Cosa sono Wikidata e Wikibase

Wikidata è una **Knowledge-Base collaborativa** ("collaboratively edited knowledge base" o "free linked database"), mantenuta dalla Wikimedia Foundation e aperta in consultazione e modifica a **utenti** e **robot** automatizzati.

Wikidata svolge una funzione di archiviazione centrale dei "dati puntuali" ("structured data", come date o luoghi) contenuti in altri progetti Wikimedia quali Wikipedia, Wikiquote, Wikisource, Wikibooks ed altri.

La centralizzazione dei dati comuni è mirata a facilitare

- la loro manutenzione, in modo indipendente dalle specificità dei diversi progetti Wikimedia (per esempio le diverse lingue)
- ed il loro ri-utilizzo in nuovi contesti, con un'impostazione Linked Open Data.

L'iniziativa è stata avviata da un gruppo di ricercatori di **Wikimedia Deutschland**, con il supporto della Paul Allen Institute for Artificial Intelligence, dalla Gordon and Betty Moore Foundation e da Google.

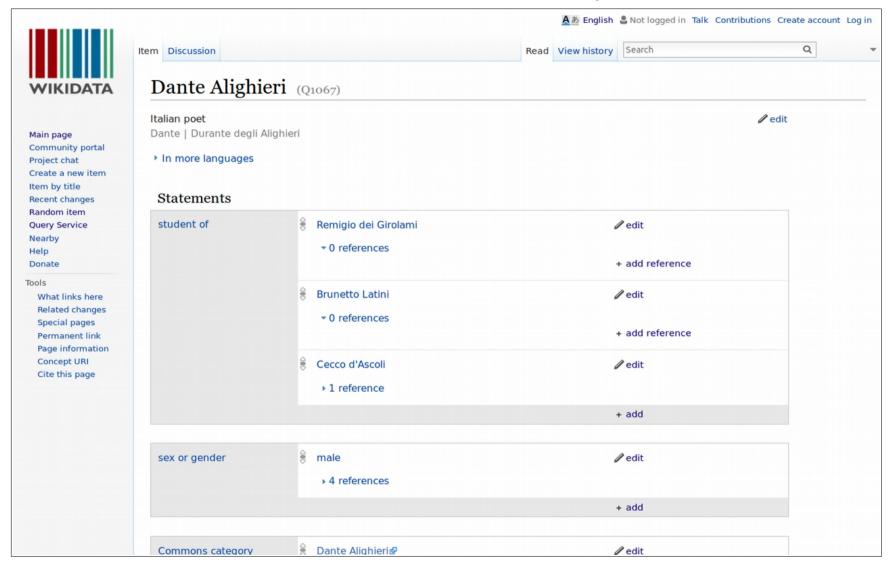
Wikibase è il software libero con il quale il servizio viene gestito

Il funzionamento di Wikidata dal punto di vista dell'utenza

Gli articoli nelle differenti lingue di Wikipedia riguardanti Dante...



...puntano al medesimo Item Wikidata, dove sono raccolti i "dati puntuali"



Il Wikibase data-model

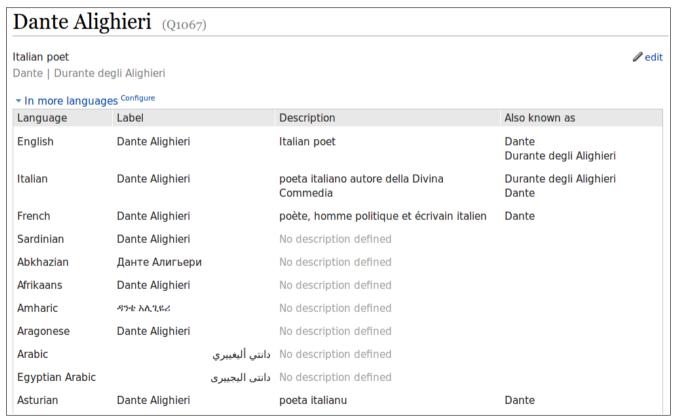
Le entità trattate da Wikibase si distinguono in Item (elementi) e Property (proprietà)

L'Item di Wikidata è identificato da un "Item Identifier": un numero preceduto da una "Q"

