



Open Data

Casi studio e flusso di lavoro

Andrea Zedda

28 febbraio 2017

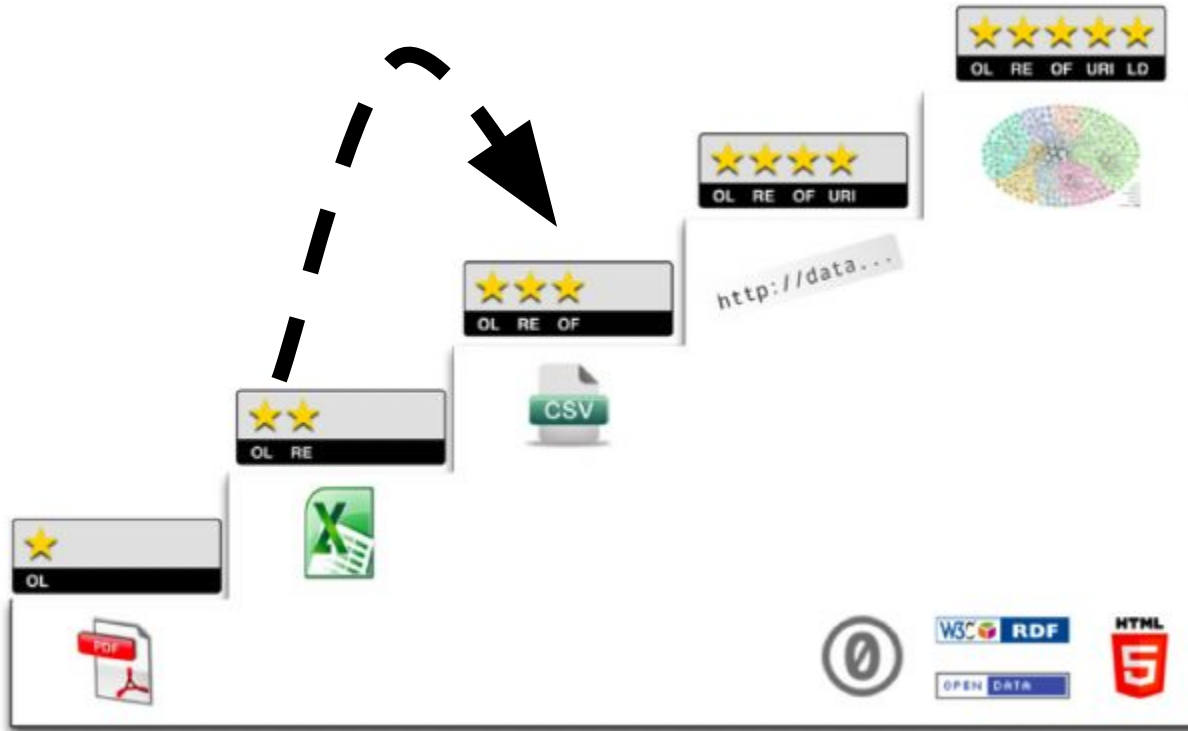


Il seminario «La governance nazionale e regionale dei dati aperti» rientra tra le attività del progetto "OpenRAS: dati aperti per la trasparenza e l'accountability" - Programma Operativo Regionale FSE 2014-2020 (Decisione C 2014 N° 10096 del 17/12/2014), Asse 4 – Capacità istituzionale e amministrativa, azione 11.1.1 "Interventi mirati allo sviluppo delle competenze per assicurare qualità, accessibilità, fruibilità, rilascio, riutilizzabilità dei dati pubblici".

SARDINIA OPEN DATA

- ❑ **Sostegno gratuito alle associazioni e ONG in elaborazione dati**
- ❑ **Formazione volontari nell'ambito IT**
- ❑ **Sensibilizzazione tematiche Open Data a istituzioni, aziende, associazioni e la società civile**
- ❑ **Sviluppo di idee, progetti e applicazioni in ottica Open Data**

OPENINPS



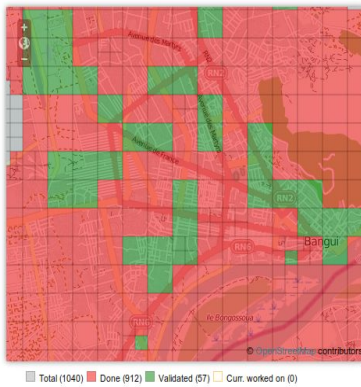
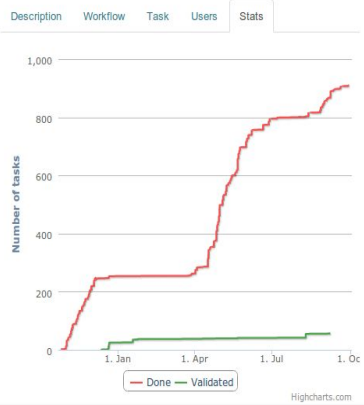
MAPPING MEETINGS



