

Data coherence between OSM and Wikipedia

Cristian Consonni
Fondazione Bruno Kessler



State of the Map 2013 - Birmingham
September 2013

Outline

- 1 Introduction
- 2 The Problem
- 3 Proposing a Solution
 - Wikipedia-OSM comparator
 - Nut4Nuts
- 4 Conclusions
- 5 Questions

Collecting Information About the Real World



WIKIPEDIA
L'enciclopedia libera

- Pagina principale
- Ultime modifiche
- Una voce a caso
- Vetrina
- Aiuto
- Comunità
 - Portale Comunità
 - Bar
 - Il Wikipediano
 - Fai una donazione
 - Contatti
- Stampa/esporta
- Strumenti
 - Puntano qui
 - Modifiche correlate
 - Carica un file
 - Carica su Commons
 - Pagine speciali
 - Link permanente
 - Informazioni sulla pagina
 - Elemento dati

CristianCantoro [discussioni](#) [preferenze](#) [osservati speciali](#) [contribuì](#) [esci](#)

Voce [Discussione](#) [Leggi](#) [Modifica](#) [Modifica sorgente](#) [Cronologia](#) [Ricerca](#)

Palazzo Vecchio

Da Wikipedia, l'enciclopedia libera.
Wikipedia: Palazzo Vecchio (0271928), museo italiano
Altre: Palazzo della Signoria, Palazzo Vecchio di Firenze, Museo di Palazzo Vecchio

Coordinate: 43°7′53″N 11°25′17″E﻿ / ﻿43.798315°N 11.256174°E﻿ / 43.798315; 11.256174 (Mappa)

Disambiguazione – Se stai cercando altri significati, vedi **Palazzo Vecchio (disambigua)**.

Palazzo Vecchio si trova in piazza della Signoria a Firenze ed è la sede del comune della città. Rappresenta la migliore sintesi dell'*architettura civile trecentesca* cittadina ed uno dei palazzi civici più conosciuti d'Italia.

Chiamato in origine "Palazzo dei Priori" o "Palagio Novo", divenne nel XV secolo "Palazzo della Signoria", dal nome dell'organismo principale della *Repubblica fiorentina*; nel 1540 divenne Palazzo Ducale, quando il duca Cosimo I de' Medici ne fece la sua residenza; infine il nome Vecchio lo assunse nel 1565 quando la corte del Duca Cosimo si spostò nel "nuovo" Palazzo Pitti. Dal 1865 al 1871 fu sede del *Parlamento italiano*, mentre oggi ospita il Sindaco di Firenze e vari uffici comunali. Vi si trova inoltre un museo, che permette di visitare le magnifiche sale dove lavorarono, tra gli altri, Agnolo Bronzino, Ghirlandaio, Giorgio Vasari, e dove sono esposte opere di Michelangelo Buonarroti, Donatello, Verrocchio.

L'edificio si è gradualmente ingrandito verso est, arrivando ad occupare un isolato intero e allungando l'iniziale parallelepipedo trecentesco fino a quadruplicarne le dimensioni, con una pianta che ricorda un trapezoido del quale la facciata è solo il lato più corto. Sulla facciata principale a bugnato, la Torre di Arnolfo è uno degli emblemi della città.

Indice [nascondi]

- 1 Storia
 - 1.1 Preesistenze
 - 1.2 Il nuovo palazzo dei Priori
 - 1.3 La residenza del Duca
 - 1.4 Storia contemporanea
- 2 L'esterno
 - 2.1 Arengario ed entrata
 - 2.2 Gli stemmi sulla facciata


Palazzo Vecchio - Quartieri monumentali



Palazzo Vecchio

Tipo Arte
Indirizzo Piazza della Signoria, Firenze
Sito www.museivicinofirentini.it

Collecting Information About the Real World



WIKIPEDIA
L'enciclopedia libera

Pagina principale
Ultime modifiche
Una voce a caso
Vetrina
Aiuto


Comunità

- Portale Comunità
- Bar
- Il Wikipediano
- Fai una donazione
- Contatti

Stampa/esporta

Strumenti

- Puntano qui
- Modifiche correlate
- Carica un file
- Carica su Commons
- Pagine speciali
- Link permanente
- Informazioni sulla pagina
- Elemento dati



OpenStreetMap
The Free Wiki World Map

Search

examples: "Adnan", "Regent Street, Cambridge", "CB2 5AQ", or "post offices near Lüben"
more examples... Where am I?