Nell'esempio: Q1067

Si noti che qui "item" non ha il significato FRBR di esemplare ma di elemento

...e dal Fingerprint: cioè Label, Alias e Description nelle diverse lingue



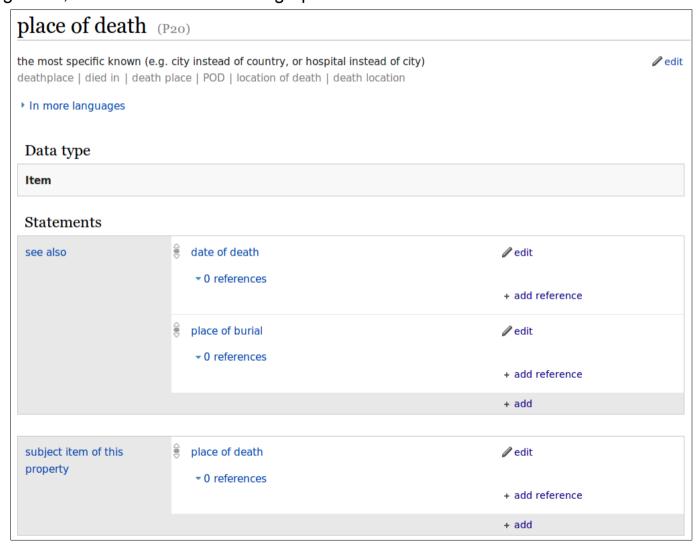
I puntatori alle diverse pagine dei progetti Wikimedia sono gestiti nella sezione Link dell'Item



Gli Item sono "descritti" mediante una serie di Statement

place of death		Ravenna		/ edit
		▶ 2 references		
				+ add
occupation	QOD	poet		∥ edit
		▶ 1 reference		
	000	politician		∥ edit
	_			
date of birth	40 0	May 1265		
		earliest date	21 May 1265	
		latest date	1 June 1265	
		▼ 5 references		
		stated in	Integrated Authority File	
		retrieved	9 April 2014	
		stated in	Dizionario Biografico degli Italiani	
		reference URL	http://www.treccani.it /enciclopedia/dante- alighieri_%28Dizionario_Biogr afico%29/&	
		retrieved	8 February 2015	

Gli Statement si basano sulle Property, come avviene di consueto nelle triplette del web-semantico Le Property, come gli Item, sono identificate con Fingerprint e descritte con Statement



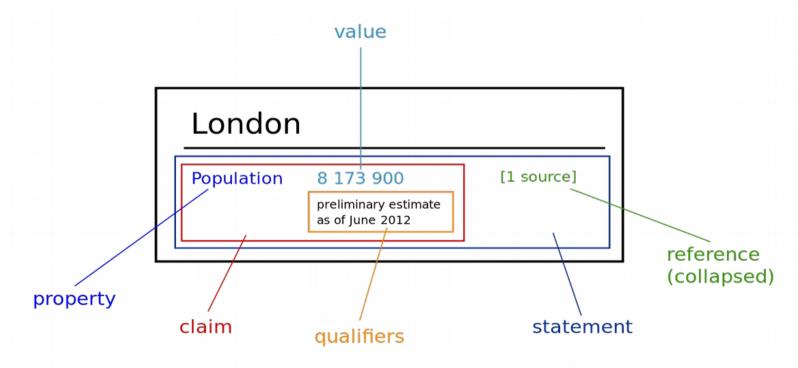
..ma gli Statement di Wikibase integrano la tripletta semantica con

Qualifier

descritti anch'essi con Property e Target: qualificano l'informazione

e Source ("Reference")

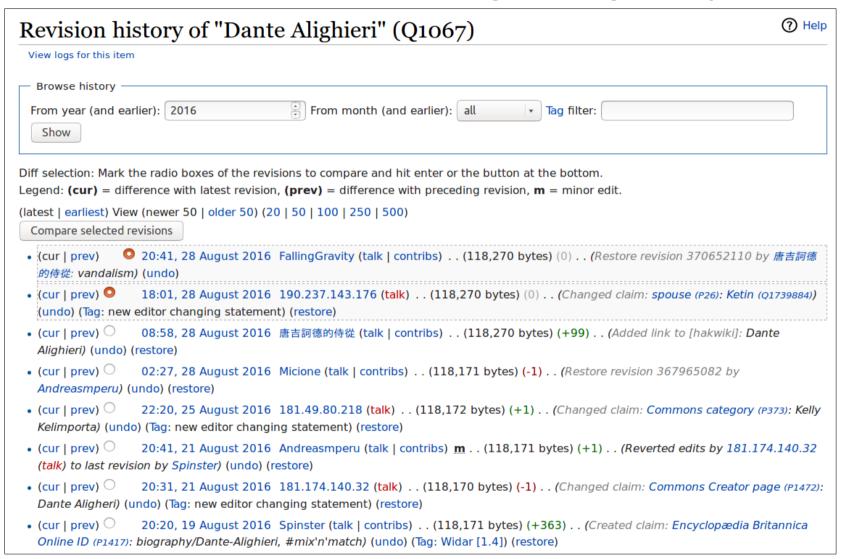
anche questi descritti con Property e Target: indicano la fonte dell'informazione