MAPPING UMANITARIO

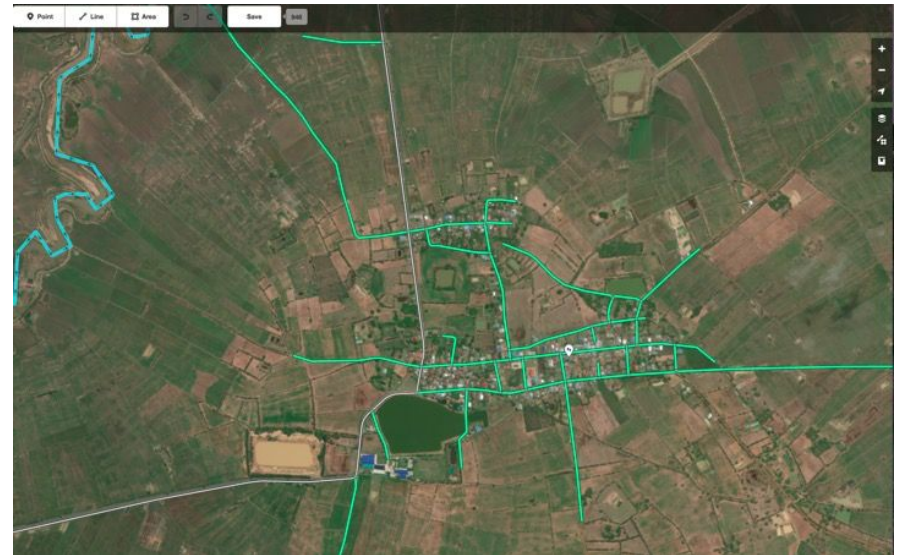
OSM Tasking Manager

You are Andria Tzedda

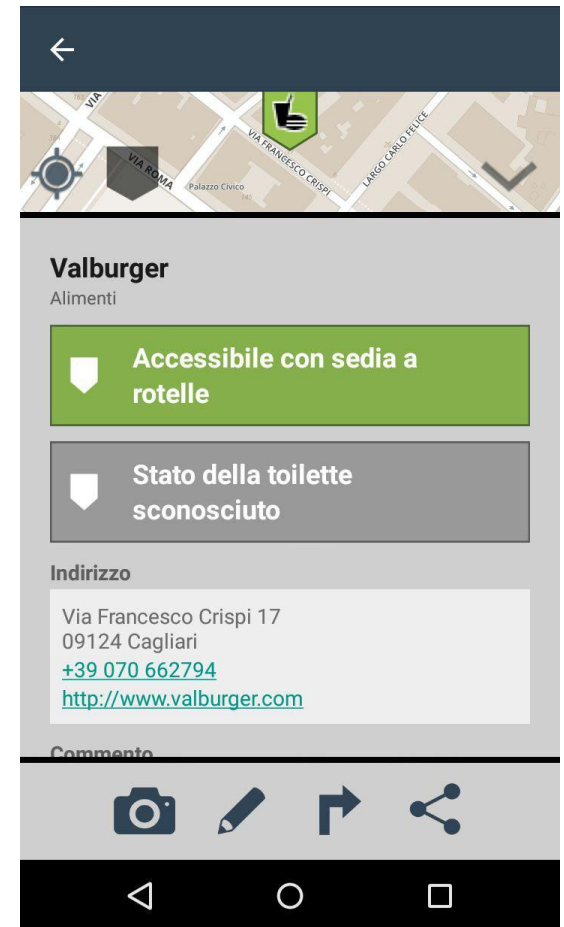
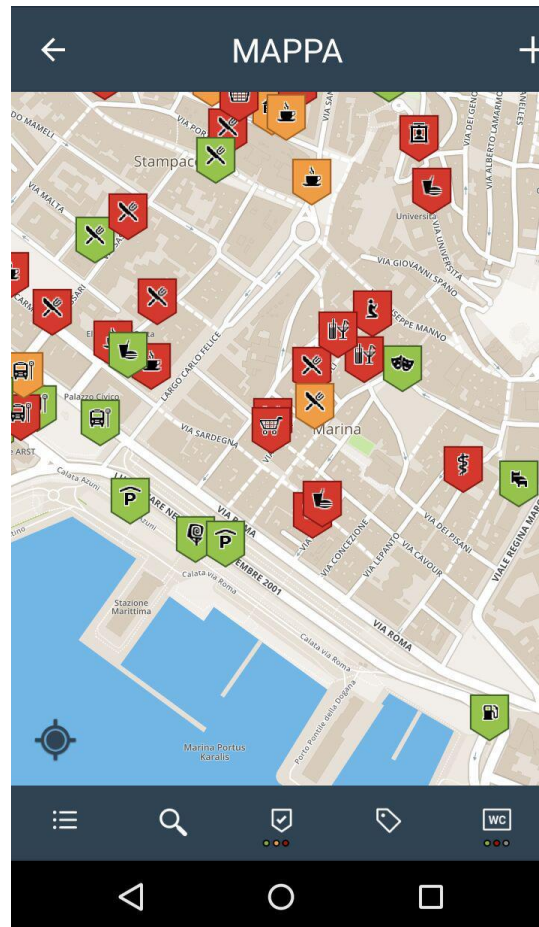
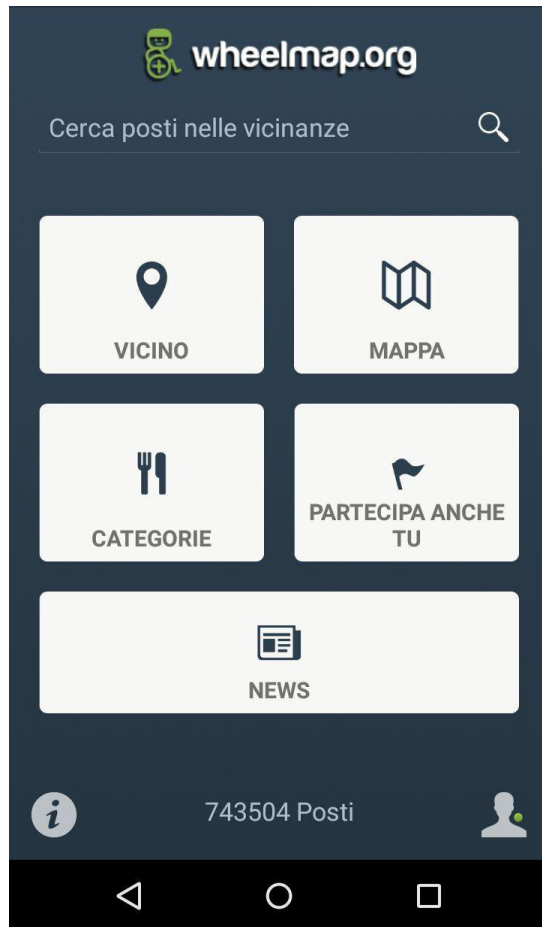


Designed and built for the Humanitarian OpenStreetMap Team with initial sponsorship from the Australia-Indonesia Facility for Disaster Reduction. See [this github page](#) for complete information.

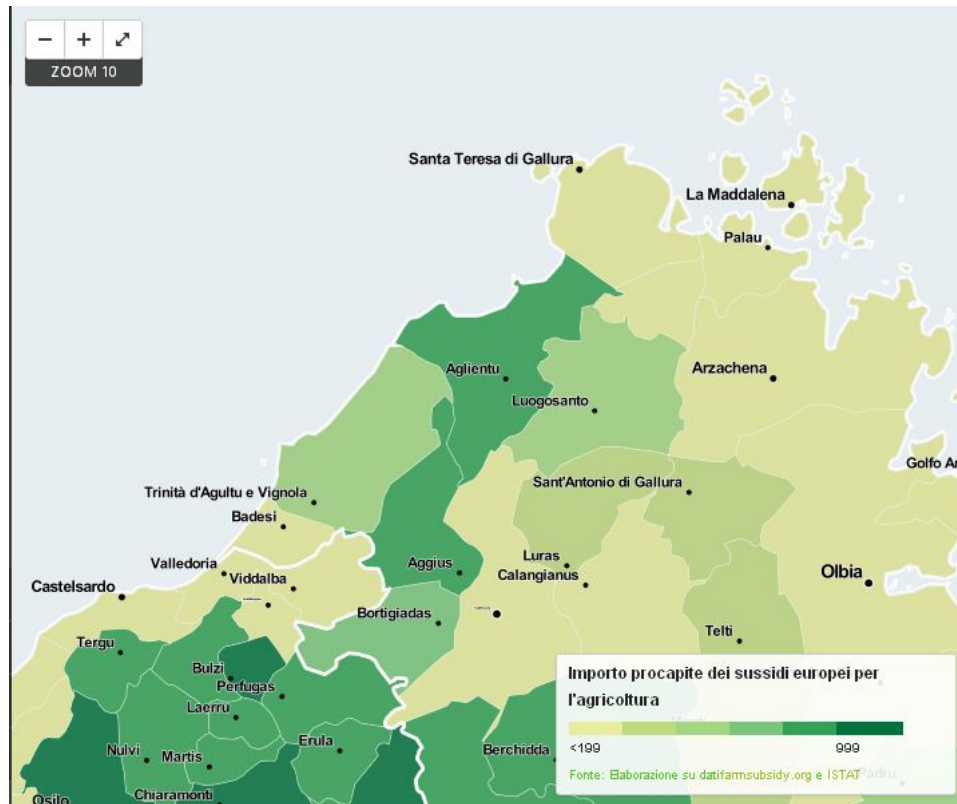
Fork the code on [github](#).



