POLITECNICO DI MILANO
GEO Laboratory



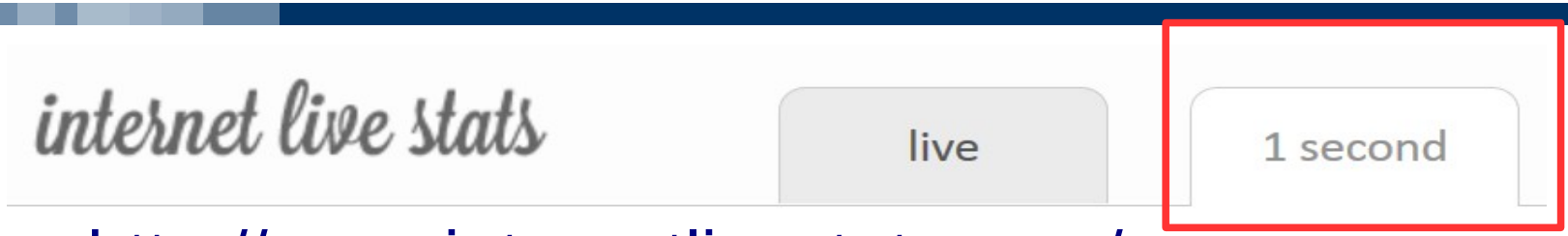


Citizen Generated Content and FOS Participative Platforms: geocrowdsourced data

Maria Antonia Brovelli

Politecnico di Milano, DICA – GEO Laboratory

Geocrowdsourcing (Big Geo Data)



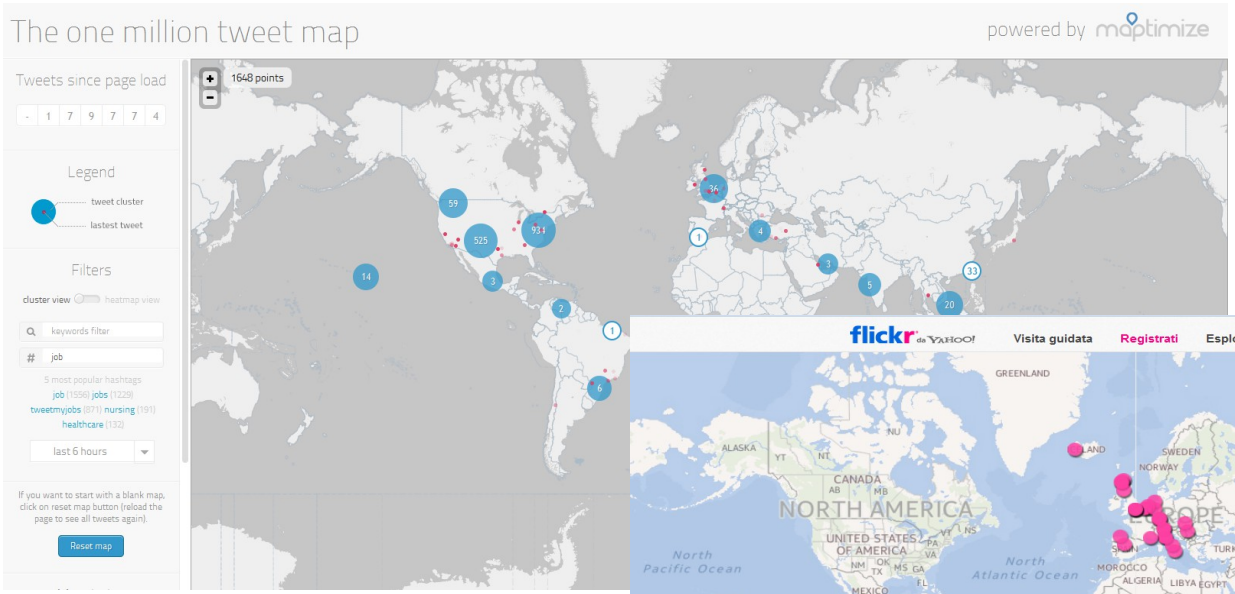
http://www.internetlivestats.com/

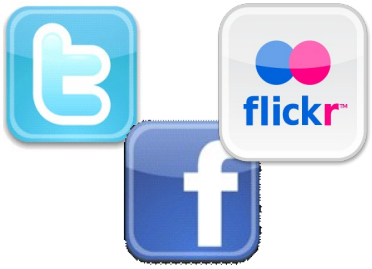
- 7,290 Tweets sent in 1 second
- 734 Instagram photos uploaded in 1 second
- 1,142 Tumblr posts in 1 second
- 2,214 Skype calls in 1 second
- 36,800 GB of Internet traffic in 1 second
- 55,467 Google searches in 1 second
- 127,793 YouTube videos viewed in 1 second
- 2,510,411 Emails sent in 1 second

Retrieved July 15, 2002 from http://www.internetlivestats.com/one-second/

Location based social network

- http://onemilliontweetmap.com/
- http://www.flickr.com/map





Citizen- Generated Geographic Information

Contributed Geographic Information (CGI) refers to geographic information "that has been collected without the immediate knowledge and explicit decision of a person using mobile technology that records location"

Table 1: Typology of Citizen-Generated Geographic Information

	Explicitly Geographic	Implicitly Geographic
Explicit or Active Volunteering	This is "True" Volunteered Geographic Information in the strictest sense. Examples include Open Street Map.	Volunteered (geo)spatial information (VSI). Examples would include Wikipedia articles about non-geographic topics, which contain place names
Implicit or Passive Volunteering	Citizen-generated geographic content (CGGC). Examples would include any public Tweet referring to the properties of an Identifiable place.	Citizen-generated (geo)spatial content (CGSC) such as a Tweet simply mentioning a place in the context of another (non-geographic) topic.