La soluzione Wikibase è comunque compatibile con lo standard RDF:

è una semplificazione ritenere che per la "grammatica del web semantico = RDF" siano sufficienti solo Soggetto Predicato e Oggetto

Tutte le revisioni dell'Item sono conservate nella cronologia, e rimangono sempre accessibili



Oggi molti sono impegnati a disegnare e a sperimentare percorsi di transizione

Per "tradurre" il MARC e i suoi dialetti in "Linked data /RDF" esistono varie strategie:

- 1. ristrutturazione dell'informazione (senza garantire un ritorno al formato di origine)
- 2. **mappatura** di tutti gli elementi del MARC (con garanzia di ritorno): realizzata per il MARC21 ma non sperimentata in nessun progetto → http://www.marc21rdf.info/
- 3. ristrutturazione dell'informazione e mappatura (garantendo un ritorno al formato di origine)

Le "ristrutturazioni" possono essere di vario tipo: l'analisi FRBR è il modello di dati con il quale tutti i progetti si confrontato

Wikibib https://it.wikipedia.org/wiki/Progetto:GLAM/Biblioteca_Nazionale_Centrale_di_Firenze#Wikibib è un progetto ancora informale nato in incontri GLAM alla BNCF: si propone di usare la strategia 3 attraverso l'import dei dati bibliografici nella piattaforma collaborativa Wikibase

Test di caricamento MARC in Wikibase: il caso UNIMARC proveniente da SBN

Replica in Wikibase di tutto il dettaglio informativo del record bibliografico unimarc

La sperimentazione qui descritta mira ad ottenere, come base per futuri ri-utilizzi dei dati, una replica in Wikibase di tutto il dettaglio informativo del record bibliografico UNIMARC (così come esposto dall'Indice).

LEADER 05771nam0 2201813 i 4500 001 IT\ICCU\CFI\0893220 003 http://id.sbn.it/bid/CFI0893220 005 20150525130956.6 010 \$a978-88-581-1170-3 020 \$aIT\$b2014-2498 100 \$a20140514d2014 ||||0itac50 ba 101 \$aita\$cita 102 \$aIT 181 1\$6z01\$ai \$bxxxe 182 1\$6z01\$an 183 1\$6z01\$cnc\$2rdacontent 200 1 \$a<<Lo >>Stato innovatore\$esfatare il mito del pubblico contro il privato\$fMariana Mazzucato\$gtraduzione di Fabio Galimberti 210 \$aRoma\$aBari\$cGLF editori Laterza\$d2014 215 \$aXXVI, 351 p.\$d21 cm 410 0\$1001IT\ICCU\RAV\1797816\$12001 \$aAnticorpi\$v43 500 10\$a<<The >>entrepreneurial State\$3IT\ICCU\CFI\0893222\$9Mazzucato, Mariana 606 \$alnnovazione tecnologica\$xInterventi statali\$2FN \$3IT\ICCU\CFIC\227454 676 \$a338.45\$cINDUSTRIE SECONDARIE E SERVIZI. EFFICIENZA PRODUTTIVA\$v23 700 1\$aMazzucato\$b, Mariana\$3IT\ICCU\UFIV\143190 702 1\$aGalimberti\$b. Fabio\$f <1972- >\$3IT\ICCU\CFIV\224780 801 3\$aIT\$bICCU\$c20160705 899 \$1AL0114\$2TO0 49\$fP/G\$qN

I campi MARC sono archiviati come Statement, gli indicatori ed i sotto-campi come Qualifier