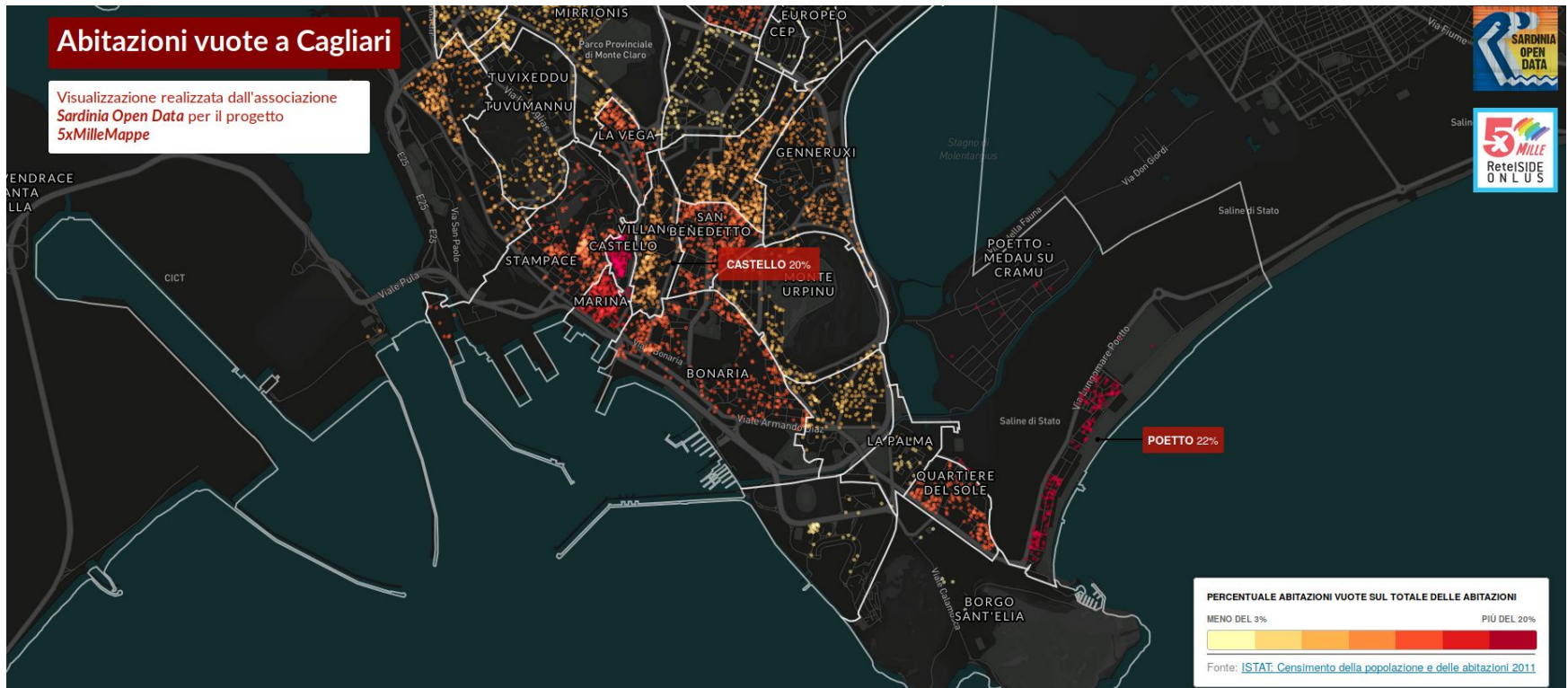
MAPPA LUOGHI ACCESSIBILI



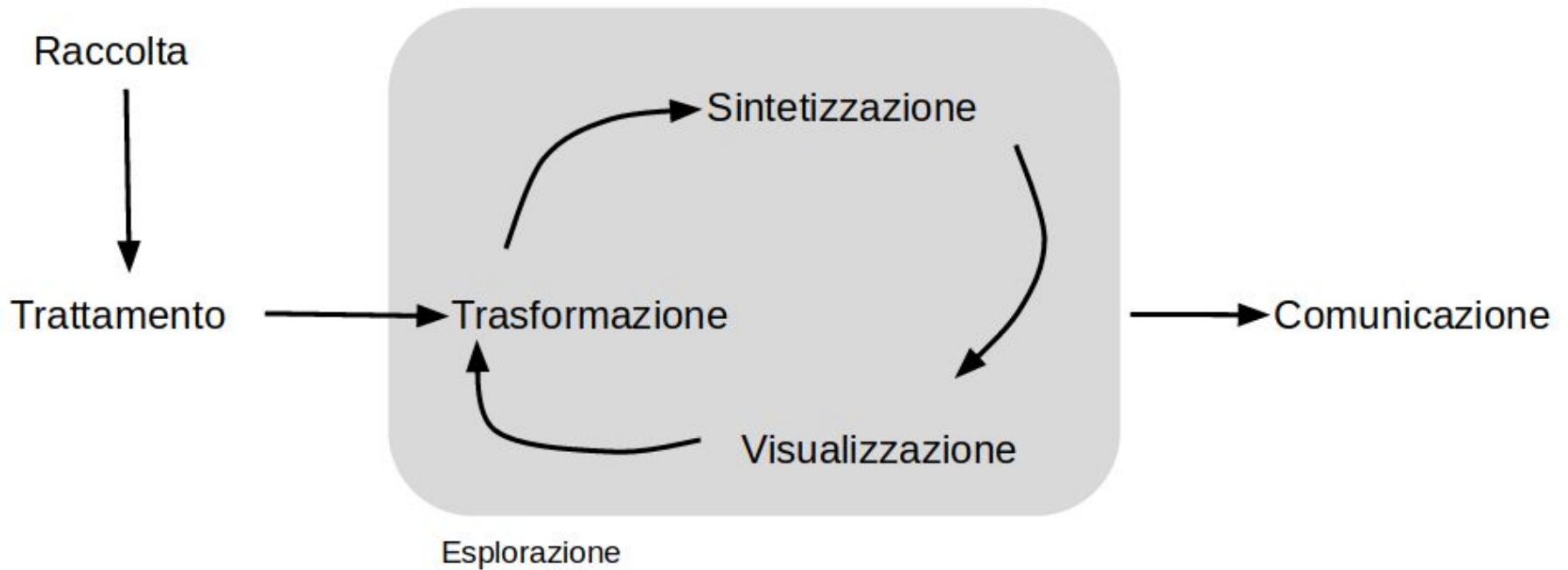
MAPPA INTERATTIVA SUSSIDI AGRICOLTURA



MAPPA SFITTI DI CAGLIARI



WORKFLOW DATA SCIENCE



RACCOLTA: STATISTICA UFFICIALE

Dati provenienti da istituti di statistica riconosciuti e certificati.
In Italia dagli enti appartenenti al circuito **SISTAN** (ISTAT, uffici di statistica nelle amministrazioni, Banca d'Italia...)

PRO

- sono certificati e hanno una alta affidabilità
- indagini a cadenze definite

CONTRO

- son rilasciati in forma aggregata
- spesso son dati non più tanto freschi

RACCOLTA: STATISTICA UFFICIALE



I.Stat versione di prova
il tuo accesso diretto
alla statistica italiana

[Login](#) | [English Version](#) | [FAQs e Contatti](#) | [Manuale utente](#) | [Hor](#)

Ricerca [Per Iniziare](#)

- Previsioni
- Condizioni economiche delle famiglie e disuguaglianze
- Salute e sanità
- Assistenza e previdenza
- Istruzione e formazione
- Cultura, comunicazione, tempo libero, uso del tempo
- Giustizia e sicurezza
- Opinioni dei cittadini e soddisfazione per la vita
- Partecipazione sociale
- Conti nazionali**
 - Conti e aggregati economici nazionali annuali
 - Sequenza dei conti (milioni di euro)**
 - Principali aggregati del Prodotto interno lordo (milioni di euro)
 - Produzione e valore aggiunto per branca di attività (milioni di euro)
 - Redditi da lavoro dipendente e componenti per branca di attività

Sequenza dei conti (milioni di euro)

Personalizza | Esportazioni | Grafici | La tua interrogazione

→ Territorio	Italia													
→ Valutazione	prezzi correnti													
→ Correzione	dati grezzi													
→ Edizione	Mar-2015													
→ Anno	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
→ Tipo														
aggregato														
conto economico delle risorse e degli impieghi	(t)	(t)	(t)	(t)	(t)	(t)	(t)	(t)	(t)	(t)	(t)	(t)	(t)	(t)
risorse	1 424 027.4	1 547 304.2	1 617 334.3	1 665 666.5	1 709 895.2	1 789 102.3	1 859 297.5	1 968 271.7	2 057 541.9	2 085 911.9	1 937 501.5	2 041 438.7	2 106 789.2	2 060 155.0
prodotto interno lordo ai prezzi di mercato	1 172 365.3	1 239 758.8	1 299 411.8	1 346 360.2	1 391 312.8	1 449 016.0	1 490 409.4	1 549 188.0	1 610 304.9	1 632 933.4	1 573 655.1	1 605 694.4	1 638 857.3	1 615 131.2
importazioni di beni (fob) e servizi	251 662.2	307 545.4	317 922.5	319 306.3	318 582.4	340 086.3	368 888.1	419 083.6	447 237.0	452 978.4	363 846.4	435 744.3	467 931.9	445 023.8
importazioni di														

RACCOLTA: OPEN DATA PUBBLICI

Provengono da fonte amministrativa e sono correlati da licenze libere che permettono il riutilizzo e la redistribuzione da parte dell'utente



PRO

- le licenze libere tutelano sia l'utente che l'autore della base dati
- grande varietà
- facilmente scaricabili e accessibili



CONTRO

- non c'è uno standard che regoli la struttura interna dei dataset
- il tasso di rilascio è molto variabile tra le amministrazioni
- manca ancora una cultura radicata



RACCOLTA: OPEN DATA PUBBLICI



dati.gov.it
I dati aperti della PA



Governo italiano
Presidenza del Consiglio dei Ministri
Ministero per la semplificazione e la pubblica amministrazione

[Home](#) | [Dati](#) | [Voglio capire](#) | [Applicazioni](#) | [Condivido dati](#) | [Condivido applicazioni](#) | [Notizie](#)

[Home](#) / Dataset

Organizzazioni [No Filtri](#)

- Comune Albano Laziale (1089)
- Provincia Autonoma ... (696)
- Regione Lombardia (551)
- INPS (537)
- Regione Piemonte (477)
- Comune di Roma (466)
- Comune di Bologna (462)
- Regione Liguria (361)

🔍

8.843 dataset trovati

Ordina per: Rilevanza ▼

Formez Linea Amica
Enciclopedia delle Domande e Risposte 🔥 43 visite

Nella banca dati sono archiviate tutte le Domande e Risposte (D&R) della sezione Cerco le risposte, ovvero le FAQ e che piu di frequente vengono poste al servizio Linea...

