* SPARQL per recuperare nodi Linked Data, e su *queries* REST per recuperare dati da fonti più tradizionali. I *Web Services* dai quali WeME estrae i dati sono sostanzialmente tre:

- DBpedia<sup>19</sup>: un progetto aperto e collaborativo per l'estrazione e il riutilizzo di informazioni semistrutturate (Linked Data) da Wikipedia, che costituisce un importante riferimento all'interno della *Linked Data Cloud* per il collegamento tra diverse banche di dati. Il motore di ricerca si connette a DBpedia per recuperare dati riguardanti persone e luoghi (ad esempio date e luoghi di nascita, date e luoghi di morte, biografie, ecc.);
- VIAF<sup>20</sup>: il *Virtual International Authority File* è un progetto internazionale che ha dato vita a un unico Database contenente voci di autorità controllate, in cui l'utente può cercare un autore utilizzando il nome nella propria lingua, accedendo immediatamente ai cataloghi corrispondenti, nazionali o internazionali, per visualizzare la lista delle opere. Il motore di ricerca si connette a VIAF per recuperare le variazioni del nome di una persona in ciascuna lingua, insieme al suo identificativo all'interno del *dataset* di VIAF;
- GeoNames<sup>21</sup>: un progetto che punta alla creazione di un Database geografico mondiale, che possa fornire agli utenti gli strumenti per tradurre semplici nomi geografici (ad esempio il nome di una città) in dati che li rappresentino: latitudine, longitudine, popolazione, ecc. Il motore di ricerca si connette a GeoNames per recuperare le variazioni del nome di un luogo nelle varie lingue, insieme a una breve descrizione e informazioni sulle coordinate geografiche.

---

<sup>19</sup> <http://wiki.dbpedia.org>

<sup>20</sup> <https://viaf.org>

<sup>21</sup> <http://www.geonames.org>



Una volta recuperati i dati, l'utente è libero di modificarli o accettarli per salvarli nella base di conoscenza, oppure di scartarli. Infine, l'utente può visualizzare, modificare e organizzare le proprie risorse in collezioni.

Uno dei problemi principali nel rapportarsi con diverse fonti riguarda la disambiguazione delle risorse: può accadere che, tra due o più fonti, i dati riguardanti uno stesso campo (ad esempio la data di nascita) differiscano ed entrino in conflitto. Attualmente, WeME lascia all'utente il compito della disambiguazione, ma come futuro lavoro sarebbe possibile evitare automaticamente il problema organizzando le fonti secondo una scala gerarchica (ad esempio, associando un punteggio ad ogni fonte), in modo da stabilire le priorità tra dati nel caso in cui questi entrassero in conflitto.

Un ulteriore aspetto di WeME riguarda l'indipendenza della base di conoscenza costruita: pur mantenendo i collegamenti verso le fonti esterne, il Database dell'applicazione è totalmente autonomo e non risente dei cambiamenti esterni. Un altro possibile approccio, evitando di costruire una nuova *knowledge base*, consisterebbe nell'aggiornare o arricchire direttamente le fonti esistenti, come DBpedia e GeoNames; tuttavia, si è preferito seguire la strategia dell'indipendenza per quattro ragioni essenziali:

- gli utenti sono in grado di richiedere la paternità del loro lavoro (ad esempio in ambito accademico o alle agenzie di finanziamento);
- gli utenti hanno il controllo su aggiornamenti che potrebbero compromettere il loro lavoro;
- si è in grado di evitare ritardi e blocchi nell'aggiornamento dei dati, dovuto al processo di validazione;
- le risorse rappresentate sono molto specifiche e pertanto di scarso interesse per un'enciclopedia generica.