Source: Craglia, M., & Granell Canut, C. (2014). Citizen Science and Smart Cities.

Applications

- **Sensing Slow Mobility and Interesting Locations for Lombardy Region (Italy)**: a Case Study Using Pointwise Geolocated Open Data. An approach for collecting, unifying and analysing pointwise geolocated open data available from different sources with the aim of identifying the main locations and destinations of slow mobility activities.

- **Land Coverage Platform**. A **WebGIS platform** designed to publish the available land use and land cover maps of Europe at continental scale, were users can add to the platform photos from popular photo sharing services, in order to have a visual assessment of the available land coverages based on other user-generated contents available on the Internet.

Applications

- **Sensing the City**. A series of applications and procedures for the visualization and analysis of Social Media and Telecommunications Data (user-generated mobile network traffic).

  - **Sensing the city: calls and tweets.** A Web application for visualizing the number of calls exchanged between callers located in Milan and receivers located in other provinces in Italy
  - **Social media data management with Rasdaman**: Web application for testing the Web Coverage Processing Service (WCPS) OGC standard provided by Rasdaman
  - **Big data to netCDF**: Web application for creating netCDF files from time series telecommunications data
  - **Visualizing social media data with EST-WA:** EST-WA is a tool developed by GEOlab @Polimi for representing 4D variables (3D location of the variable values at different times) provided in netCDF format
  - **Relationships Between Telecommunications and Weather Data** Meteorological measurements of precipitation and temperature, as well as user-generated mobile network traffic is being analysed on a common space-time basis with a Two-Way Analysis of variance ANOVA on the city of Milan

Applications - 1

Sensing Slow Mobility and Interesting Locations for Lombardy Region (Italy): a Case Study Using Pointwise Geolocated Open Data.

Sensing Slow Mobility - 1

Aim of the study

The analysis purpose is to identify attractive locations and destinations of slow mobility activities (e.g. hiking, biking, etc.) within Lombardy Region (Italy) according to user's reported activities

Sensing Slow Mobility - 2

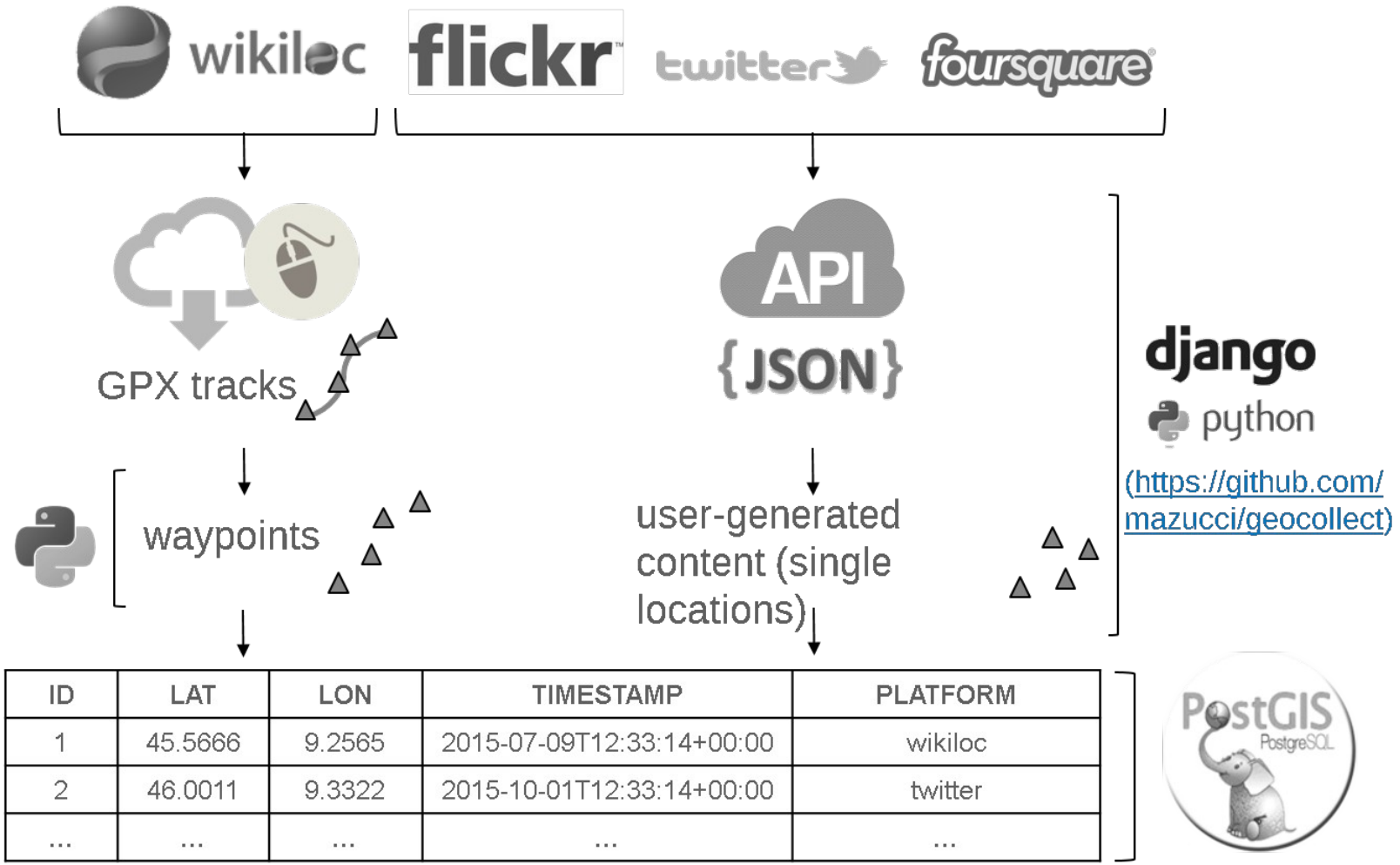
Selected CGI platforms:

- **Wikiloc (http://www.wikiloc.com)**: specialized platform for sharing and gathering insights on outdoor activities. Content is mainly GPX tracks. The collection is allowed only through manual download

- **Flickr (https://www.flickr.com)**, **Twitter (https://twitter.com)** and **Foursquare (https://it.foursquare.com)**: general purpose platforms that allow sharing different kind of content (pictures, check-ins, text messages, etc.). APIs are available to obtain content metadata in JSON format

Sensing Slow Mobility - 3

Data Collection and Storing

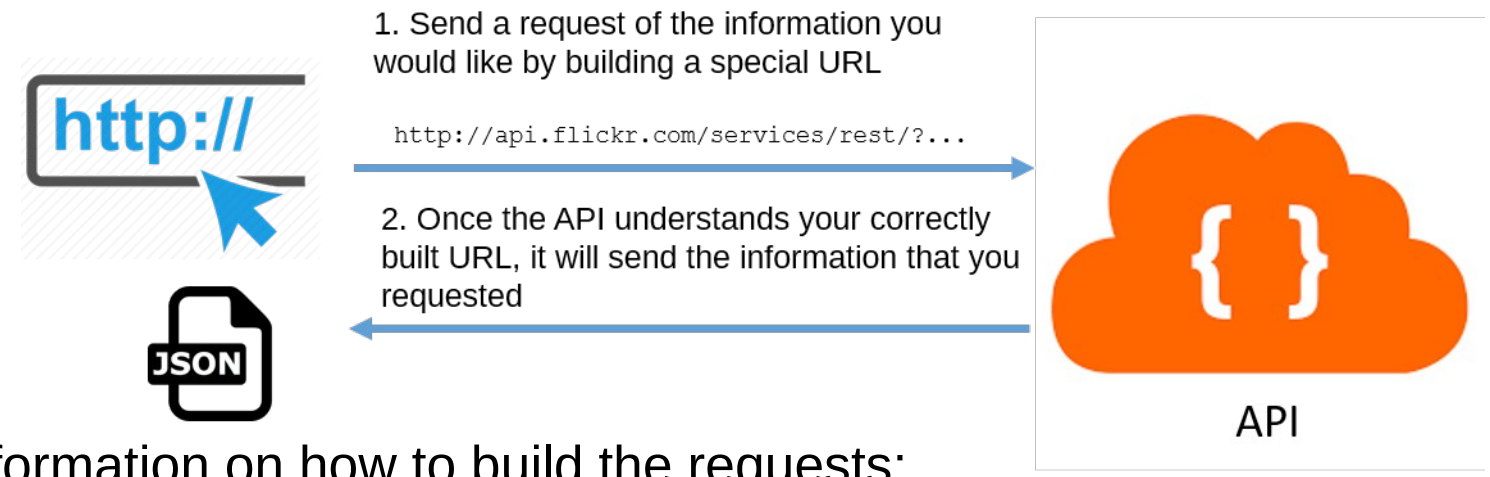


ID	LAT	LON	TIMESTAMP	PLATFORM
1	45.5666	9.2565	2015-07-09T12:33:14+00:00	wikiloc
2	46.0011	9.3322	2015-10-01T12:33:14+00:00	twitter
...	...	...	...	...

Data collection period: winter 2015/2016, n° of points ~ 2,300,000

Sensing Slow Mobility - 4

Application Programming Interface connections



Information on how to build the requests:

- **Flickr:** **https://www.flickr.com/services/api/.**
- **Twitter:** **https://dev.twitter.com/overview/documentation**
- **Foursquare:** **https://developer.foursquare.com/**
- Check : https://github.com/mazucci/geocollect for all the information on how to connect to the APIs

Sensing Slow Mobility - 5

Data filtering

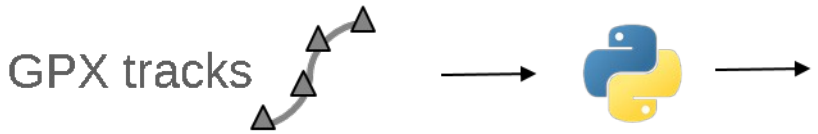
- Tracks speed was calculated with a python script: the difference between position of the beginning and end of the track gave the distance traveled, same approach for the timestamp gives time traveled. With distance and time the speed was calculated.

ID	LAT	LON	TIMESTAMP	PLATFORM
1	45.5666	9.2565	2015-07-09T12:33:14+00:00	wikiloc
2	46.0011	9.3322	2015-10-01T12:33:14+00:00	twitter
...	...	...	...	...