CSV
XML

RACCOLTA: USER GENERATED DATA

Informazioni rese disponibili dalle corporation che offrono servizi web-based (social network, motori di ricerca, app, mappe)

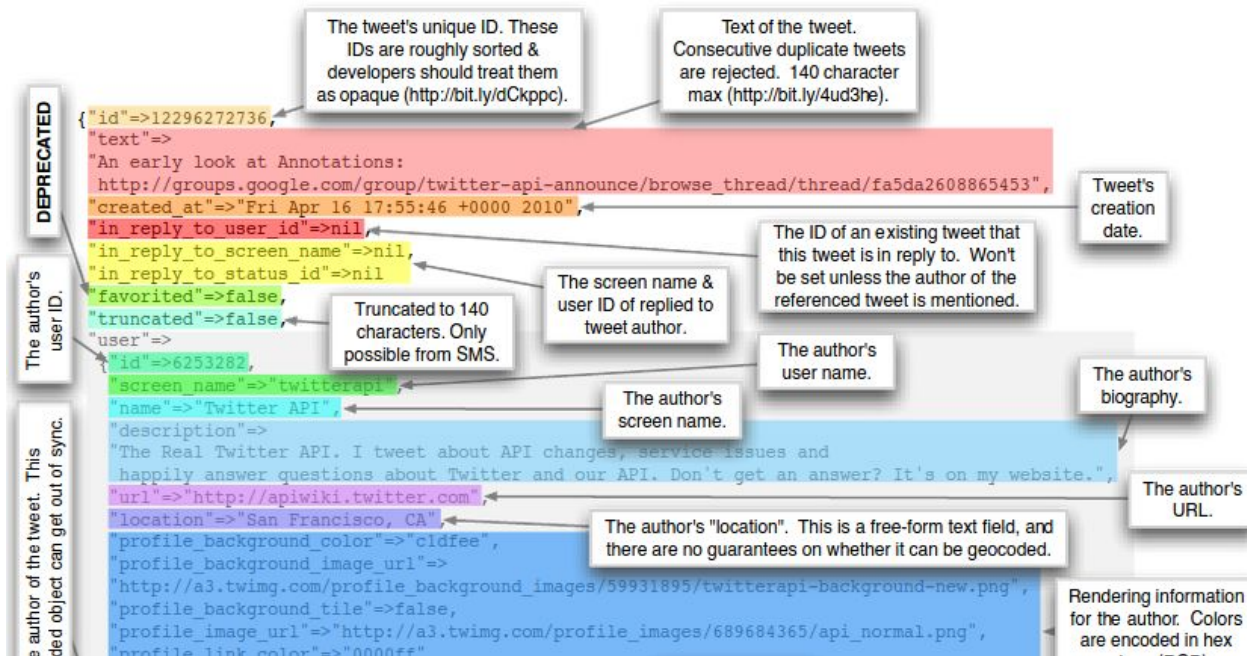
↑ PRO

- enormi quantità di informazioni
- API efficienti e ben documentate
- dettaglio e puntualità

↓ CONTRO

- sono di proprietà dell'azienda che gestisce il servizio
- la maggior parte delle informazioni è a pagamento

RACCOLTA: USER GENERATED DATA



RACCOLTA: COMMUNITY DATA

Dati generati consapevolmente da comunità di utenti attraverso piattaforme appositamente sviluppate (Wikipedia, OpenStreetMap, Ushaidi ..)

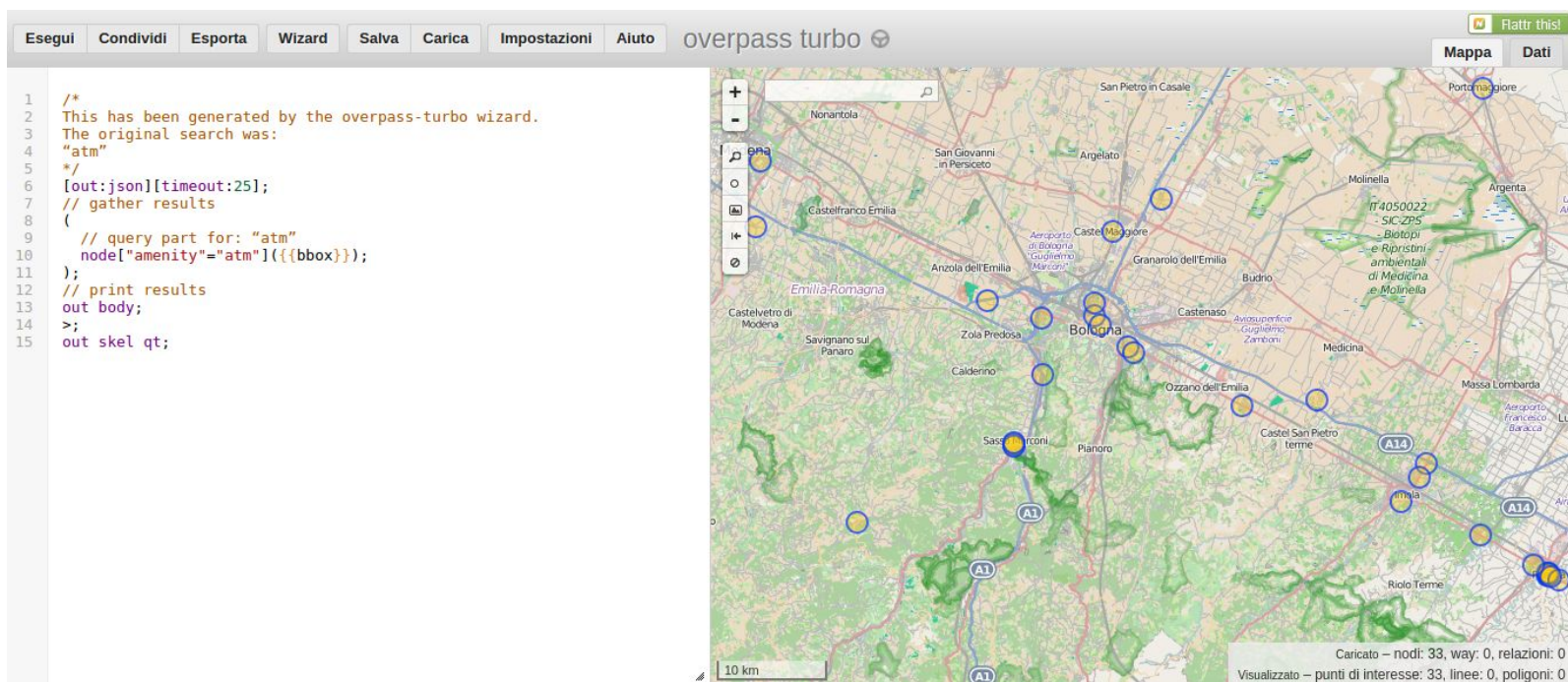
↑ PRO

- enormi quantità di informazioni
- completa e integrale disponibilità
- dettaglio e puntualità
- i dati sono open e vanno attribuiti a chi li ha generati

↓ CONTRO

- la qualità dipende dalla tipologia della comunità dei contributori
- copertura variabile tra i vari temi e territori trattati