## 4.2 Layout

### 4.2.1 Homepage

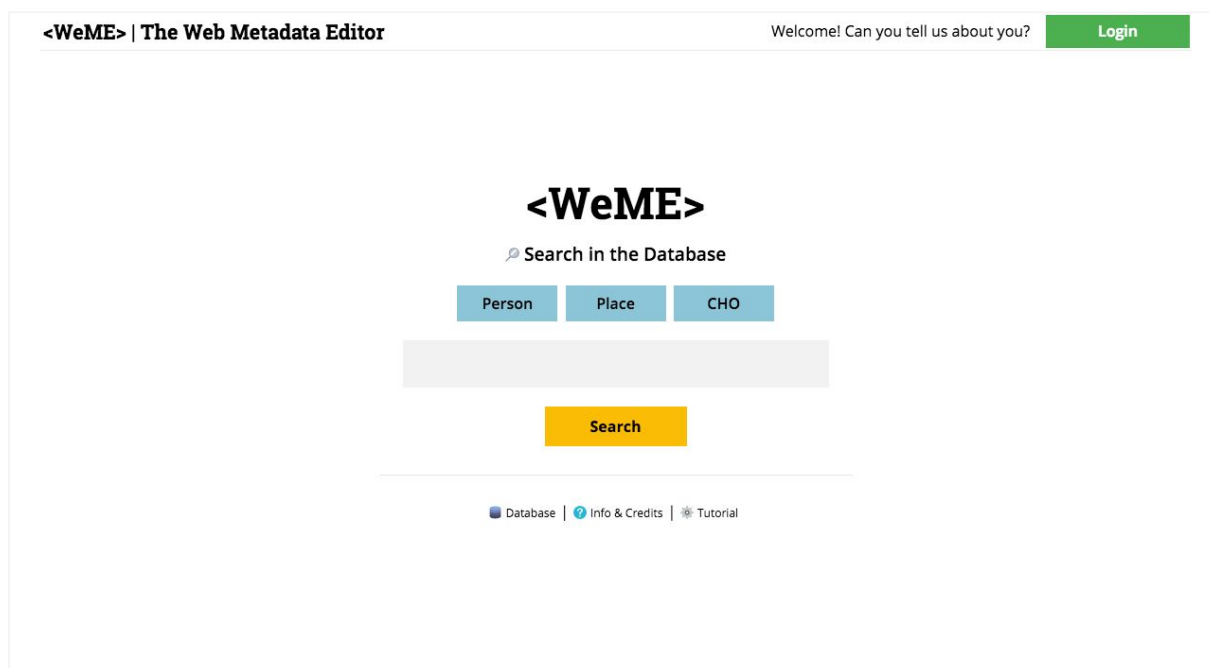


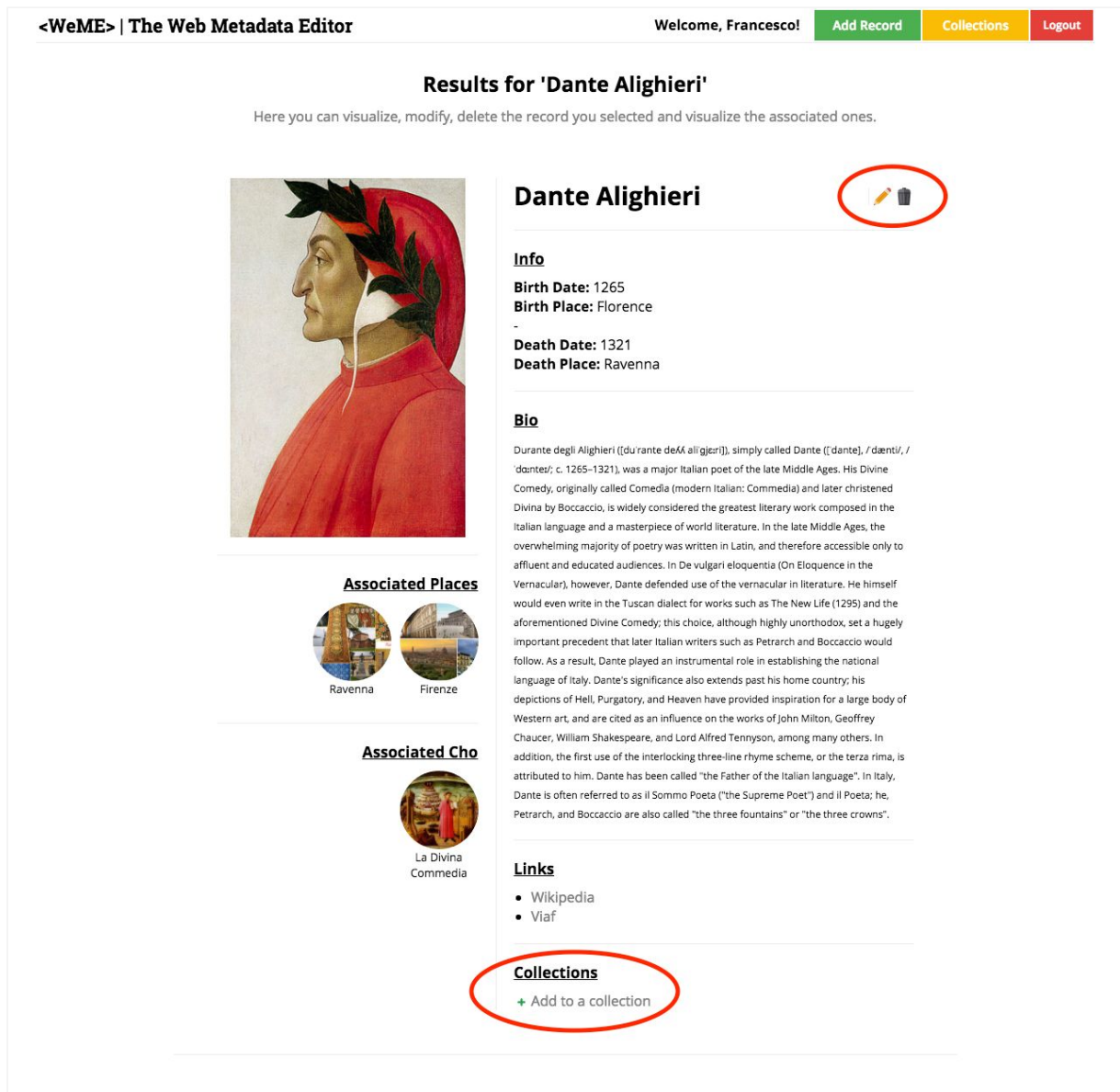
Figura 5. Homepage di WeME

La *homepage* è suddivisa in due sezioni:

- lo *header*: in questa sezione l'utente può effettuare il *login* al suo account personale, sia manualmente, sia usando i suoi *social media accounts* (Google o Facebook). E' inoltre possibile creare un nuovo profilo, nel caso che l'utente non sia ancora registrato. Una volta effettuato il *login*, compariranno nuovi pulsanti: il pulsante "*Add Record*" permette all'utente di aprire un menu per selezionare il tipo di record da aggiungere; il pulsante "*Collections*" porta l'utente alla pagina in cui può gestire le sue collezioni; il pulsante "*Logout*" scollega l'utente dalla piattaforma;
- la *search console*: utilizzando la consolle di ricerca, l'utente può cercare risorse all'interno del Database, specificando la classe della risorsa (*person*, *place* o *CHO*) e la stringa su cui basare la ricerca. Una volta selezionata la classe, digitando la stringa, il sistema suggerirà delle risorse presenti nel Database; un altro modo per visualizzare i suggerimenti è fare doppio click sulla barra di ricerca. Al di sotto, sono presenti altri pulsanti: il pulsante "*Database*" porta alla pagina di visualizzazione dell'intero

Database; il pulsante “*Info & Credits*” porta alla pagina delle informazioni e dei crediti; il pulsante “*Tutorial*” porta alla pagina Wiki del progetto su GitHub.


## 4.2.2 Pagina di visualizzazione delle risorse



<WeME> | The Web Metadata Editor Welcome, Francesco! [Add Record](#) [Collections](#) [Logout](#)

### Results for 'Dante Alighieri'

Here you can visualize, modify, delete the record you selected and visualize the associated ones.



## Dante Alighieri

**Info**  
**Birth Date:** 1265  
**Birth Place:** Florence  
**Death Date:** 1321  
**Death Place:** Ravenna

**Bio**  
Durante degli Alighieri (['duːrante deˈalɪɡjɛri]), simply called Dante (['dante], /ˈdænti/, /ˈdɑːntɛz/; c. 1265–1321), was a major Italian poet of the late Middle Ages. His Divine Comedy, originally called *Comedia* (modern Italian: *Commedia*) and later christened *Divina* by Boccaccio, is widely considered the greatest literary work composed in the Italian language and a masterpiece of world literature. In the late Middle Ages, the overwhelming majority of poetry was written in Latin, and therefore accessible only to affluent and educated audiences. In *De vulgari eloquentia* (On Eloquence in the Vernacular), however, Dante defended use of the vernacular in literature. He himself would even write in the Tuscan dialect for works such as *The New Life* (1295) and the aforementioned *Divine Comedy*; this choice, although highly unorthodox, set a hugely important precedent that later Italian writers such as Petrarch and Boccaccio would follow. As a result, Dante played an instrumental role in establishing the national language of Italy. Dante's significance also extends past his home country; his depictions of Hell, Purgatory, and Heaven have provided inspiration for a large body of Western art, and are cited as an influence on the works of John Milton, Geoffrey Chaucer, William Shakespeare, and Lord Alfred Tennyson, among many others. In addition, the first use of the interlocking three-line rhyme scheme, or the *terza rima*, is attributed to him. Dante has been called "the Father of the Italian language". In Italy, Dante is often referred to as *il Sommo Poeta* ("the Supreme Poet") and *il Poeta*; he, Petrarch, and Boccaccio are also called "the three fountains" or "the three crowns".

**Links**

- Wikipedia
- Viat

**Collections**  
[+ Add to a collection](#)

Figura 6. Pagina di visualizzazione delle risorse di WeME

In questa pagina è possibile visualizzare le informazioni riguardanti le risorse richieste; la visualizzazione dei dati è organizzata in due colonne:

- la colonna di sinistra contiene l'immagine, la mappa (nel caso dei luoghi) e le risorse associate con quella che si sta visualizzando (ad esempio, una persona può essere associata con il luogo di nascita, il luogo di morte e i CHO di cui è autore);

- la colonna di destra contiene tutte le informazioni riguardanti la risorsa: per una persona, ad esempio, si possono visualizzare i dati riguardanti la nascita, la morte, la biografia, i links associati (Wikipedia, VIAF e GeoNames), e le collezioni in cui la risorsa è stata inserita, con la possibilità di aggiungerla ad una nuova. Inoltre, in alto sulla destra sono presenti due pulsanti per modificare o eliminare la risorsa visualizzata.