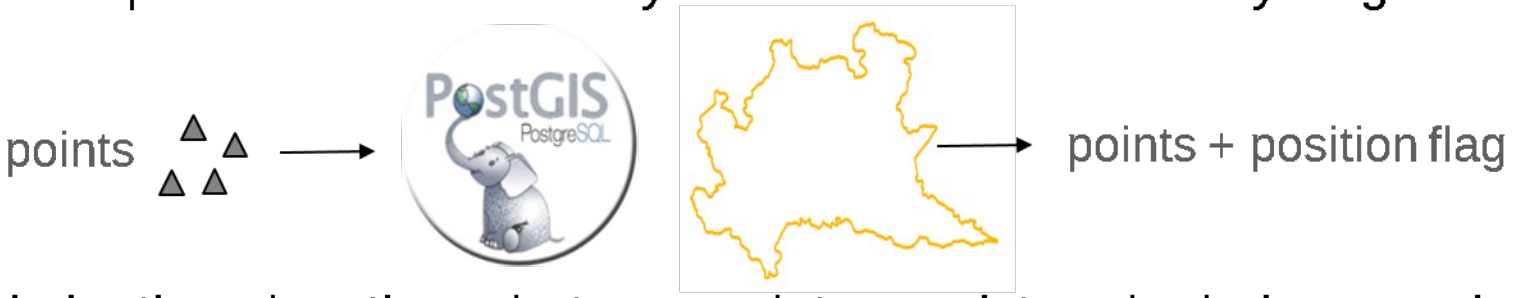
Sensing Slow Mobility - 6

Data filtering

- ✓ Check on Wikiloc tracks average speeds to include only non-motorized transportation (*speed threshold set to 22 km/h):



- ✓ Check on position to include only data within the Lombardy Region



- ✓ Discriminating by time between data registered during weekends (We-Wd=1) and weekdays (We-Wd=0)

*Gilani, H., 2005. Automatically Determining Route and Mode of Transport Using a GPS Enabled Phone. PhD thesis, University of South Florida.

ID	LAT	LON	TIMESTAMP	PLATFORM	SPEED CHECK	POSITION CHECK	We - Wd
1	45.5666	9.2565	2015-07-09T12:33:14+00:00	wikiloc	0	1	0
2	46.0011	9.3322	2015-10-01T12:33:14+00:00	twitter	1	0	1
...	...	...	...	...	...	...	...

Sensing Slow Mobility - 7

Data Analysis

Purpose: Identification of the most visited locations by looking for atypical spatial patterns as well as concentration of user-generated content within the study region

- ✔ Comparison between different platforms
- ✔ Comparison between user activities during weekdays and weekends
- ✔ Comparison between different spatial analysis techniques

Selected techniques:

- ✔ Concentration Maps
- ✔ Hot-spot Analysis (Exploratory Spatial Data Analysis)

Sensing Slow Mobility - 8

Techniques overview

Concentration Maps

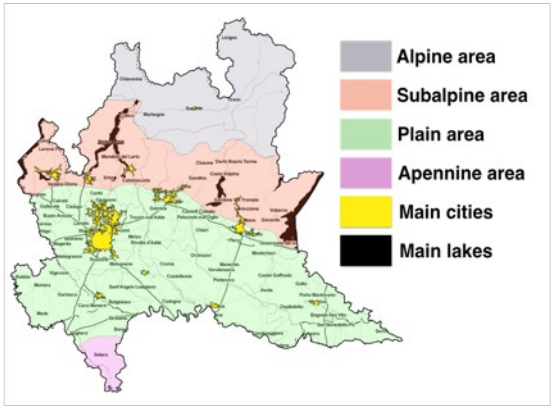
- ✔ Interpolated surface showing the density of occurrence of sparse point trough a color gradient or patches
- ✔ Requires to define an interpolating function and influence radius to compute the density surface

Hot-Spot Analysis

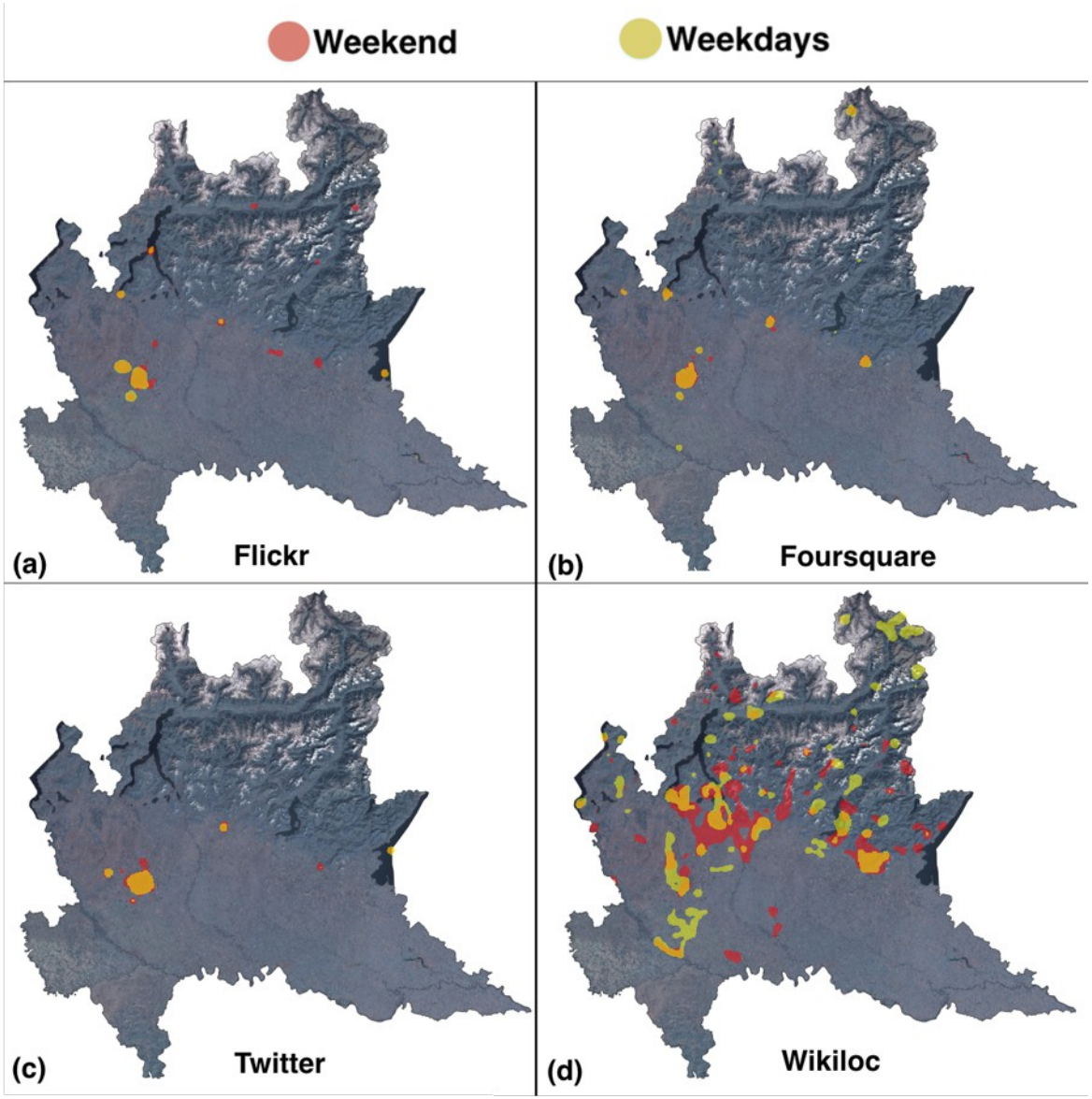
- ✔ Underlines where locational similarity is matched by attribute correlation in a spatial dataset by mean of statistical analysis (i.e. Getis-Ord GI* local statistic)
- ✔ Requires sparse points aggregation into representative points for any parcels of the study area as well as the identification of distance threshold to compute the local statistics in a defined region surrounding any points