Titolo e indicazione di responsabilità [bib]	Lo Stato innovatore : sfat privato / Mariana Mazzuo Galimberti	care il mito del pubblico contro il cato ; traduzione di Fabio	[modifica]
	Indicatore di titolo significativo [bib]	Titolo significativo (Indicatore titolo significativo)	
	Titolo proprio [bib]	Lo Stato innovatore	
	Complemento del titolo (Titolo e indicazione di responsabilità) [bib]	sfatare il mito del pubblico contro il privato	
	Prima indicazione di responsabilità' [bib]	Mariana Mazzucato	
	Altre indicazioni di responsabilità' [bib]	traduzione di Fabio Galimberti	
	MARC ordine sottocampi	a[1] e[1] f[1] g[1]	
	▶ 1 riferimento		
			[aggiungi]

Tutti gli elementi MARC (campi, indicatori e sottocampi) sono mappati con specifiche Property



Il Fingerprint di ciascun Item contiene l'dentificativo (BID) del record

In questo modo i record/Item potranno essere individuati in base all'identificativo assegnato nel sistema di origine.



Ciascun campo (Statement) contiene la fonte (Source, Reference) da cui proviene il dato

Vengono registrati, con apposite Property:

- l'istituzione da cui proviene il record, con un puntatore al relativo Item
- il puntatore al record originale, con permanent URL alla fonte



Un apposito campo/Qualifier registra l'esatto ordine dei sottocampi di origine.

I campi MARC contenenti voci di autorità sono gestiti con puntatori ad Item esterni

Nell'Item del record bibliografico, per gestire le intestazioni ai nomi si registrano

- il **puntatore** all'Item esterno (si noti che Wikibase espone automaticamente la Label dell'Item collegato)
- e i dati specifici (se presenti) come il Relator-code dell'intestazione al nome

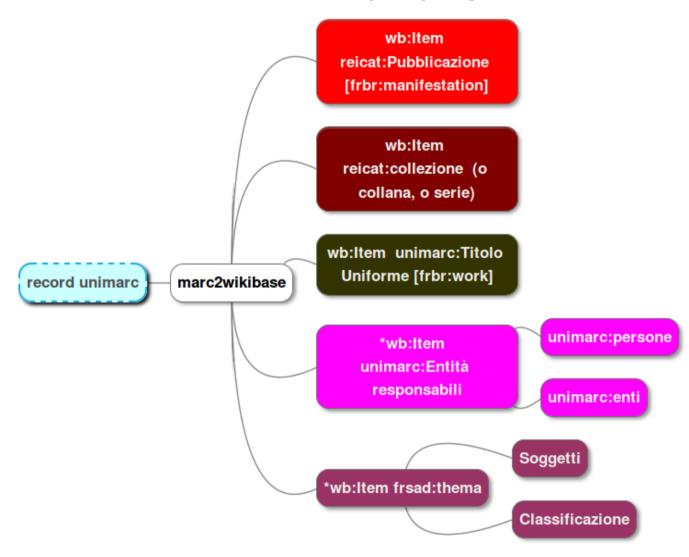


La voce della lista di autorità diventa un Item separato

L'identificativo della voce di autorità (in questo caso il VID) viene registrato nel Fingerprint dell'Item. L'authority-record unimarc viene registrato con Statement e Property come il record bibliografico.



Ristrutturazione del record unimarc in Item distinti per tipologia di dato



Oltre 160 Entity (fra Item e Property) per caricare tutti i dati del record unimarc di esempio

Il record di esempio ha il proprio Item:

Record bibliografico: Lo Stato innovatore

Tutte le Entity collegate al record sono destinate ad essere riutilizzate in connessione con altre Entity:

Item per i record collegati e le voci di autorità:

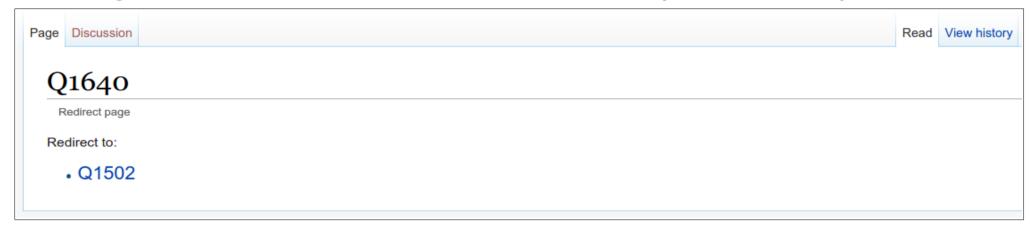
- Record bibliografico (serie): Anticorpi
- Responsabilità: Galimberti, Fabio <1972- >
- Responsabilità: Mazzucato, Mariana
- Voce classificazione: 338.45 (ed. 23) INDUSTRIE SECONDARIE E SERVIZI. EFFICIENZA PRODUTTIVA
- Voce soggetto: Innovazione tecnologica Interventi statali
- Titolo uniforme: The entrepreneurial State