### 4.2.3 Editor delle risorse

The screenshot shows the 'New Person' form in the WeME interface. At the top, there is a navigation bar with '<WeME> | The Web Metadata Editor' on the left, 'Welcome, Francesco!' in the center, and three buttons: 'Add Record' (green), 'Collections' (yellow), and 'Logout' (red). The main heading is 'New Person' with a sub-instruction: 'Insert the data manually or click the buttons below to check for linked data.' Below this are two buttons: 'Check with DBPedia' and 'Check Viaf ID'. The form fields are arranged in two columns: 'Name\*' and 'Surname\*', 'Birth Date\* (yyyy-mm-dd)' and 'Birth Place', 'Death Date (yyyy-mm-dd)' and 'Death Place'. There is a checkbox labeled 'Still Alive' below the 'Death Date' field. Below the 'Still Alive' checkbox is a text area for 'Bio'. Below the 'Bio' field is a text input for 'Image Link'. Below 'Image Link' are two text inputs: 'Wikipedia Link' and 'Viaf Link'. At the bottom center is a yellow 'Send' button.

Figura 7. Pagina dell'editor delle risorse di classe "persona"

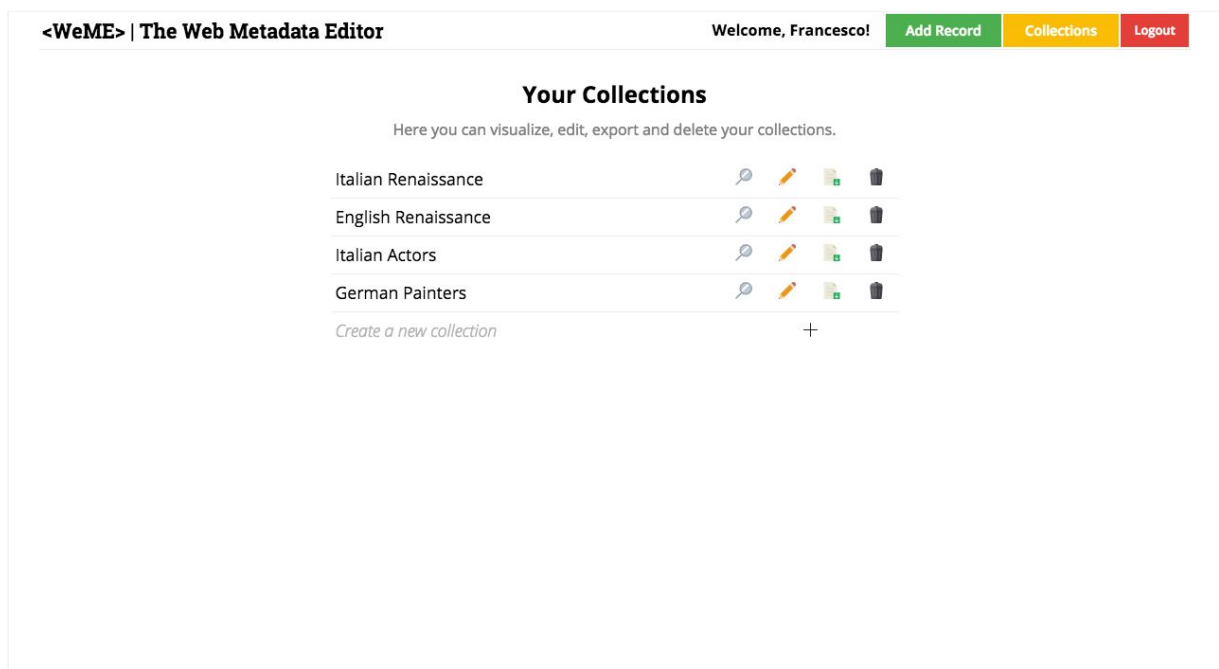
Per quanto riguarda il menu di inserimento dei dati, il *layout* è composto da tre sezioni:

- il *person box*: l'editor dà la possibilità di aggiungere o modificare dati riguardanti una nuova persona, specificando i seguenti campi: nome, cognome, data di nascita, luogo di nascita, data di morte, luogo di morte, biografia, link all'immagine, link a Wikipedia e link a VIAF. E' inoltre presente una *checkbox* per specificare se la persona è ancora in vita o meno. L'utente può inserire i dati manualmente, o premere i pulsanti "*Check with DBpedia*" e "*Check VIAF ID*" per popolare i campi con informazioni provenienti dalle due fonti Web, se presenti. Quando la risorsa è pronta, l'utente può premere il pulsante "*Send*" per memorizzare la persona nel Database;

- il *place box*: l'editor dà la possibilità di aggiungere o modificare dati riguardanti un nuovo luogo, specificando i seguenti campi: nome originale, nome inglese, nazione, regione, popolazione, latitudine, longitudine, descrizione, link all'immagine, link a Wikipedia e link a GeoNames. L'utente può inserire i dati manualmente, o premere i pulsanti “*Check with DBpedia*” e “*Check with GeoNames*” come specificato prima;
- *CHO box*: l'editor dà la possibilità di aggiungere o modificare dati riguardanti un nuovo *Cultural Heritage Object* (ad esempio una lettera, un dipinto, un file audio, ecc.), specificando i seguenti campi: titolo originale, titolo inglese, autore, data di creazione, data di pubblicazione, tipo (*text, video, sound, image, 3D*), lingua, descrizione, link all'immagine e link a Wikipedia. In questo caso, tutti i campi devono essere aggiunti manualmente.

Alla fine del processo, un *popup* mostra all'utente il risultato della procedura di inserimento, indicando la presenza di eventuali errori. Nel caso in cui l'inserimento vada a buon fine, l'utente può visualizzare la nuova risorsa. Se invece il *record* da inserire è già presente nel Database, l'utente può visualizzare la risorsa già inserita e, se necessario, modificarla.

#### 4.2.4 Pagina delle collezioni



**Figura 8.** Pagina delle collezioni di WeME

In questa pagina, l'utente può visualizzare, modificare, esportare, aggiungere o eliminare una collezione. A fianco del titolo di ogni collezione ci sono vari pulsanti: il pulsante “View” (lente di ingrandimento) permette di visualizzare la collezione, dividendo le risorse per classe; il pulsante “Edit” (matita) permette di modificare la collezione, aggiungendo o rimuovendo risorse; il pulsante “Export” (documento con freccia) permette di esportare la collezione in due diversi formati (CSV o JSON-LD); il pulsante “Delete” (cestino) permette di eliminare la collezione.

#### 4.2.5 Pagina del Database

The screenshot shows the WeME interface with a navigation bar at the top containing the logo, user name, and buttons for 'Add Record', 'Collections', and 'Logout'. The main content area displays '34 records in the Database:' with an 'Export: CSV or JSON-LD' link. Below this are three yellow filter bars: '15 Persons', '12 Places', and '7 Cultural Heritage Object'. The '12 Places' filter is active, showing a table with columns: Name, Region, Country, Population, Lat,Long, and View. The table lists 12 cities with their respective data.