Sensing Slow Mobility - 9

Concentration maps

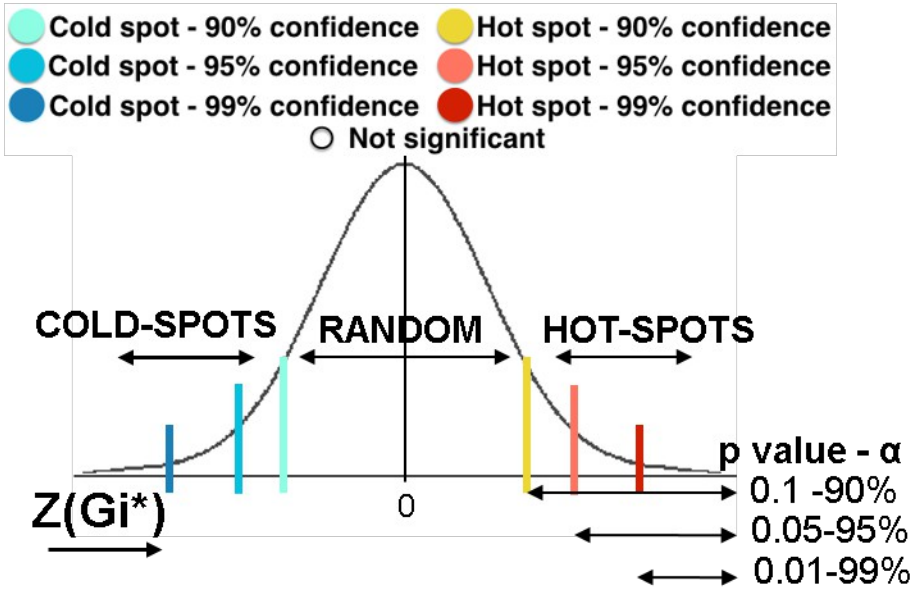


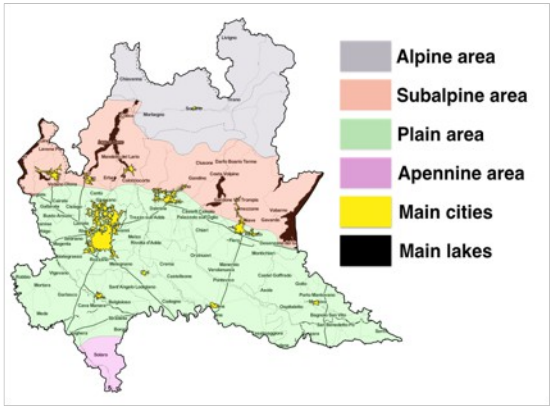
Schematic reference map including the main territorial features of Lombardy Region.