RACCOLTA: COMMUNITY DATA



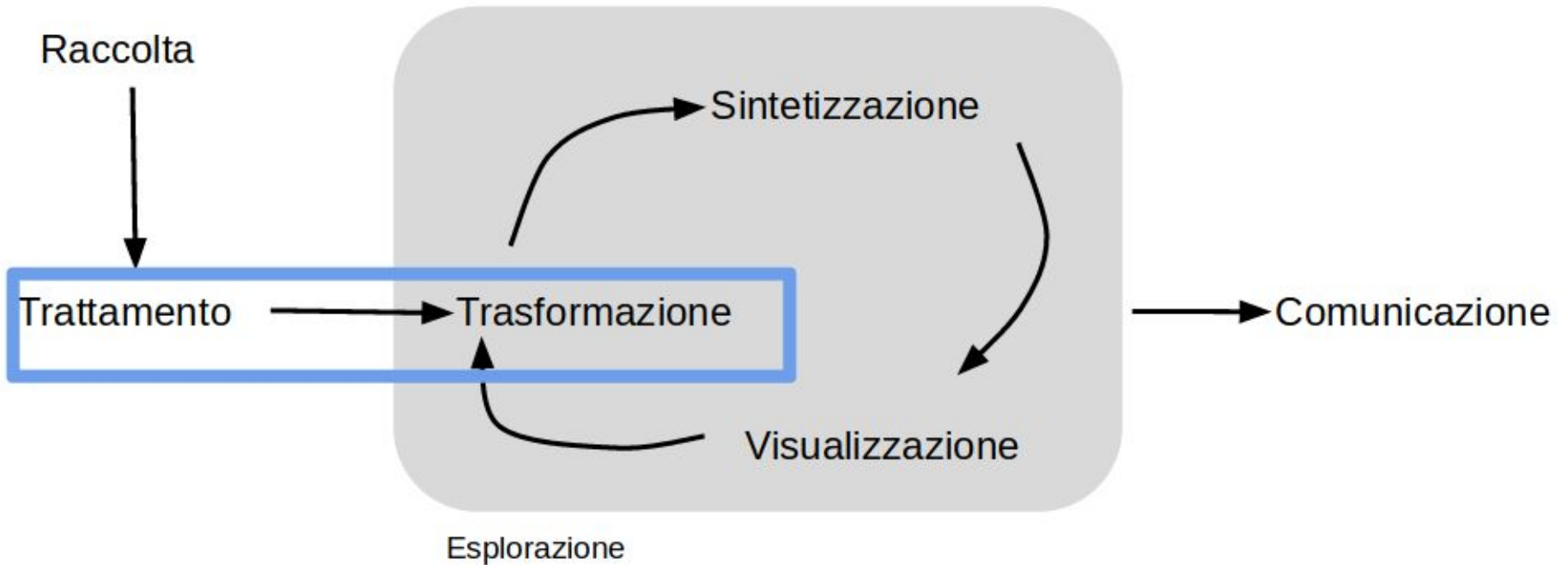
The screenshot displays the OpenStreetMap Overpass Turbo interface. On the left, there is a code editor with a query for ATM locations in the Bologna area. The query is as follows:

```
1 /*  
2 This has been generated by the overpass-turbo wizard.  
3 The original search was:  
4 "atm"  
5 */  
6 [out:json][timeout:25];  
7 // gather results  
8 (  
9 // query part for: "atm"  
10 node["amenity"="atm"]({{bbox}});  
11 );  
12 // print results  
13 out body;  
14 >;  
15 out skel qt;
```


On the right, a map of the Bologna region is shown with several blue circular markers indicating ATM locations. The map includes various geographical features, roads, and place names. In the bottom right corner, the following statistics are displayed:

Caricato – nodi: 33, way: 0, relazioni: 0
Visualizzato – punti di interesse: 33, linee: 0, poligoni: 0

TRATTAMENTO E TRASFORMAZIONE



TRATTAMENTO E TRASFORMAZIONE



**DATI NON PRONTI
PER ESSERE
ELABORATI**



**DATI GIÀ PRONTI
PER ESSERE
ELABORATI**

I DATASET

Sono serie di valori disposti in maniera strutturata cioè in osservazioni e divisi per variabili.

- Una **variabile** è una quantità, oppure una qualità o ancora una proprietà che si possono misurare.
- Un **valore** è la misura della variabile.
- Una **osservazione** è l'insieme di variabili e valori relative ad un oggetto, persona o eventi.

I DATASET

country	year	cases	population
Afghanistan	1999	7745	19987071
Afghanistan	2000	2666	20593360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	127291272
China	2000	213766	128042583

variables

country	year	cases	population
Afghanistan	1999	7745	19987071
Afghanistan	2000	2666	20593360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	127291272
China	2000	213766	128042583

observations

country	year	cases	population
Afghanistan	99	75	987071
Afghanistan	00	66	593360
Brazil	99	737	2006362
Brazil	00	488	504898
China	99	2258	291272
China	00	3766	42583

values

TRATTAMENTO

- ❑ spazi vuoti
- ❑ celle unite
- ❑ errori di digitazione
- ❑ maiuscole/minuscole
- ❑ numeri digitati come testo
- ❑ valori digitati in modo eterogeneo

TRASFORMAZIONE



tidyr::gather(cases, "year", "n", 2:4)
Gather columns into rows.



tidyr::spread(pollution, size, amount)
Spread rows into columns.



tidyr::separate(storms, date, c("y", "m", "d"))
Separate one column into several.



tidyr::unite(data, col, ..., sep)
Unite several columns into one.

L'IMPORTANZA DEGLI STANDARD

- ❑ Codici univoci (ad esempio il codice ISTAT ed EUROSTAT per gli aggregati territoriali)
- ❑ Specificare l'encoding del file
- ❑ Per i dati georiferiti utilizzare un sistema di riferimento delle coordinate diffuso (ad es WGS84)
- ❑ Documentare i dataset (metadata)

L'IMPORTANZA DEGLI STANDARD:

GTFIS shapes.txt

```
shape_id,shape_pt_lat,shape_pt_lon,shape_pt_sequence,shape_dist_traveled  
A_shp,37.61956,-122.48161,1,0  
A_shp,37.64430,-122.41070,2,6.8310  
A_shp,37.65863,-122.30839,3,15.8765
```

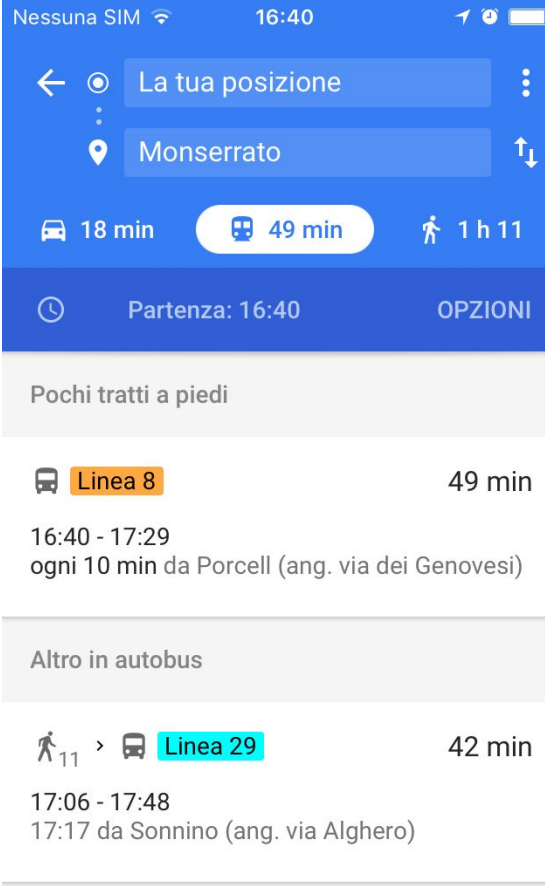
stops.txt

```
stop_id,stop_name,stop_desc,stop_lat,stop_lon,stop_url,location_type,parent_station  
S1,Mission St. & Silver Ave.,The stop is located at the southwest corner of the intersection.,37.  
S2,Mission St. & Cortland Ave.,The stop is located 20 feet south of Mission St.,37.74103,-122.422  
S3,Mission St. & 24th St.,The stop is located at the southwest corner of the intersection.,37.752  
S4,Mission St. & 21st St.,The stop is located at the northwest corner of the intersection.,37.757  
S5,Mission St. & 18th St.,The stop is located 25 feet west of 18th St.,37.761829,-122.419382,,  
S6,Mission St. & 15th St.,The stop is located 10 feet north of Mission St.,37.766629,-122.419782,  
S7,24th St. Mission Station,,37.752240,-122.418450,,S8  
S8,24th St. Mission Station,,37.752240,-122.418450,http://www.bart.gov/stations/stationguide/stat
```

stop_times.txt

```
trip_id,arrival_time,departure_time,stop_id,stop_sequence,pickup_type,drop_off_type  
AWE1,0:06:10,0:06:10,S1,1,0,0  
AWE1,,S2,2,1,3  
AWE1,0:06:20,0:06:30,S3,3,0,0  
AWE1,,S5,4,0,0  
AWE1,0:06:45,0:06:45,S6,5,0,0  
AWD1,0:06:10,0:06:10,S1,1,0,0  
AWD1,,S2,2,0,0  
AWD1,0:06:20,0:06:20,S3,3,0,0  
AWD1,,S4,4,0,0  
AWD1,,S5,5,0,0  
AWD1,0:06:45,0:06:45,S6,6,0,0
```