Name	Region	Country	Population	Lat,Long	View
Caprese, Caprese Michelangelo	Tuscany	Italy	1,551	43.633, 11.983	
Caserta, Caserta	Campania	Italy	76,693	41.067, 14.333	
Cecina, Cecina	Tuscany	Italy	28,152	43.306, 10.520	
Firenze, Florence	Tuscany	Italy	381,037	43.783, 11.250	
Londra, London	England	United Kingdom	7,556,900	51.509, -0.126	
Milano, Milan	Lombardy	Italy	1,343,163	45.467, 9.183	
Parigi, Paris	Île-de-France	France	2,240,621	48.857, 2.351	
Ravenna, Ravenna	Emilia-Romagna	Italy	80,868	44.413, 12.201	
Roma, Rome	Latium	Italy	2,318,895	41.892, 12.511	
Urbino, Urbino	Marche	Italy	15,566	43.717, 12.633	
Venice, Venice	Veneto	Italy	264,579	45.438, 12.336	
Verona, Verona	Veneto	Italy	259,069	45.433, 10.983	

Figura 9. Pagina del Database di WeME

In questa pagina l'utente può visualizzare l'intero Database: in alto viene indicato il numero totale dei records presenti, mentre nelle barre gialle viene indicato il numero di risorse per classe. Cliccando sulle barre, comparirà una tabella contenente tutte le risorse appartenenti ad una certa classe, con le informazioni basilari e un pulsante “View” (lente di ingrandimento) per visualizzare la risorsa nella pagina di visualizzazione. Inoltre, sono presenti i pulsanti per esportare l'intero Database in due diversi formati (CSV o JSON-LD).

## 4.2.6 Pagina dell'importazione

<WeME> | The Web Metadata Editor Welcome, Francesco! [Add Record](#) [Collections](#) [Logout](#)

### ➔ Import CSV files into the Database

Here you can import a .csv file into the Database. Be sure the structure of your file is correct.

**Instructions:**

- be sure your file contains the *correct number of fields*;
- the *field delimiter* must be a comma (,); all the commas (,) within the fields must be escaped correctly;
- the *the row delimiter* must be a end-of-line (\n); all the end-of-line (\n) within the fields must be escaped correctly;
- the field must be *enclosed in double quotes* ("); all the double quotes (") within the fields must be escaped correctly;

Persons key\_id,person\_id,name,surname,name\_surname,was\_born,was\_born\_year,died,died\_year,still\_alive,t

**Which type do you want to import?**

Select your option

**Upload your CSV File**

Scegli file Nessun file selezionato

Figura 10. Pagina dell'importazione di WeME

In questa pagina l'utente può importare un file CSV<sup>22</sup> all'interno del Database. Dato che l'importazione può causare diversi errori, la procedura segue alcune regole precise:

- è necessario specificare il tipo di dati che si vuole importare (*person*, *place*, o *CHO*);
- è opportuno verificare che il numero di campi nel file sia corretto;
- la prima riga del file deve contenere i campi della tabella corrispondente;
- il delimitatore dei campi deve essere una virgola (,);
- il delimitatore delle righe deve essere un a-capo (\n);
- tutto i campi devono essere contenuti entro delle virgolette doppie (“);
- tutte le virgole (,), gli a-capo (\n) e le virgolette (“) all'interno dei campi devono essere manipolati correttamente.

<sup>22</sup> Il *comma-separated values* (abbreviato in CSV) è un formato basato su file di testo, utilizzato per l'importazione ed esportazione (ad esempio da fogli elettronici o Database) di una tabella di dati. In questo formato, ogni riga della tabella (o record della base dati) è normalmente rappresentata da una linea di testo, che a sua volta è divisa in campi (le singole colonne) separati da un apposito carattere separatore, ciascuno dei quali rappresenta un valore (Wikipedia, voce *Comma-separated values*).

### 4.3 Tipologie di utenti

WeME può essere usato da varie categorie di utenza: una distinzione potrebbe essere fatta tra bibliotecari, archivisti e studiosi. Dal punto di vista dell'applicazione, bibliotecari e archivisti possono essere raggruppati nella stessa categoria, avendo specifiche abilità nella creazione di basi di conoscenza per collezioni di documenti; inoltre, è possibile che le persone appartenenti a queste due categorie siano specializzate in un determinato settore, per esempio la storia, i beni culturali, ecc. Il loro interesse principale è raccogliere tutti metadati rilevanti per facilitare la scoperta, la classificazione e l'esplorazione di cataloghi. Gli studiosi, invece, sono più orientati verso la redazione di una *knowledge base* per rispondere alle loro esigenze di ricerca, senza necessariamente essere specializzati in un certo settore.

Alcuni progetti sul Web, come l'*Irish Record Linkage* (Debruyne, 2016), implementano, all'interno della stessa piattaforma, due basi di conoscenza differenti: una è pensata per archivisti, mentre l'altra è adatta per studiosi. WeMe differisce da questa strategia, costruendo una sola *knowledge base* che tenta di soddisfare i bisogni di entrambe le parti. Dal punto di vista degli archivisti, l'utilizzo di Linked Data da parte dell'applicazione facilita l'interoperabilità e la connessione con altri archivi pubblicati sul Web, permettendo di stabilire nuove relazioni tra risorse informative collocate in punti differenti della rete. Per quanto riguarda gli studiosi, il collegamento con le fonti esterne e le relazioni tra risorse all'interno della base di conoscenza permettono di estrarre nuova informazione, grazie all'utilizzo di alcuni strumenti di inferenza. Attualmente, WeME non implementa tecniche di inferenza, ma potrebbe essere una funzione di particolare interesse per gli sviluppi futuri.



## 5 Test di usabilità

Per testare l'usabilità di WeME è stato predisposto un questionario per guidare gli utenti nell'uso delle varie funzioni dell'applicazione: sono stati impostati diversi scenari di utilizzo, in modo da verificare l'efficienza delle varie funzionalità. Il test proposto aveva un duplice obiettivo: capire il grado di apprezzamento dell'applicazione da parte degli utenti e ottenere suggerimenti per il miglioramento della stessa.

Il questionario è stato inoltrato in diverse *mailing list* relative alle tematiche dei beni culturali. In totale, hanno partecipato 26 utenti, di cui il 61,5% uomini e il rimanente donne. Le Figure 11 e 13 mostrano rispettivamente la distribuzione dell'età degli utenti e le loro competenze. Come si evince dalla Figura 13, il 38,5% degli utenti è esperto nel settore dell'informatica, un altro 38,5% è un ricercatore nel campo delle Digital Humanities, mentre solo il 3% ha competenze archivistiche. Ad ogni modo, di tutti gli utenti, solo il 69,2% ha mostrato un chiaro interesse nel settore dei beni culturali (v. fig. 12).