Help

- Help Centre
- Documentation


Community

- Community Blogs
- Foundation
- User Diaries

Data

- Copyright & License
- Export Data
- GPS Traces

State of the Map 2013
6-8 September
Birmingham, UK

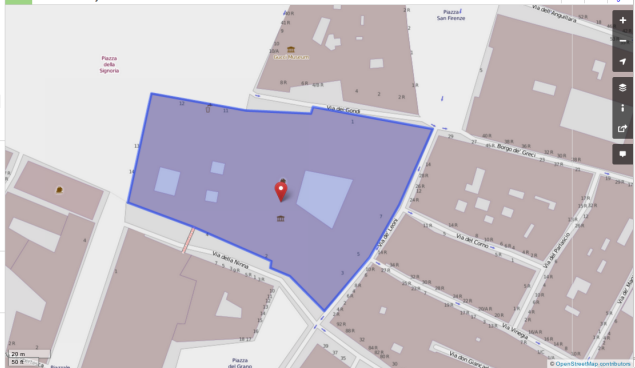


Make a Donation

Hosting is supported by the
USC, Vrije Universiteit, Imperial
College London and Syberwerk
Hamburg, and other partners.

- 1.3 La residenza del Duca
- 1.4 Storia contemporanea
- 2 L'esterno
 - 2.1 Arengio ed entrata
 - 2.2 Gli stemmi sulla facciata

View Edit History



CristianCantero home inbox logout

Tipo: Arte
Indirizzo: Piazza della Signoria, Firenze
Sito: www.museivicinofirenze.it

Collecting Information About the Real World

Wikipedia and OpenStreetMap are:

- collaborative
- volunteer-driven
- free (as in freedom *and* as in beer)

Both projects collect information about the **real world**.

Different Processes and Communities



Wikipedia



OpenStreetMap

- anonymous users can edit
- entries consist in *text* (or media)
- only encyclopedical subjects
- content can be protected from editing in case of problems

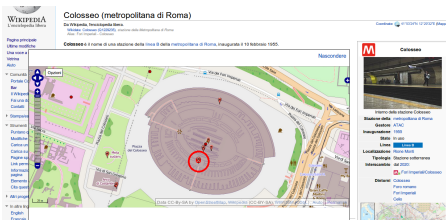
- only registered users can edit
- entries consist in *data*
- everything can be described
- content is always editable

Inconsistencies in the data

Data in Wikipedia can be **inconsistent** with data from OpenStreetMap.
We should *compare* the data and *reconcile* the differences.

Inconsistencies in the data

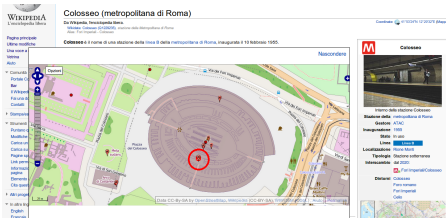
Data in Wikipedia can be **inconsistent** with data from OpenStreetMap. We should *compare* the data and *reconcile* the differences.



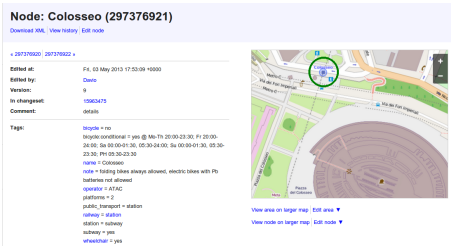
On Wikipedia the metro station “Colosseum” is **inside** the Colosseum itself.

Inconsistencies in the data

Data in Wikipedia can be **inconsistent** with data from OpenStreetMap. We should **compare** the data and **reconcile** the differences.



On Wikipedia the metro station “Colosseum” is **inside** the Colosseum itself.



On OpenStreetMap the metro station is **correctly placed outside** the monument.