Item e Property per la mappatura della sintassi Unimarc:

- Circa 40 Item separati utilizzati per archiviare i coded-data
- 116 Property

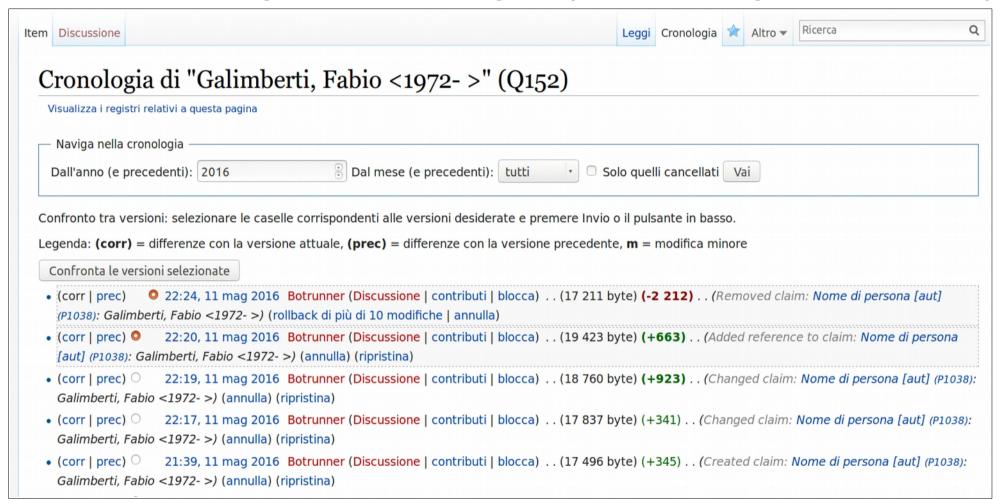
Principali potenzialità tecniche di Wikibase, per l'informazione bibliografica

Wikibase gestisce il reindirizzamento nel caso di "sostituzioni" ("schiacciamenti") di Item





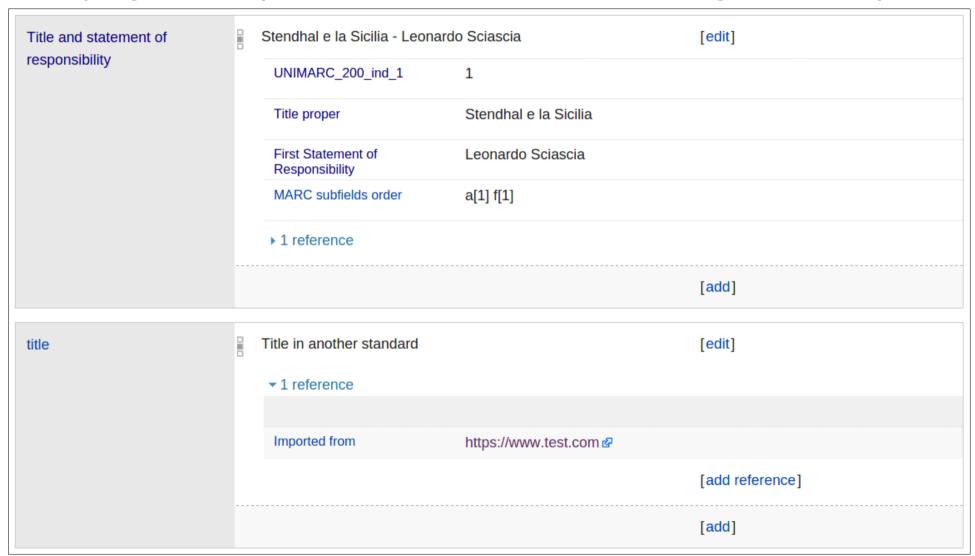
Viene conservata la cronologia di tutte le versioni degli Item (con il riferimento agli autori delle modifiche)