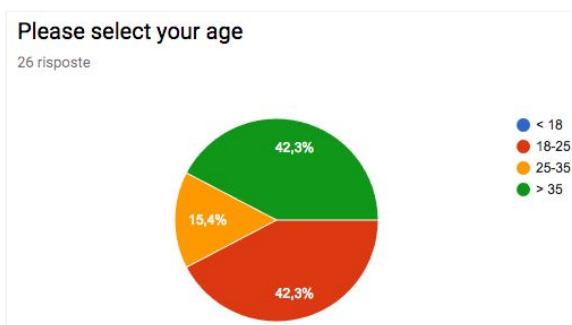


Figura 11. Distribuzione d'età degli utenti

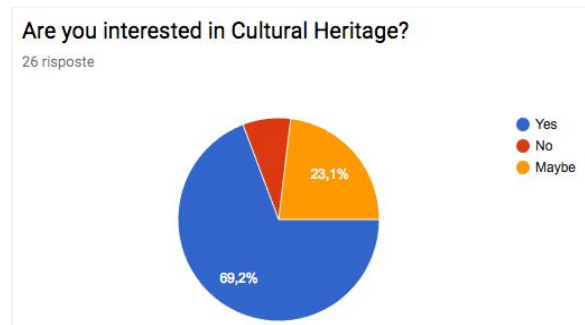


Figura 12. Interesse nel settore dei beni culturali

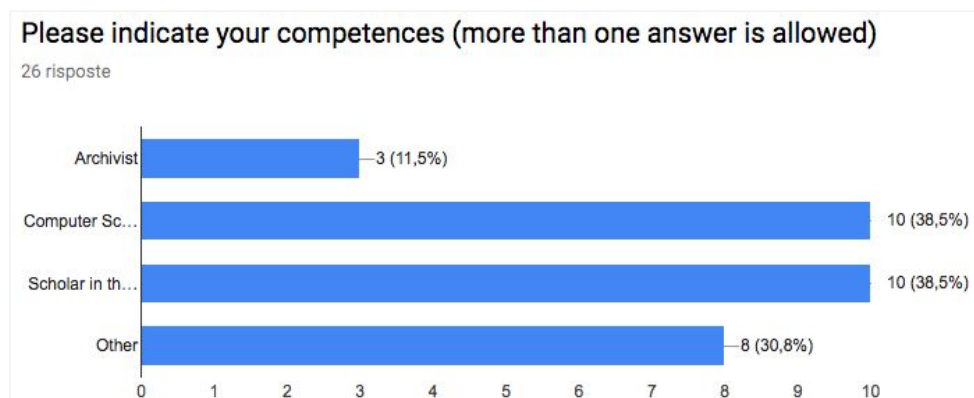


Figura 13. Settori di competenza degli utenti

Il test è stato organizzato nell'esecuzione delle seguenti attività:

- [T1] creazione di un account;
- [T2] gestione di una collezione;
- [T3] gestione di nuove risorse (persona, luogo o CHO);
- [T4] giudizio generale.

Task	Scenario	Domande	Tipo di risposta
T1	Creazione manuale dell'account all'interno dell'applicazione	a) Quanto è stato difficile creare l'account?	a) Scala da "molto difficile" a "molto facile"
T2	Creazione di una nuova collezione	a) Quanto è stato difficile creare una collezione?	a) Scala da "molto difficile" a "molto facile"
	Aggiunta di elementi alla collezione (una persona, un luogo e un CHO)	a) Quanto è stato difficile aggiungere elementi a una collezione?	a) Scala da "molto difficile" a "molto facile"
	Export della collezione in CSV	a) Quanto è stato difficile esportare una collezione?	a) Scala da "molto difficile" a "molto facile"
T3	Aggiunta di risorse al Database mediante una ricerca con DBpedia, VIAF e GeoNames (una persona, un luogo e un CHO)	Per ogni categoria: a) Che risorsa hai aggiunto? (facoltativa) b) Hai trovato i dati usando DBpedia? c) Hai trovato i dati usando VIAF? d) Hai trovato i dati usando GeoNames?	a) Risposta breve b-c-d) Scelta multipla tra "sì", "no" e "in parte"
		e) Quanto è stato difficile aggiungere elementi al Database?	e) Scala da "molto difficile" a "molto facile"
	Visualizzazione delle risorse aggiunte in precedenza	a) Quanto è stato difficile cercare le risorse? b) Pensi che la base di conoscenza sia organizzata bene?	a) Scala da "molto difficile" a "molto facile" b) Scala da "molto confusionaria" a "molto chiara"
T4	Considerazioni generali su WeME	a) Quanto è stato difficile navigare in WeME? b) Suggerimenti per migliorare la navigazione (facoltativa)	a) Scala da "molto difficile" a "molto facile" b) Risposta aperta
		c) Che cosa ne pensi della grafica di WeME? d) Suggerimenti per migliorare la grafica (facoltativa)	c) Scala da "scadente" a "eccellente" d) Risposta aperta
		e) Pensi che le collezioni siano ben organizzate?	e) Scelta multipla tra "sì", "no" e "forse"
		f) Giudizio complessivo sull'applicazione	f) Scala da 1 a 5

**Tabella 3.** Scenari di utilizzo e rispettive domande

## 5.1 Creazione dell'account

E' stato richiesto agli utenti di creare il proprio account nell'applicazione, utilizzando l'apposito menu. Una volta creato, è stato chiesto loro di effettuare l'accesso. In generale, non si sono manifestate difficoltà nella procedura: circa il 75% degli utenti, infatti, ha trovato il processo molto facile o facile, un altro gruppo l'ha considerato di media difficoltà, mentre solamente un utente ha incontrato delle complicazioni (v. fig. 14).

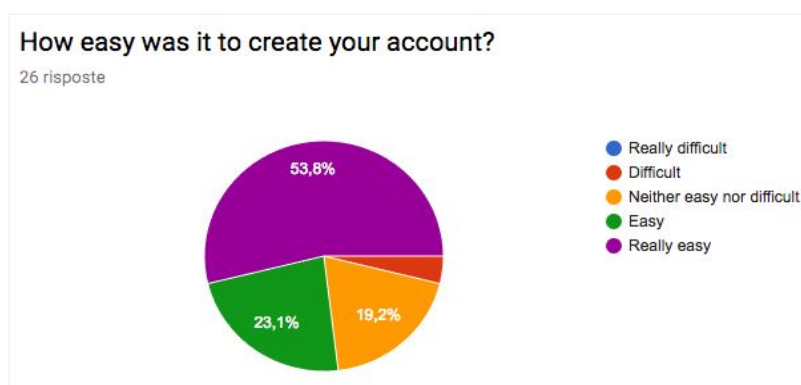


Figura 14. Difficoltà della procedura di creazione dell'account

## 5.2 Gestione di una collezione

E' stato richiesto agli utenti di testare le varie funzioni per la gestione di una collezione, ovvero creazione, inserimento ed esportazione. Come si vede nella Figura 15, la procedura di creazione non ha creato evidenti difficoltà, e per questo più dell'80% degli utenti l'ha definita facile o molto facile.

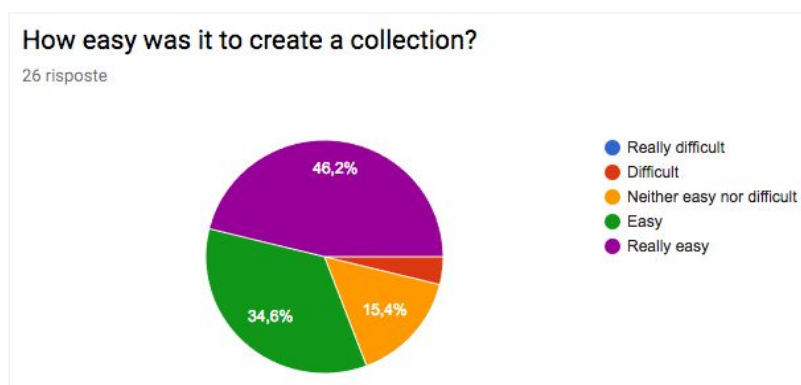


Figura 15. Difficoltà della creazione di una collezione

L'inserimento di risorse all'interno della collezione e l'esportazione della stessa (v. fig. 16 e fig. 17), invece, hanno portato più problemi: più della metà delle persone ha trovato la procedura facile o molto facile, ma un numero maggiore di utenti l'ha ritenuta difficile o molto difficile.

