

# Data coherence between OSM and Wikipedia

Cristian Consonni  
Fondazione Bruno Kessler



State of the Map 2013 - Birmingham  
September 2013

# Outline

# 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 contributi esci

Voce **Discussione**

Leggi

Modifica

Modifica sorgente

Cronologia



Ricerca



## Palazzo Vecchio

Da Wikipedia, l'enciclopedia libera.

Wikidata: [Palazzo Vecchio](#) (Q271928)

Altre: [Palazzo della Signoria](#), [Palazzo Vecchio di Firenze](#), [Museo di Palazzo Vecchio](#)

Coordinate: 43.769315°N 11.256174°E﻿ / ﻿(Mappa)

**Dissambiguazione** – 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](http://www.museivicinofirentini.it)

# Collecting Information About the Real World

**View** Edit History CristianCarloro home inbox logout

**OpenStreetMap**  
The Free Wiki World Map

Search

Example: "Akshar", "Regent Street, Cambridge", "CB2 5AQ", or "post offices near Lünen"  
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, V&I Centre, Imperial  
College London and Bytemark  
Hosting, and other partners.

1.3 **La residenza del DUCS**

1.4 Storia contemporanea

2 **L'esterno**

2.1 Arengario ed entrata

2.2 Gli stemmi sulla facciata

**Tipo** Arte

**Indirizzo** Piazza della Signoria, Firenze

**Sito** [www.museivicinotrentini.it](http://www.museivicinotrentini.it)

© OpenStreetMap contributors

# 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

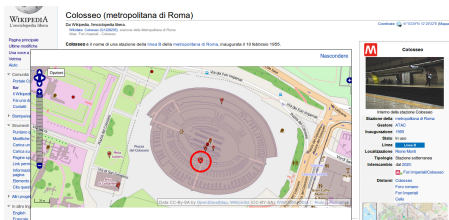
- only registered users can edit
- entries consist in *data*
- everything can be described
- content is always editable

# 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.

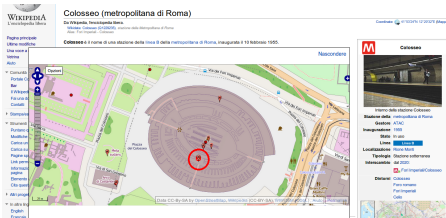


On Wikipedia the metro station “Colosseum” is **inside** the Colosseum itself.

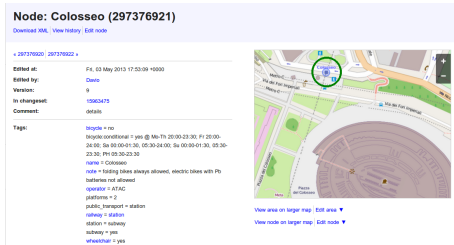


# Inconsistencies in the data

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

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}}>

**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. At the top, there is a navigation bar with the Dandelion logo and a network icon, and a menu with links for 'Data & Services', 'How it works', 'API Docs', 'FAQ', and 'Logout'. The main content area features the Dandelion logo and the tagline: 'The one-stop shop for smart, high-quality data. Geo & Linked Data from trusted sources. Public and private.' Below this is a call to action: 'Are you a data provider?' with a 'write us!' button. The page is divided into two main sections: 'Freshly baked data, ready to eat' and '... and for dessert'. The first section highlights 'Eurogeo' (with a globe icon) and 'Places (Italy)' (with a location pin icon). The second section highlights 'DataTXT entity extraction' (with a document icon) and 'Coming soon' (with a heart icon). A vertical 'Feedback' button is visible on the left side of the page.

https://dandelion.eu

dandelion private beta

Data & Services - How it works API Docs FAQ Logout

dandelion

The one-stop shop for smart, high-quality data.  
Geo & Linked Data from trusted sources. Public and private.

Are you a data provider? write us!

Feedback

Freshly baked data, ready to eat

**Eurogeo**

A easy-to-use API with administrative & geo relations of countries, cities & regions in Europe from trusted, official sources.

**Places (Italy)**

Tens of thousand of places in Italy from OpenStreetmap, Wikipedia, Geonames and our partners.

... and for dessert

**DataTXT entity extraction**

An API to extract & link entities (like persons, places, organizations) from texts in English and

**Coming soon**

We're starting with local data about Italy, but we'll soon extend on other verticals and

# 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**



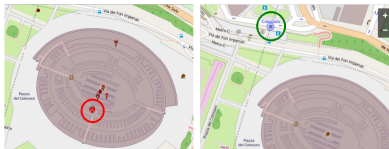


# Conclusions

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



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

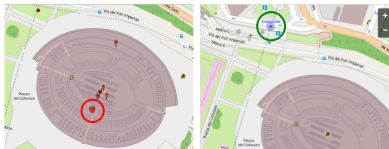


# Conclusions

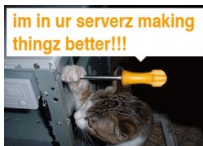
- Wikipedia and OSM collect information about the **real world**



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



- We should **fix** them



## Questions?

mail: [consonni@fbk.eu](mailto:consonni@fbk.eu)

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

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

# 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](https://creativecommons.org/licenses/by-sa/3.0/).



Further info:

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

Logos and trademarks are of the respective owners.