L'IMPORTANZA DEGLI STANDARD



Nessuna SIM 16:40

← La tua posizione

⋮

📍 Monserrato

🚗 18 min **🚌 49 min** 🚶 1 h 11

🕒 Partenza: 16:40 OPZIONI

Pochi tratti a piedi

🚌 **Linea 8** 49 min

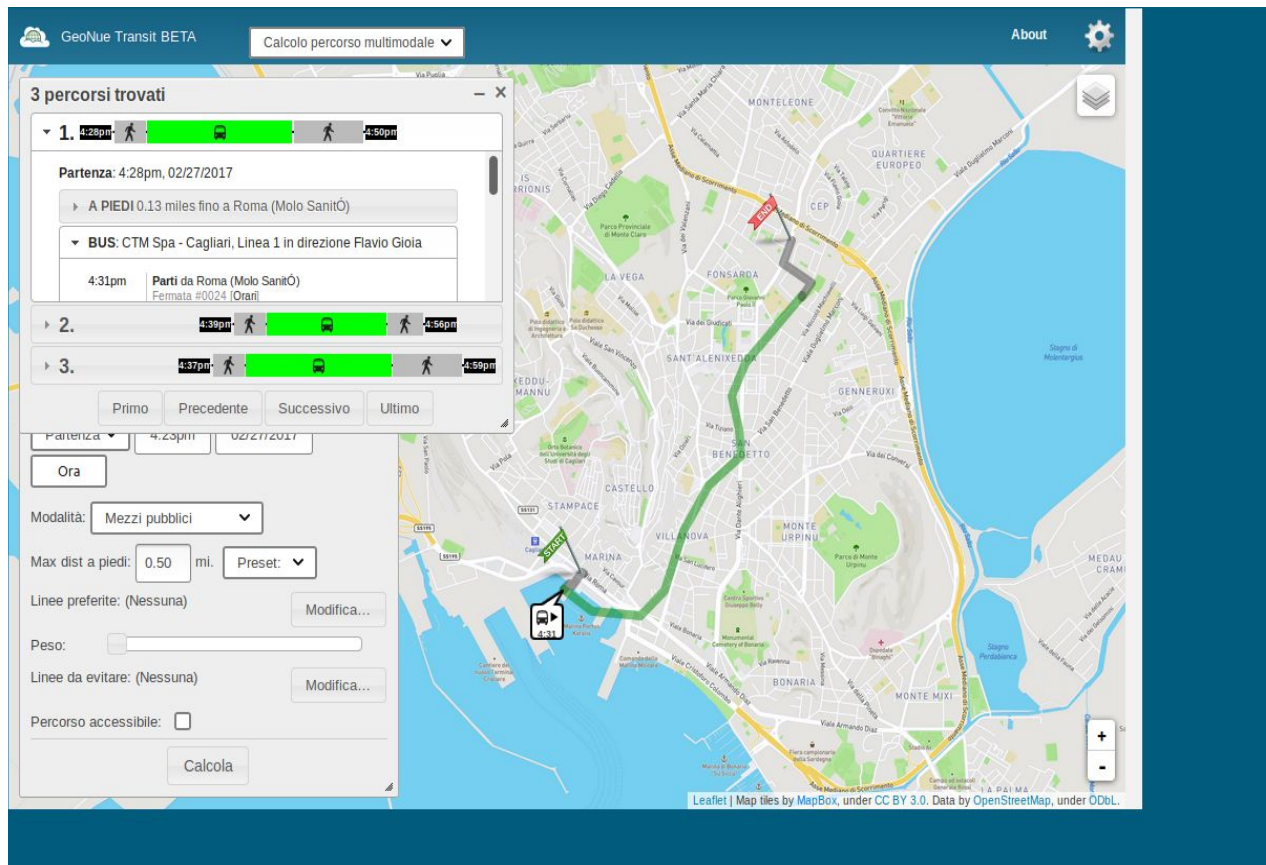
16:40 - 17:29
ogni 10 min da Porcell (ang. via dei Genovesi)

Altro in autobus

🚶₁₁ > 🚌 **Linea 29** 42 min

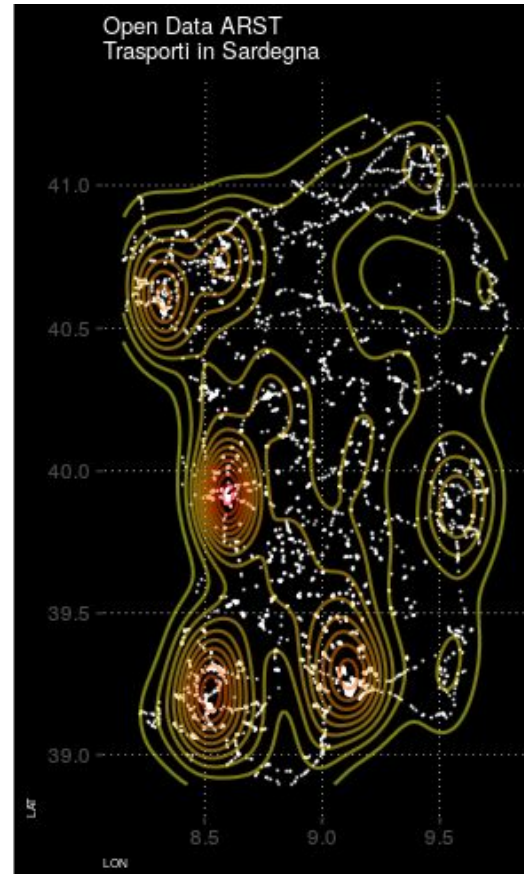
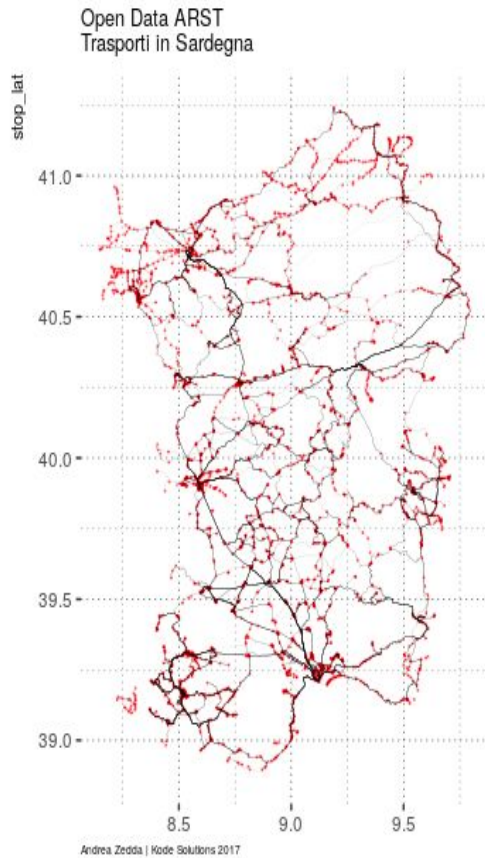
17:06 - 17:48
17:17 da Sonnino (ang. via Alghero)