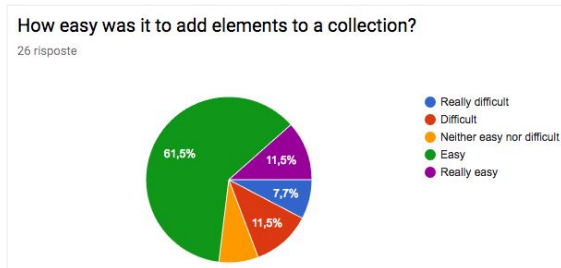


Figura 16. Difficoltà dell'inserimento in una collezione

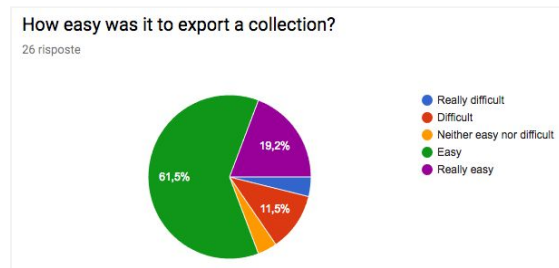


Figura 17. Difficoltà dell'esportazione di una collezione

## 5.3 Gestione di nuove risorse

### 5.3.1 Ricerca e inserimento

E' stato richiesto agli utenti di inserire delle risorse nel Database tentando di recuperare i dati da DBpedia, VIAF e GeoNames. In particolare, è stato richiesto di aggiungere una persona, un luogo e un CHO che fossero correlati tra loro (ad esempio "Dante Alighieri", "Firenze", "La Divina Commedia"), in modo da visualizzare il collegamento tra i vari *records*. Alla fine, è stato chiesto di valutare la complessità dell'intera procedura: come si vede nella Figura 18, quasi il 70% degli utenti ha ritenuto l'operazione facile o molto facile, mentre la restante parte ha incontrato problemi di natura tecnica, che sono stati specificati nella sezione dedicata ai suggerimenti.

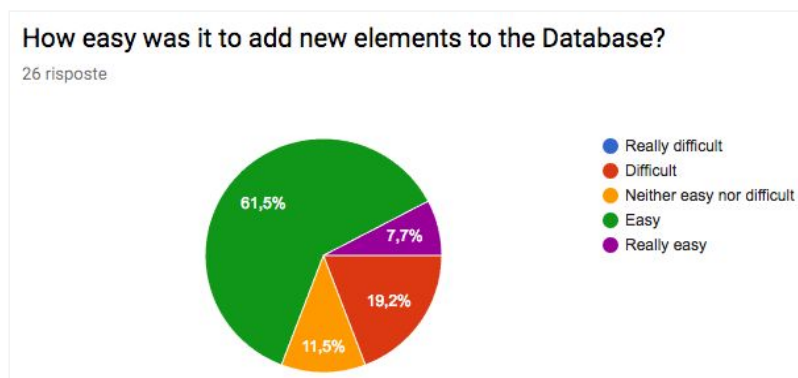
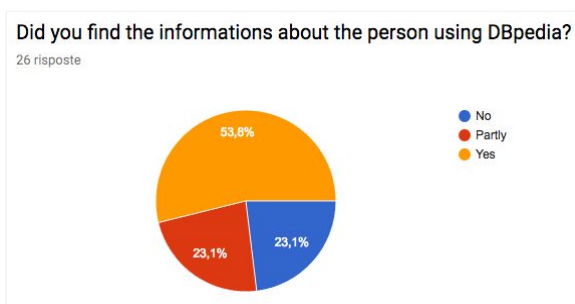


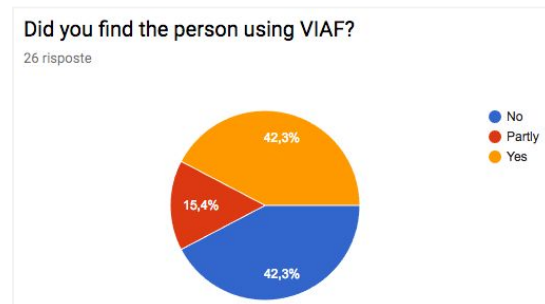
Figura 18. Difficoltà della procedura di inserimento delle risorse nel Database

### 5.3.1.1 Ricerca e inserimento di persone

Per quanto riguarda le persone, sono state aggiunti personaggi appartenenti a diverse categorie (ad esempio Miguel de Cervantes, Barack Obama, Eugenio Montale, Stephen King, ecc.) Come si vede nella Figura 19, più del 50% degli utenti è riuscito a ottenere informazioni tramite DBpedia, e un'altra buona parte degli utenti ha ottenuto almeno delle informazioni parziali. La ricerca tramite VIAF, invece, si è rivelata meno proficua, mostrando una divisione più equa tra persone che hanno ottenuto informazioni e altre che non le hanno ottenute (v. fig. 20).



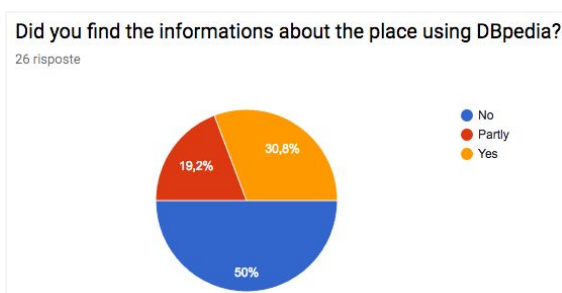
**Figura 19.** Informazioni sulle persone recuperate tramite DBpedia



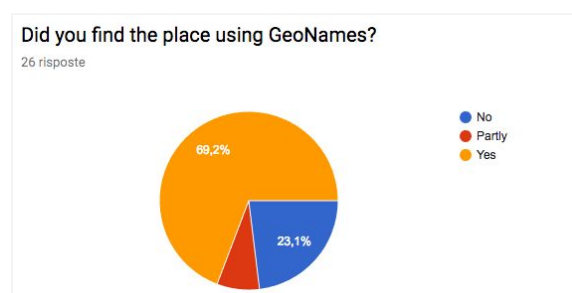
**Figura 20.** Informazioni sulle persone recuperate tramite VIAF

### 5.3.1.2 Ricerca e inserimento di luoghi

Gli utenti hanno aggiunto diverse località correlate ai personaggi inseriti in precedenza (ad esempio Madrid, Milano, New York, Portland, ecc.) Per quanto riguarda le risorse di tipo geografico, DBpedia si dimostra meno efficace: come si vede nella Figura 21, infatti, la metà degli utenti non è riuscita a recuperare informazioni. Tuttavia, la ricerca tramite GeoNames ha funzionato molto bene, permettendo di recuperare dati nel 70% dei casi (v. fig. 22).



**Figura 21.** Informazioni sulle località recuperate tramite DBpedia



**Figura 22.** Informazioni sulle località recuperate tramite GeoNames

### 5.3.1.3 Ricerca e inserimento di CHO

Questa fase non prevede ricerche automatiche, pertanto è stato semplicemente chiesto agli utenti di inserire manualmente un CHO, in modo che la persona aggiunta in principio ne fosse l'autore. Tra le risorse aggiunte possiamo citare "Don Quijote de la Mancha", "Ossi di Seppia", "IT", ecc.