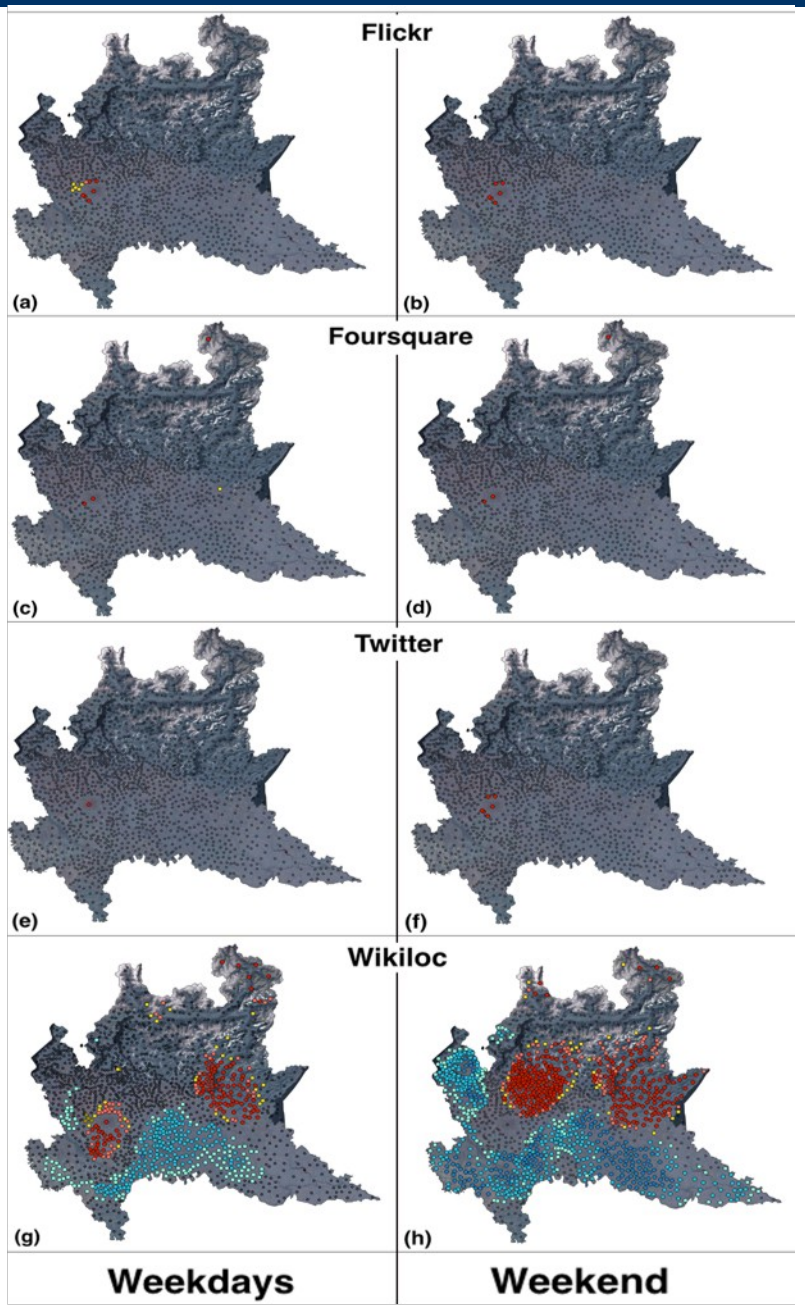
Sensing Slow Mobility - 10

Hot-Spot maps





Schematic reference map including the main territorial features of Lombardy Region.



Sensing Slow Mobility - 11

Results and Discussions

- ✔ Wikiloc data better describes locations for slow mobility activities with respect to the other platforms
- ✔ Flickr, Foursquare and Twitter data shows redundant places of interest across the region focused on the main cities (which are reasonably popular locations)
- ✔ Hot-spots concentrate around some of the main cities as well as in the alpine area. During weekend a strong hot-spot concentration appears all along the subalpine area and lakes.
- ✔ Cold-spots are located mainly in the plain area.
- ✔ Concentration maps retrace closely the patters highlighted by the hot-spot maps

Further improvements

- ✔ Results may be improved for the general purpose platforms performing specific data filtering (e.g. through keywords, hashtags, venues category etc.)
- ✔ Inclusion of Explanatory Spatial Data Analysis tools into QGIS*

*Oxoli D., Zurbarán M.A., Shaji S., Muthusamy A.K. (2016) Hotspot analysis: a first prototype Python plugin enabling exploratory spatial data analysis into QGIS. PeerJ Preprints 4:e2204v2 https://doi.org/10.7287/peerj.preprints.2204v2

Applications - 2

Land Coverage Platform. **A WebGIS platform designed to publish the available land use and land cover maps of Europe**

Land Coverage Platform - 1

Aim of the study

Implementing an open-source WebGIS aiming to collect, visualize, analyze and compare the land use and land cover datasets freely available for the Europe area in a single platform.

Research topics:

- ✔ Comparison between the LULC datasets in order to detect similarities and discrepancies
- ✔ Assessment of the classification quality of the LULC datasets

Land Coverage Platform - 2

Selected CGI platforms:

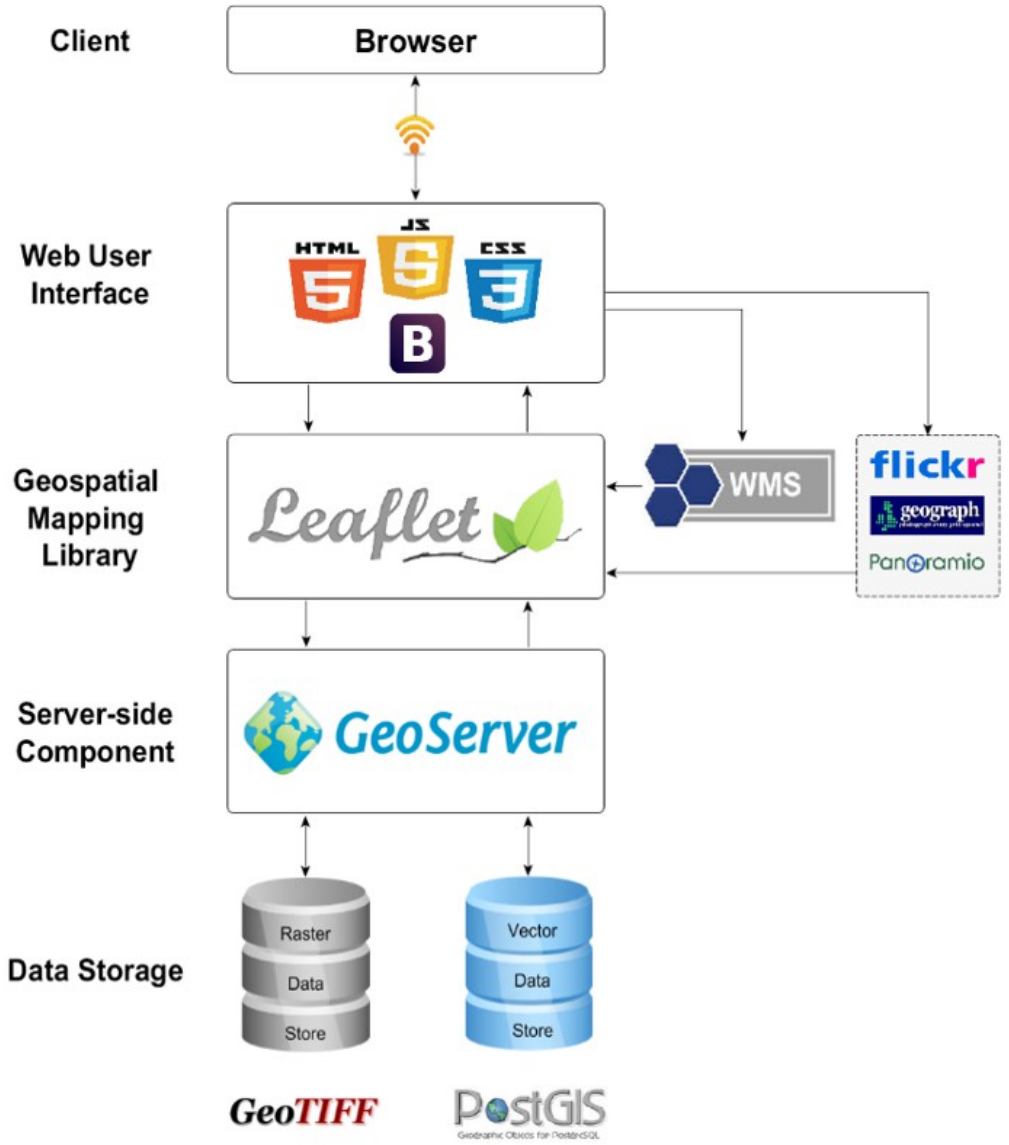
- **Geograph (http://www.geograph.org.uk):** a project limited to UK and Ireland which aims to collect geographically representative photographs at each node of a square grid with 1 km side

- **Flickr (http://flickr.com):** probably the most popular image hosting website for sharing personal photos.

- **Panoramio (http://www.panoramio.com):** a popular photo sharing website owned by Google. User submitted photos are published on the platform upon acceptance. The collection currently counts over 90 million photos

Land Coverage Platform - 3

Application Architecture



Land Coverage Platform - 4

Application Programming Interface connections

Information on how to build the requests:

- **Flickr:** **https://www.flickr.com/services/api/.**
- **Panoramio:** **http://www.panoramio.com/api/widget/api.html**
- **Geograph:** **http://www.geograph.org.uk/help/api**

Example of a Geograph request url:

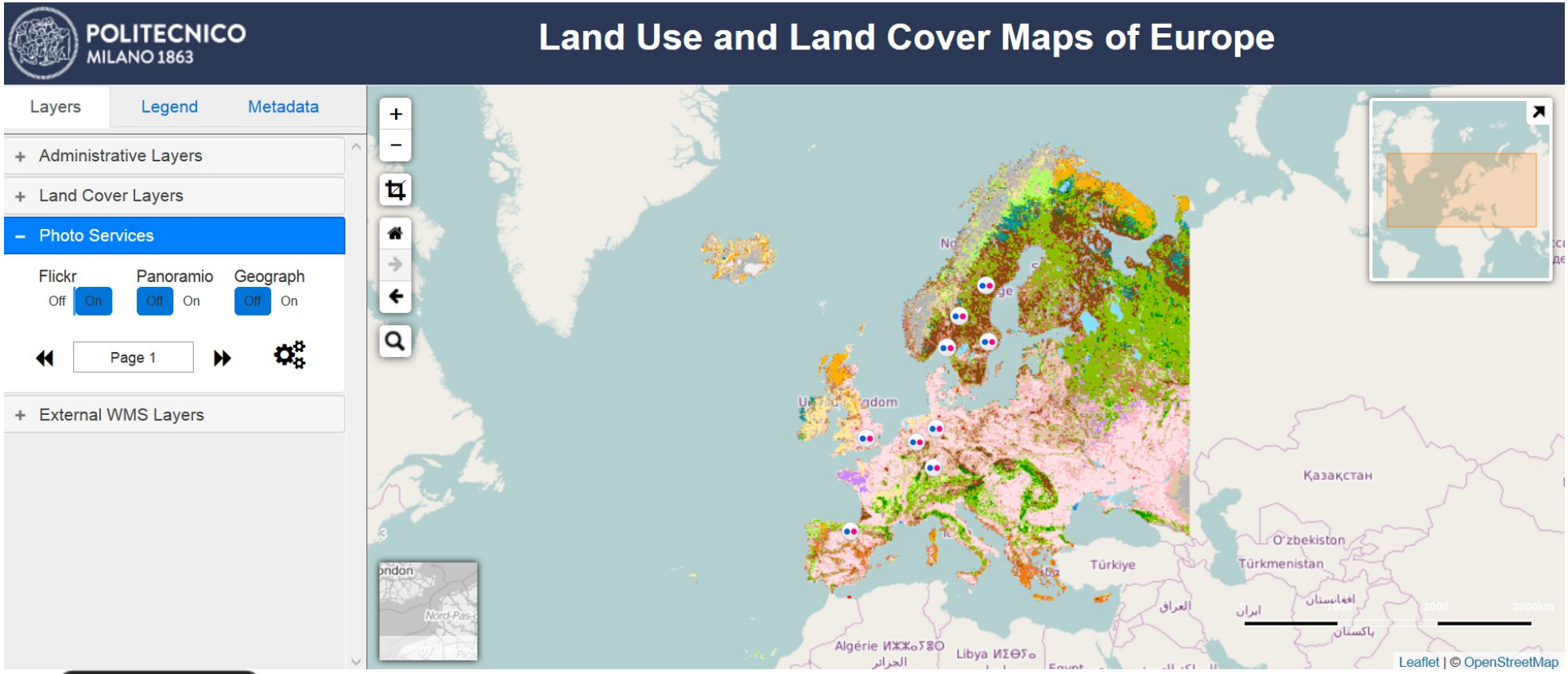
```
var geograph_URL =   "http://api.geograph.org.uk/api/photo/" + obj +
"/" + service_Photo.key + "?output=json";
```