L'IMPORTANZA DEGLI STANDARD

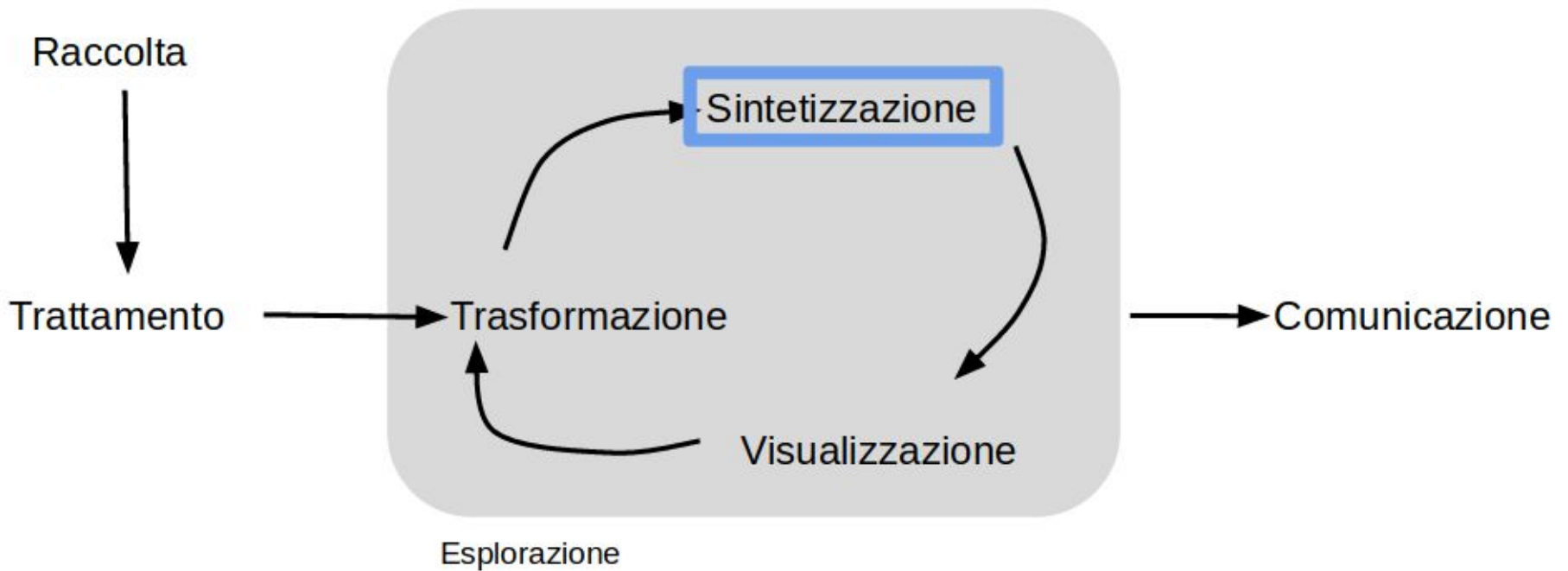


The screenshot displays the GeoNue Transit BETA application interface. At the top, it shows the logo and name 'GeoNue Transit BETA', a dropdown menu for 'Calcolo percorso multimodale', and an 'About' button with a gear icon. The main content area is titled '3 percorsi trovati' and lists three route options. The first route is selected and detailed: it starts at 4:28pm on 02/27/2017, involves a 0.13-mile walk to Roma (Molo Sanitò), and then a bus ride on CTM Spa - Cagliari, Linea 1 towards Flavio Gioia. The bus starts at 4:31pm from the Molo Sanitò stop. The interface includes various filters and settings such as 'Modalità: Mezzi pubblici', 'Max dist a piedi: 0.50 mi', and 'Percorso accessibile'. A map on the right shows the route highlighted in green and yellow over a street map of Cagliari. At the bottom, there is a 'Calcola' button and a footer with attribution: 'Leaflet | Map tiles by MapBox, under CC BY 3.0. Data by OpenStreetMap, under ODbL.'

L'IMPORTANZA DEGLI STANDARD



LA SINTESI DEI DATI





LA SINTESI DEI DATI: STATISTICA DESCRITTIVA

Si dispongono le informazioni in distribuzioni e le sintetizza descrittivamente attraverso tipologie di indici (indici di posizione, indici di forma, rapporti statistici ecc..)

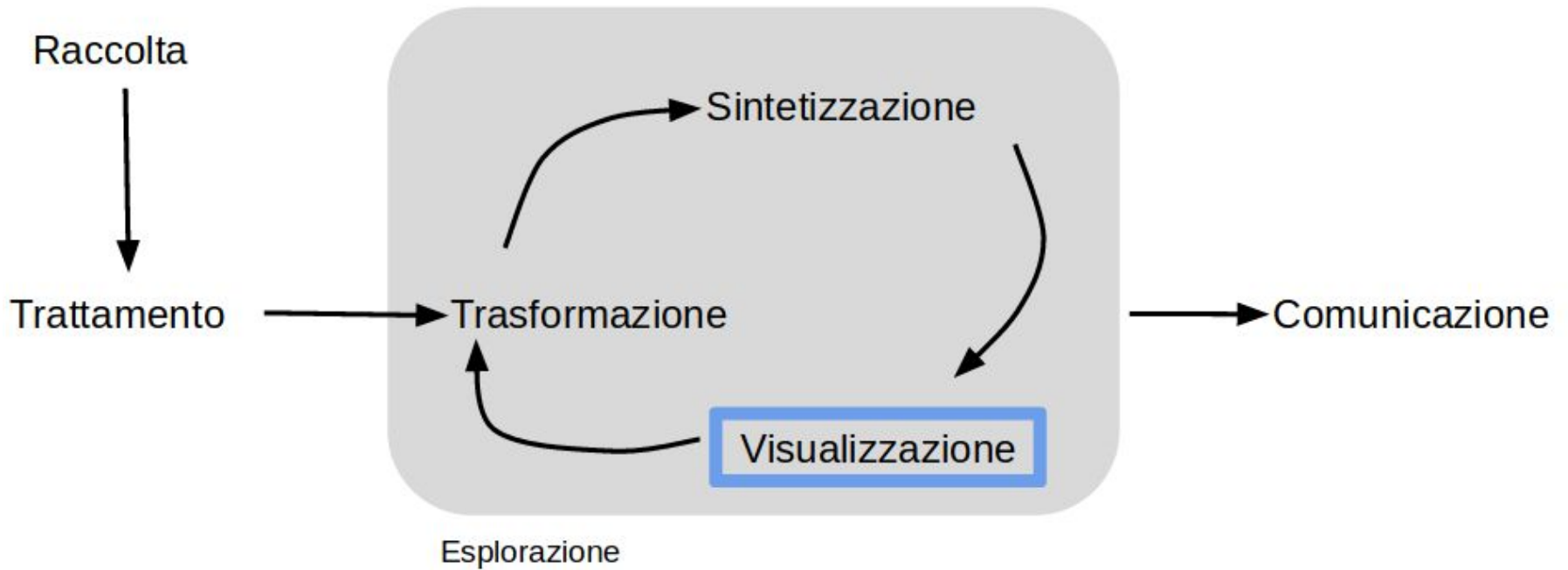
I risultati ottenuti sono una fotografia che descrive la sintesi dell'esistente