### 5.3.2 Visualizzazione

Una volta completata la procedura di inserimento, è stato chiesto agli utenti di cercare, nella base di conoscenza, le risorse appena aggiunte, per poi valutare la difficoltà del processo e dare un giudizio sull'organizzazione della *knowledge base*. La ricerca ha generato pareri contrastanti, ed è l'argomento per il quale sono stati forniti più suggerimenti: circa il 50% delle persone ha trovato il processo facile o molto facile, mentre il resto degli utenti ha incontrato dei problemi o ha suggerito dei miglioramenti (v. fig. 23). Nonostante ciò, come si evince dalla Figura 24, gran parte delle persone ha trovato chiara l'organizzazione delle informazioni, la quale non ha creato complicazioni significative.

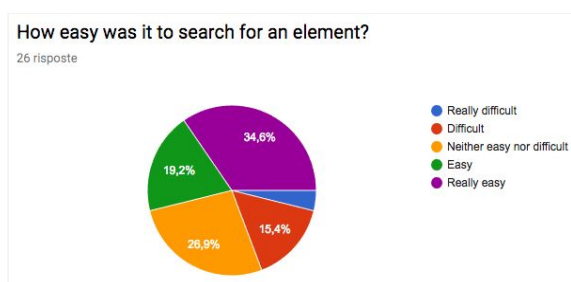


Figura 23. Difficoltà della procedura di ricerca all'interno della base di conoscenza

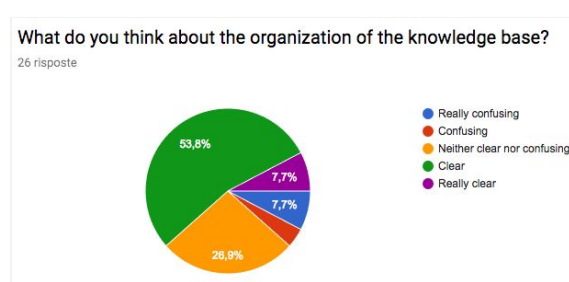


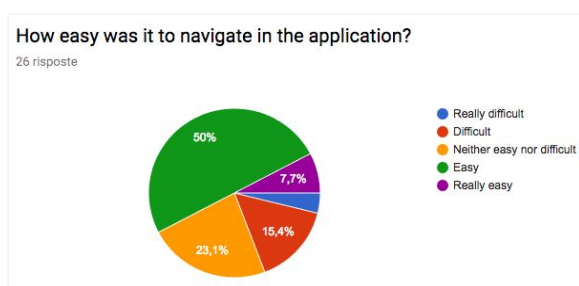
Figura 24. Giudizio sull'organizzazione delle informazioni nella base di conoscenza

## 5.4 Giudizio generale

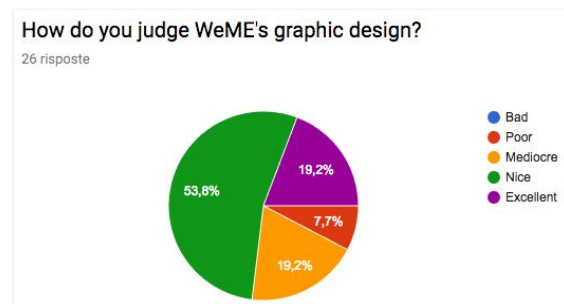
Nell'ultima fase del test è stato chiesto agli utenti di esprimere dei giudizi generali sull'applicazione, in particolare per quanto riguarda la navigazione, la grafica e l'organizzazione dei documenti. Inoltre, sono stati inseriti dei campi facoltativi in cui poter inserire ogni tipo di suggerimento, dalla segnalazione di problemi fino alla proposta di nuove funzionalità. Quest'ultima sezione ha evidenziato la presenza di qualche *bug* all'interno del codice, che talvolta ha impedito ad alcuni utenti di completare la procedura richiesta.

Per quanto riguarda la navigazione, più del 50% degli utenti è rimasta soddisfatta (v. fig. 25). Tuttavia, un numero consistente di persone ha segnalato complicazioni di vario tipo, e ha comunicato dei suggerimenti per risolverli.

L'impostazione grafica dell'applicazione, come si vede nella Figura 26, è stata molto apprezzata: solo 7 utenti l'hanno definita mediocre o scadente, mentre il resto delle persone ha espresso un parere positivo. In generale, si è manifestata la necessità di rendere l'applicazione più *responsive* e adattabile a dispositivi di diverse dimensioni.

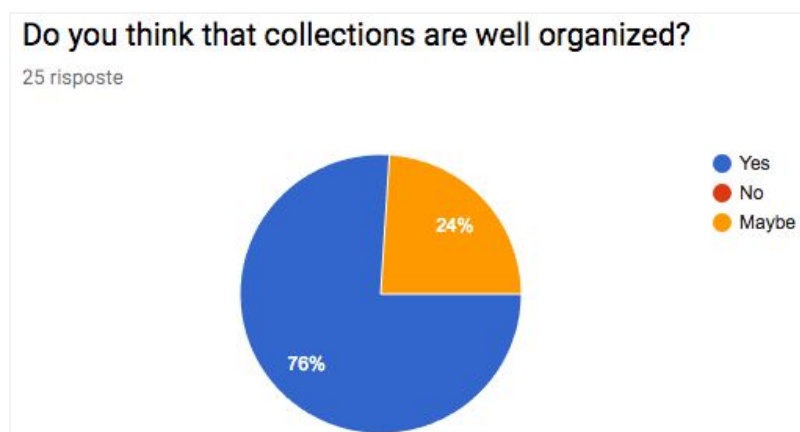


**Figura 25.** Giudizio sulla navigazione nell'applicazione



**Figura 26.** Giudizio sulla grafica dell'applicazione

L'organizzazione delle collezioni è stata molto soddisfacente, dal momento che nessun utente ha espresso giudizi negativi. Alcune persone hanno fornito dei suggerimenti per migliorarne l'efficacia, ma non hanno evidenziato problemi di alcun tipo.



**Figura 27.** Giudizio sull'organizzazione dei documenti

La tabella seguente riassume l'insieme dei suggerimenti che sono stati forniti dagli utenti, per il miglioramento della grafica, della navigazione e dell'organizzazione. In generale, si è manifestata la necessità di correggere alcuni errori e di rendere la piattaforma più intuitiva per

ogni tipo di utente. Inoltre, sono stati comunicati suggerimenti più specifici per il miglioramento e l'estensione delle funzioni testate.

Tema	Idee Proposte
Qualità della grafica	Rendere l'applicazione più <i>responsive</i> e adattabile a diversi dispositivi
Navigazione dei contenuti	Migliorare la chiarezza del tasto "Home"
	Migliorare la navigazione tramite browser sfruttando il tasto "Indietro"
	Dare la possibilità di cercare una risorsa senza specificare la classe
	Avviare la ricerca di una risorsa nel Database premendo il tasto "Invio"
	Effettuare una ricerca all'interno del Database anche con stringhe non esatte, ad esempio "montale" invece di "Eugenio Montale"
	Mostrare un <i>popup</i> di avvertimento quando la ricerca non dà risultati
Gestione dei documenti	Dare la possibilità di modificare una collezione gestendo più risorse contemporaneamente

Tabella 4. Suggerimenti degli utenti suddivisi per categoria

L'ultima domanda chiedeva di esprimere un giudizio complessivo su WeME, valutando l'applicazione in una scala da 1 a 5 (v. fig. 28). Molti utenti hanno dato un giudizio positivo (4 o 5), esprimendo interesse per le potenzialità del progetto. Altre persone hanno scelto un giudizio intermedio (3), sottolineando la presenza di alcuni problemi che, tuttavia, non hanno scoraggiato dal considerare WeME uno strumento con un grande potenziale di miglioramento.