OpenStreetMap maps on Wikipedia provided by **WIWOSM** tool by User:Master and User:Kolossos, check it out on:

<http://wiki.openstreetmap.org/wiki/WIWOSM>

Proposal of the Solution

Two steps towards a solution:

- 1 **Compare** the data
 - Identify **links** between Wikipedia pages and OSM entities
 - Extract all the available **geographical information**
 - Define **metrics** to calculate if the data are “close” or not
- 2 **Reconcile** the differences
 - Provide the **communities** with the result of previous analysis
 - Creating **tools** to facilitate the reconciliation

Comparing the data

Wikipedia-OpenStreetMap comparator

Proof-of-concept: comparing data about churches in Italy:

Wikipedia-OpenStreetMap comparator

source code: <https://github.com/CristianCantoro/WOcomparator>

Easy case:

- pre-defined category of items (selection on a set of features in OSM, articles with a given template in Wikipedia)
 - only entities with a (it:)Wikipedia attribute were selected
- ⇒ linking is straightforward.

Comparing the data

Wikipedia-OpenStreetMap comparator

OSM2Wiki [modifica | modifica sorgente]



Pagine di Wikipedia con coordinate [modifica | modifica sorgente]

pagename	osm_id	osm_type	osm_lon	osm_lat	wiki_lon	wiki_lat	distance	Intersects
Chiesa di San Marcella	138843630 ↗	polygon	12.3286704089	45.4429166423	12.328654	45.442754	18.1216659172	False
Chiesa di Sant'Antonio in Boccadasse	44116050 ↗	polygon	8.97334128599	44.3897596801	8.97339722222	44.3899222222	18.6034425236	False
Chiesa del Redentore	154856677 ↗	polygon	12.3323886161	45.4249690455	12.332379	45.424901	7.59993555287	True
Chiesa del Santissimo Crocifisso (San Miniato)	166060331 ↗	polygon	10.8523050828	43.6799885906	10.8523333333	43.6799722222	2.9149147783	True
Chiesa di San Francesco della Vigna	138804531 ↗	polygon	12.3479744656	45.4382162417	12.348638	45.438153	52.3909044727	False
Basilica di San Paolo Maggiore	-1686645 ↗	polygon	14.2569631374	40.8513077357	14.25683	40.85144	18.487236244	True
Chiesa di Santa Caterina (Livorno)	121984159 ↗	polygon	10.3082178851	43.5548635531	10.308287	43.554955	11.5937993656	True
Chiesa di Santa Maria presso San Satiro	45671470 ↗	polygon	9.18778642155	45.4627111186	9.18803333333	45.4626472222	20.5752721195	False
Cattedrale di Santa Maria Assunta e di San Genesio	166057438 ↗	polygon	10.8517038323	43.6796514892	10.8515305556	43.6796333333	14.1172904303	True
Battistero di San Giovanni (Volterra)	162389647 ↗	polygon	10.8583594044	43.4016133783	10.8582416667	43.4017138889	14.6855133854	False
Pieve di San Frediano a Montignoso	164912571 ↗	polygon	10.9009633095	43.495126619	10.9009694444	43.4951388889	1.45071714831	True
Basilica dei Santi Ambrogio e Carlo al Corso	25619192 ↗	polygon	12.4779597993	41.9053063237	12.478075	41.9053777778	12.4239682558	True
Basilica di San Clemente al Laterano	-326080 ↗	polygon	12.4975675257	41.8893186131	12.4975	41.8894444444	15.0579654109	True
Chiesa di San Zeno (Pisa)	137827225 ↗	polygon	10.4078238562	43.7231268785	10.4075305556	43.723	27.5186786949	False
Basilica di San Crisogono	139900092 ↗	polygon	12.4735007941	41.8891086638	12.47379	41.889178	25.207695433	True
Chiesa di San Matteo (Livorno)	153593152 ↗	polygon	10.3274093535	43.560885571	10.3269388889	43.5610166667	40.7063344519	False
Chiesa di San Matteo (Livorno)	153593152 ↗	polygon	10.3274093535	43.560885571	10.3269388889	43.5610166667	40.7063344519	False

<http://it.wikipedia.org/wiki/Utente:CristianCantoro/Georeferenziazione>

Comparing the data

nuts4nuts

For the hard case (try to link every possible thing), another tool:

Nuts4Nuts

source code: <https://github.com/SpazioDati/Nuts4Nuts>

```
{
  "q0": {
    "result": [
      {
        "score": 1,
        "type": [
          {
            "id": "/LAU2",
            "name": "Basic service"
          },
          {
            "id": "http://data.spaziodati.eu/resource/1b3843bb1cdfc5b342fefa9f76aa5caf0eb2ddb3",
            "match": true,
            "name": "Firenze"
          }
        ]
      }
    ]
  }
}
```

<http://nuts4nutsrecon.spaziodati.eu/reconcile?queries={%22q0%22:%20{%22query%22:%20%22Palazzo%20Vecchio%22%20}}>

Known limitations:

- limited to Italy
- uses of external services

grab the source code: <https://github.com/SpazioDati/Nuts4Nuts>

Dandelion

Nuts4Nuts is built using the infrastructure provided by

Dandelion (<http://dandelion.eu>)

a datamarket by SpazioDati srl.

The screenshot shows the Dandelion website homepage. The browser address bar displays <https://dandelion.eu>. The website has a dark navigation bar with the logo "dandelion private beta" and links for "Data & Services", "How it works", "API Docs", "FAQ", and "Logout". The main content area features the Dandelion logo, a tagline "The one-stop shop for smart, high-quality data. Geo & Linked Data from trusted sources. Public and private.", and a call to action "Are you a data provider? write us!". Below this, there are four featured services: "Eurogeo" (described as an easy-to-use API for administrative & geo relations), "Places (Italy)" (described as tens of thousands of places in Italy from OpenStreetmap, Wikipedia, Geonames, etc.), "DataTXT entity extraction" (described as an API to extract & link entities from texts), and "Coming soon" (described as starting with local data about Italy). Each service is accompanied by a circular icon: a globe for Eurogeo, a location pin for Places, a document for DataTXT, and a heart for Coming soon. A vertical "Feedback" button is on the left side.

Future Work

Nuts4nuts is a step to find geographical information for Wikipedia article that have no explicit coordinates in them.

Future work:

- study new approaches to link entities between Wikipedia and OpenStreetMap
- an application to fix inconsistencies or fill in missing data, like this:

Way: Chiesa di San Matteo (153593152)

[Download XML](#) [View history](#) [Edit way](#)

it.wiki: Chiesa di San Matteo (Livorno)

[View page](#)

Edited at: Sun, 11 Mar 2012 10:50:39 +0000
Edited by: [FabC](#)
Version: 2
In changeset: 10941272
Comment: Livorno - Added/updated monuments and links to Wikipedia

Tags: amenity = place_of_worship
building = church
denomination = catholic
name = Chiesa di San Matteo
religion = christian
wikipedia = [it:Chiesa di San Matteo \(Livorno\)](#)



Conclusions

- Wikipedia and OSM collect information about the **real world**

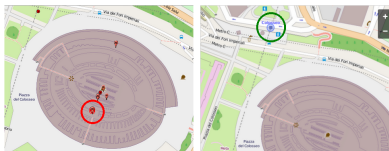


Conclusions

- Wikipedia and OSM collect information about the **real world**



- Comparing data among the two project can highlight **inconsistencies**

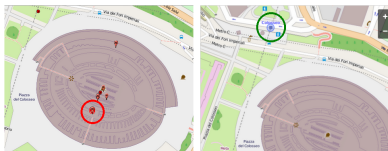


Conclusions

- Wikipedia and OSM collect information about the **real world**



- Comparing data among the two project can highlight **inconsistencies**



- We should **fix** them



Questions?

mail: consonni@fbk.eu

twitter: [@CristianCantoro](https://twitter.com/CristianCantoro)

github: <https://github.com/CristianCantoro>

Thank you

Thank you!

This work was supported by:

T2DataExchange

A project by:

- *SpazioDati srl*
- *Edizioni Curcu & Genovese*

with funds from the *European Regional Development Fund*.

More information: <http://trentino.dandelion.eu>

The following presentation is released under the licence [CC3.0-BY-SA](#).



Further info:

- <http://creativecommons.org/licenses/by-sa/3.0/>

Logos and trademarks are of the respective owners.