LA SINTESI DEI DATI: DATA MINING

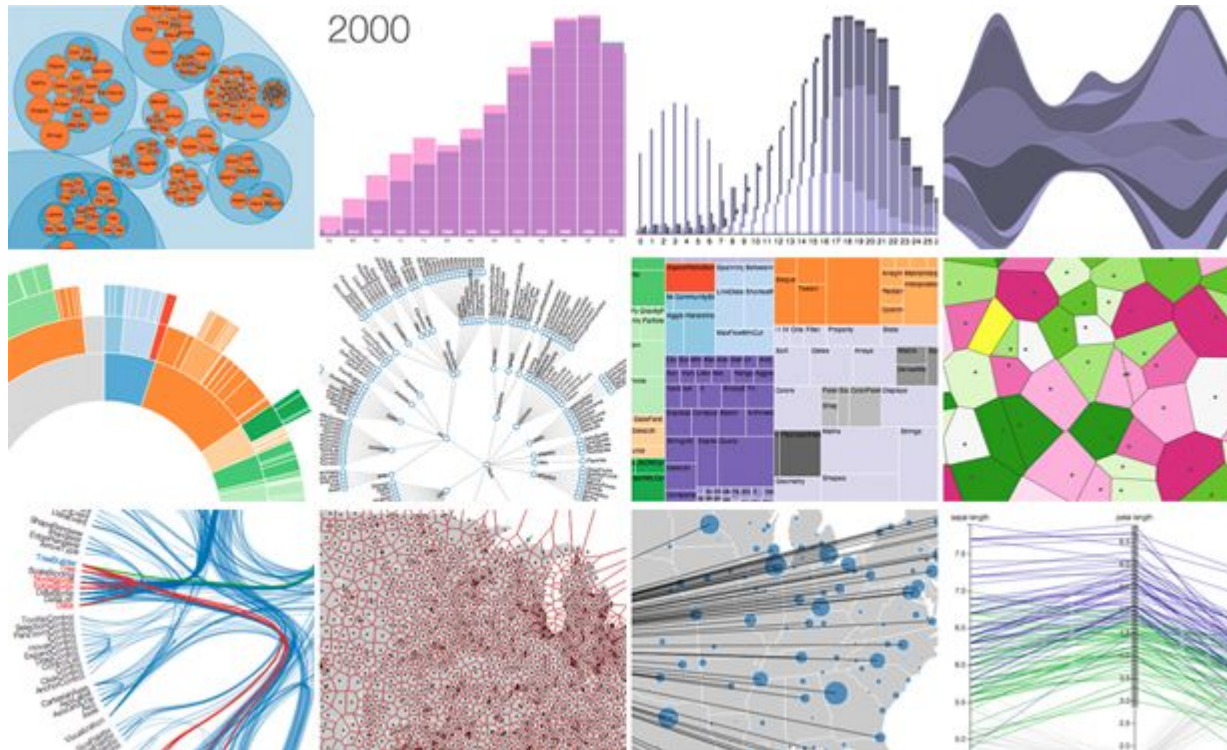
Un processo complesso d'identificazione nei dati di tendenze, strutture, modelli o trend validi, potenzialmente utili e a scopi prevalentemente predittivi.

Si prendono dei dati di “addestramento” e si individua il modello che si adatti ad essi e sia capace di funzionare anche ai dati target

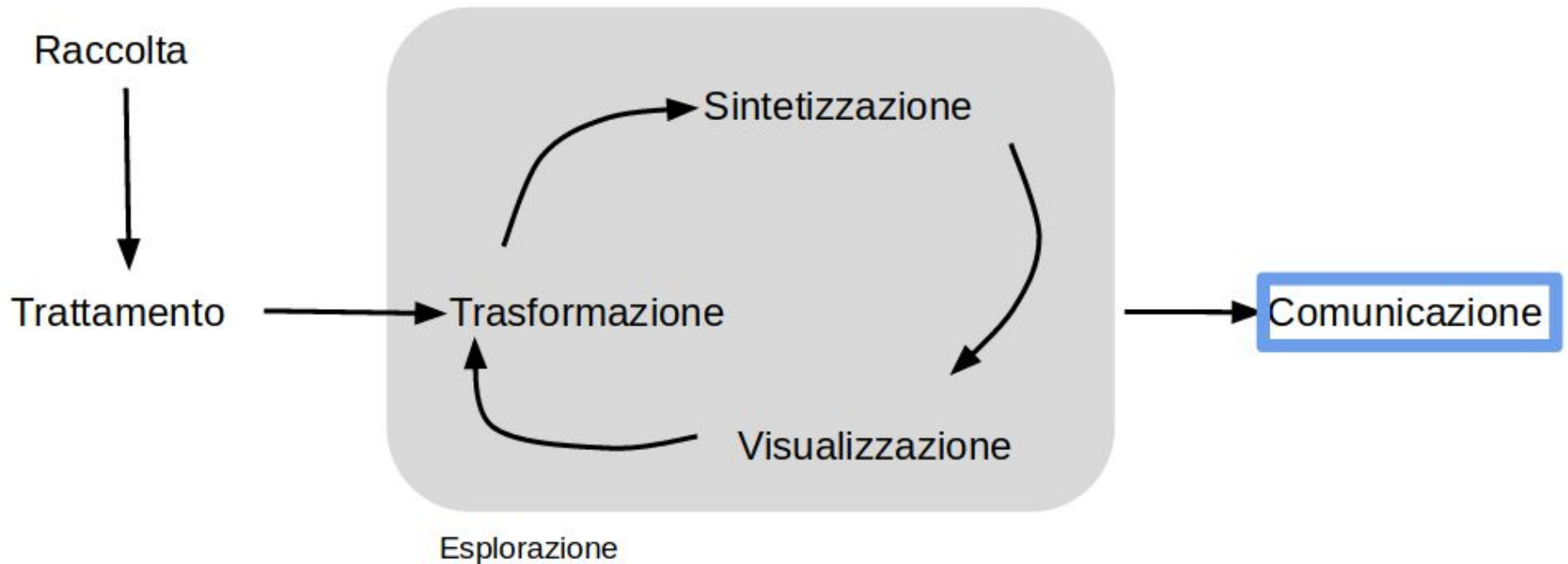
VISUALIZZAZIONE



VISUALIZZAZIONE

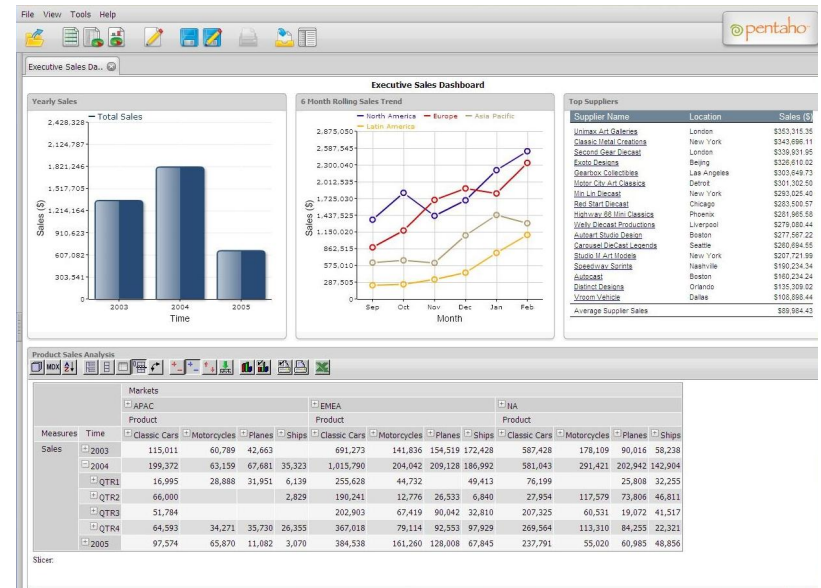
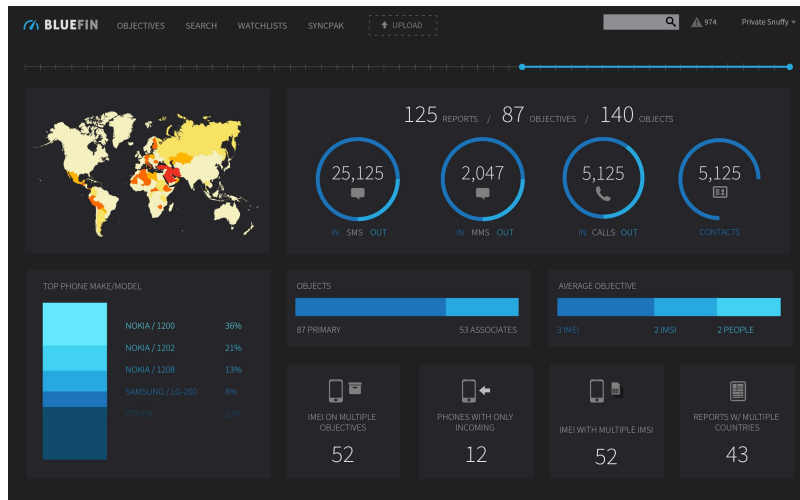


COMUNICAZIONE





COMUNICAZIONE





Grazie mille

Andrea Zedda

sito: **sardiniaopendata.org**
email: **a.zedda@kode-solutions.net**
account Twitter: **andria_tzedda**