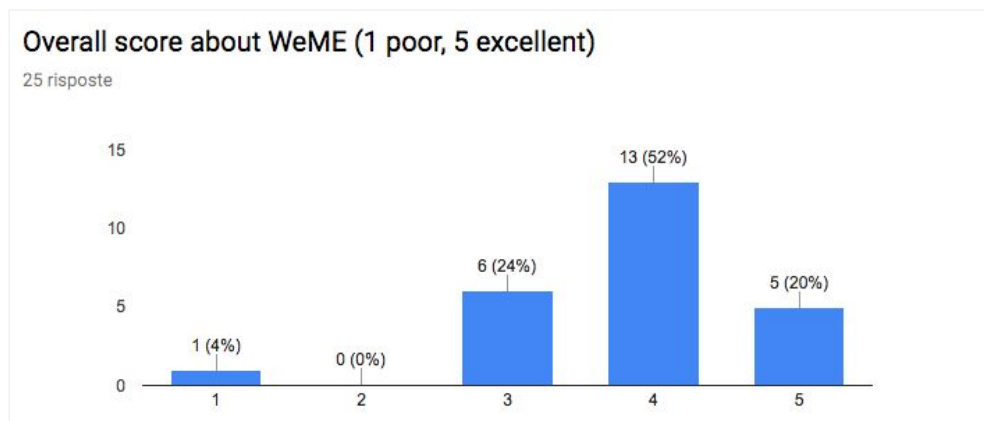


Figura 28. Giudizio complessivo sull'applicazione



## 6 Conclusione e Prossimi Lavori

In questo lavoro di tesi è stato presentato il Web Metadata Editor (WeME), un'Applicazione Web per la creazione di una base di conoscenza a corredo di documenti digitali. La piattaforma è pensata per archivisti e ricercatori, ma può essere utilizzata, in generale, da qualsiasi tipo di utente. WeME riesce a soddisfare i bisogni di tutte le categorie di utenza, velocizzando e rendendo più efficiente la raccolta di metadati, grazie alla connessione con diversi *Web Services* (DBpedia, VIAF e GeoNames). Inoltre, sfruttando i Linked Data, l'applicazione garantisce l'interoperabilità dei dati raccolti, ed entra a far parte di quell'insieme di strumenti e tecnologie che mirano al perfezionamento e al rafforzamento del Web Semantico. La procedura di test che è stata effettuata, alla quale hanno partecipato 26 utenti, ha confermato l'utilità e le potenzialità della piattaforma: il progetto ha stimolato l'interesse di utenti che si occupano di ricerca, archiviazione e catalogazione, i quali hanno sottolineato la necessità di uno strumento in grado di svolgere le funzioni implementate nell'applicazione, e per questo l'hanno ben accolta. La piattaforma può essere perfezionata estendendo le funzionalità, per rispondere a una moltitudine sempre maggiore di necessità. In particolare, si è ipotizzato di estendere l'applicazione con le seguenti caratteristiche:

- collegamento con altri nodi Linked Data e archivi online;
- disambiguazione automatica dei dati;
- esportazione della base di conoscenza in ulteriori formati (es. RDF, XML, ecc.);
- compatibilità con diverse ontologie (es. bibo);
- supporto di altre classi di risorse, come ad esempio gli eventi;
- implementazione di tecniche di inferenza;
- compatibilità con altre lingue;
- gestione più approfondita del profilo utente;
- maggiore configurabilità della piattaforma da parte dell'utente.

Ad ogni modo, l'applicazione si è rivelata già di grande utilità, specialmente per chi lavora nel settore degli archivi o in quello delle Digital Humanities, poiché dà all'utente la possibilità di aumentare, in modo non indifferente, il livello di efficienza del suo lavoro.

## Bibliografia

(Abele, McCrae, 2007-2017) *Linking Open Data cloud diagram*, by Andrejs Abele, John P. McCrae, Paul Buitelaar, Anja Jentzsch and Richard Cyganiak, 2007-2017, <http://lod-cloud.net>

(California Digital Library, 2001) California Digital Library, *Digital Object Standard: Metadata, Content and Encoding*, May 18, 2001

(Debruyne, 2016) Debruyne, C., Beyan, O.D., Grant, R., Collins, S., Decker, S., Harrower, N.: *A semantic architecture for preserving and interpreting the information contained in irish historical vital records*. International Journal on Digital Libraries 17(3), 159– 174 (2016)

(Freeney, 2014) Feeney, K.C., D. O’Sullivan, W. Tai e R. Brennan, *Improving curated web-data quality with structured harvesting and assessment*. Int. J. Semantic Web Inf. Syst. 10(2), 35–62 (Apr 2014)

(Hampson, 2012) Hampson, C., S. Lawless, B. Eoin, S. Yogev, N. Zwerdling, D. Carmel, O. Conlan, A. O’Connor e V. Wade, *CULTURA: A Metadata-Rich Environment to Support the Enhanced Interrogation of Cultural Collections*, pp. 227–238 (2012)

(Heath, 2011) Heath, Tom e Christian Bizer, *Linked Data: Evolving the Web into a Global Data Space (1st edition)*, 2011

(Tammaro, 2005) Tammaro, Anna Maria, *Che cos'è una biblioteca digitale?*, vol. 1, 2005

# Sitografia

*Cold Start KBP Track*, <https://tac.nist.gov/2016/KBP/ColdStart/index.html>

*Collection Space*, <http://www.collectionspace.org>

*Collective Access*, <http://collectiveaccess.org>

*DBpedia*, <http://wiki.dbpedia.org>

*Dublin Core*, <http://dublincore.org/documents/dces>

Europeana Website, pagina *EDM Documentation*, <http://pro.europeana.eu/page/edm-documentation>

*FREME Project*, <http://www.freme-project.eu/the-project/what-is-freme>

*GeoNames*, <http://www.geonames.org>

*Google Knowledge Graph*, <https://www.google.it/intl/it/insidesearch/features/search/knowledge.html>

*Library of Congress*, <https://www.loc.gov>

OCLC Website, pagina *CONTENTdm*, <http://www.oclc.org/en/contentdm.html>

*Omeka*, <http://omeka.org>

*Oxford Digital Library*, <http://www.odl.ox.ac.uk>

*Perseus Digital Library*, <http://www.perseus.tufts.edu>

*VIAF*, <https://viaf.org>

W3C Website, pagina *RDF*, <https://www.w3.org/TR/2014/REC-rdf-schema-20140225>

W3C Website, pagina *SPARQL*, <https://www.w3.org/TR/2013/REC-sparql11-protocol-20130321>

*Watson Web interface*, <http://watson.kmi.open.ac.uk/WatsonWUI>

Wikipedia, voce *Metadato*, <https://it.wikipedia.org/wiki/Metadato>

Wikipedia, voce *Comma-separated values*, [https://it.wikipedia.org/wiki/Comma-separated\\_values](https://it.wikipedia.org/wiki/Comma-separated_values)

Wikipedia, voce *Uniform Resource Identifier*, [http://wikipedia.org/wiki/Uniform\\_Resource\\_Identifier](http://wikipedia.org/wiki/Uniform_Resource_Identifier)

Wikipedia, voce *Web Semantico*, [https://it.wikipedia.org/wiki/Web\\_semantico](https://it.wikipedia.org/wiki/Web_semantico)

XML Namespaces Website, pagina *FOAF*, <http://xmlns.com/foaf/spec>