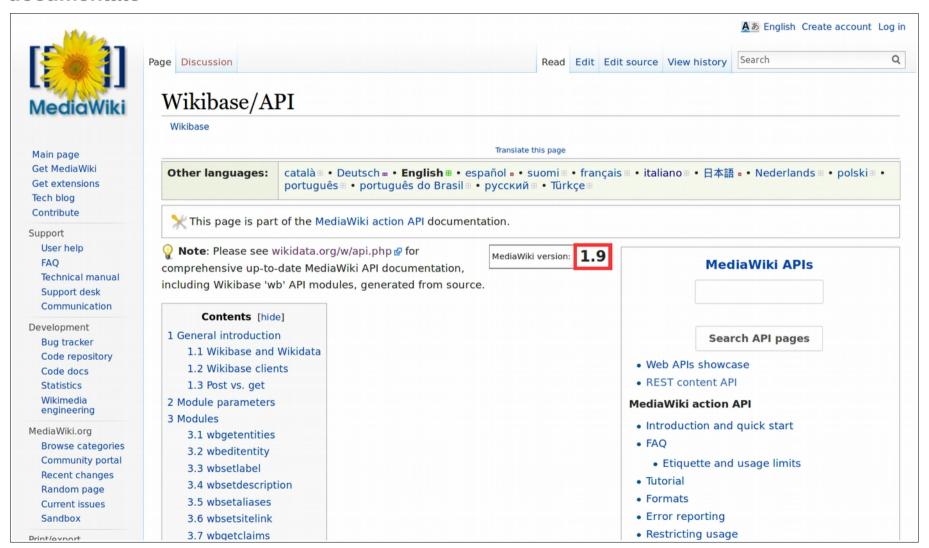
L'esposizione dei dati in RDF standard, utilizzando Wikibase si ottiene for-free

```
wdata:0156 a schema:Dataset :
    schema:about wd:0156 :
    cc:license <http://creativecommons.org/publicdomain/zero/1.0/> ;
    schema:softwareVersion "0.0.1" :
    schema:version "2480"^^xsd:integer ;
    schema:dateModified "2016-05-24T18:01:41Z"^^xsd:dateTime .
wd:0156 a wikibase:Item :
    rdfs:label "Lo Stato innovatore"@en :
    skos:prefLabel "Lo Stato innovatore"@en :
    schema:name "Lo Stato innovatore"@en;
    rdfs:label "Lo Stato innovatore"@it;
    skos:prefLabel "Lo Stato innovatore"@it :
    schema:name "Lo Stato innovatore"@it ;
    skos:altLabel "CFI0893220"@en.
        "IT\\ICCU\\CFI\\0893220"@en,
        "CFI0893220"@it,
        "IT\\ICCU\\CFI\\0893220"@it;
   wdt:P644 "IT - 2014-2498" :
   wdt:P779 "Roma ; Bari : GLF editori Laterza, 2014" ;
   wdt:P657 "20140514d2014
                             ||||0itac50
   wdt:P923 "IT - IT-FI0098 - 20140514" ;
   wdt:P754 "Lo Stato innovatore : sfatare il mito del pubblico contro il privato / Mariana Mazzucato ; traduzione di Fabio Galimberti" ;
   wdt:P793 "Anticorpi ; 43" ;
   wdt:P620 "9788858111703" ;
   wdt:P618 "CFI0893220";
   wdt:P673 "it";
   wdt:P619 "20160201124421.0";
   wdt:P671 "ita" ;
   wdt:P788 "XXVI, 351 p.; 21 cm";
   wdt:P1195 wd:Q152 ;
   wdt:P1193 wd:Q155 ;
   wdt:P1218 wd:0157;
   wdt:P1224 wd:0159;
   wdt:P1226 wd:Q160 ;
   wdt:P1228 wd:Q161;
```

Wikibase può gestire molteplici modelli descrittivi nel medesimo Item: ogni elemento è qualificato



Wikibase può gestire molteplici flussi di modifica: il *protocollo* si basa su API pubbliche e documentate



Perché una sperimentazione su larga scala potrebbe essere interessante

Questi i punti da considerare

- le potenzialità di collaborazione del modello Wiki;
- organizzazione dei dati pronta per RDF/Linked data;
- organizzazione dei dati "multistandard" (fonte e qualificatori);
- capacità nativa di trattare i dati in ambiente multilingua;
- forte integrazione con il web;
- disponibilità dei dati (in lettura e in scrittura) con protocolli aperti (SPARQL, API Wiki*, ecc);
- disponibilità di software e di programmi di utilità aperti con una larga base di sviluppo/sperimentazione a livello mondiale.