That becomes:

```
http://api.geograph.org.uk/api/photo/4342537/e79cb167fb?output=json
```

Land Coverage Platform - 5

EU-LULC platform: client overview



Prototype available at http://131.175.143.84/LULC/

Land Coverage Platform - 6

Results and Discussions

- EU-LULC WebGIS is entirely built on open source infrastructure and open standards
- It enables the visualization and visual comparison of the available LULC maps of Europe

Further improvements

- Add other LULC available datasets
- Allow the upload of user data (raster and vector maps, photos)
- Improve the platform with processing functionalities to quantitatively compare the LULC maps:
- Compute statistics on land cover classes distribution for user-defined areas
- Assess land cover changes over time
- Evaluate the accuracy of a LULC map through the confusion matrix approach

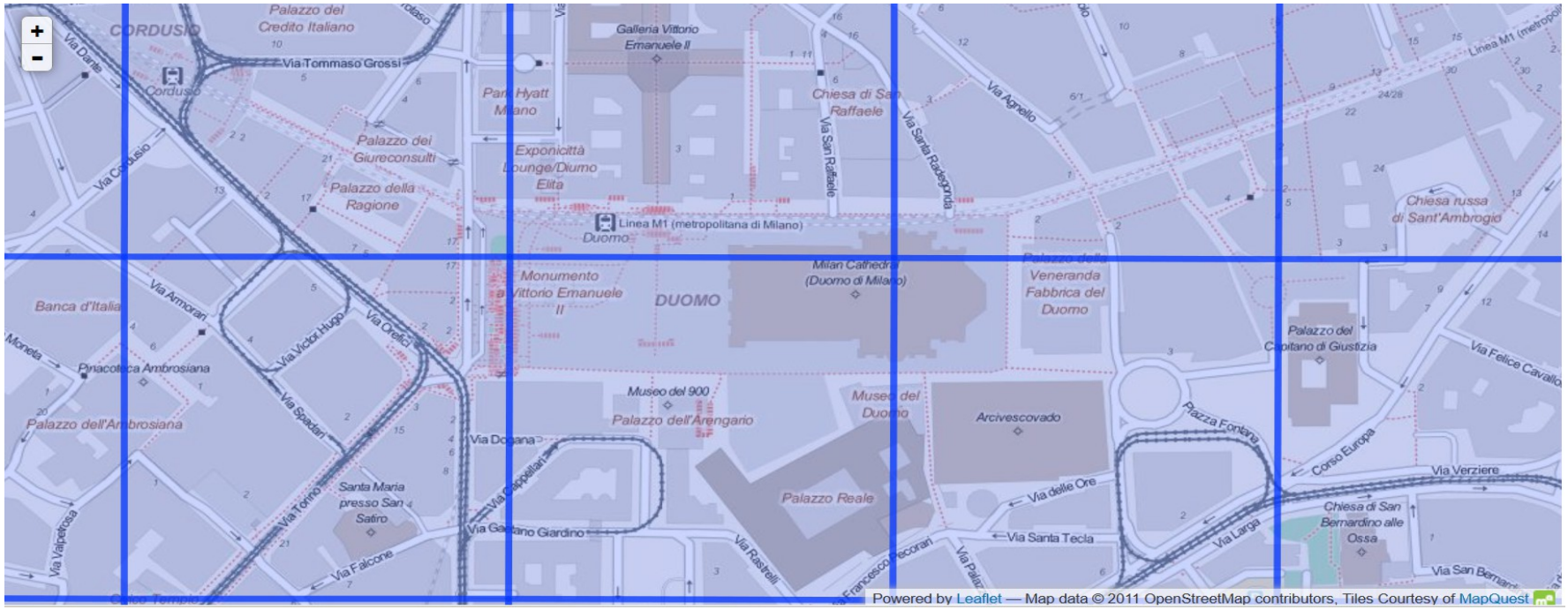
Applications - 3

Sensing the City. A series of applications and procedures for the visualization and analysis of Social Media and Telecommunications Data (user-generated mobile network traffic).

Sensing the City - 1

Geo Big Data: Milano GRID

- ✔ Two months of data, with a temporal step of 10 minutes
- ✔ Grid of 100 x 100 cells with size = 235 m



https://dandelion.eu/datamine/open-big-data/

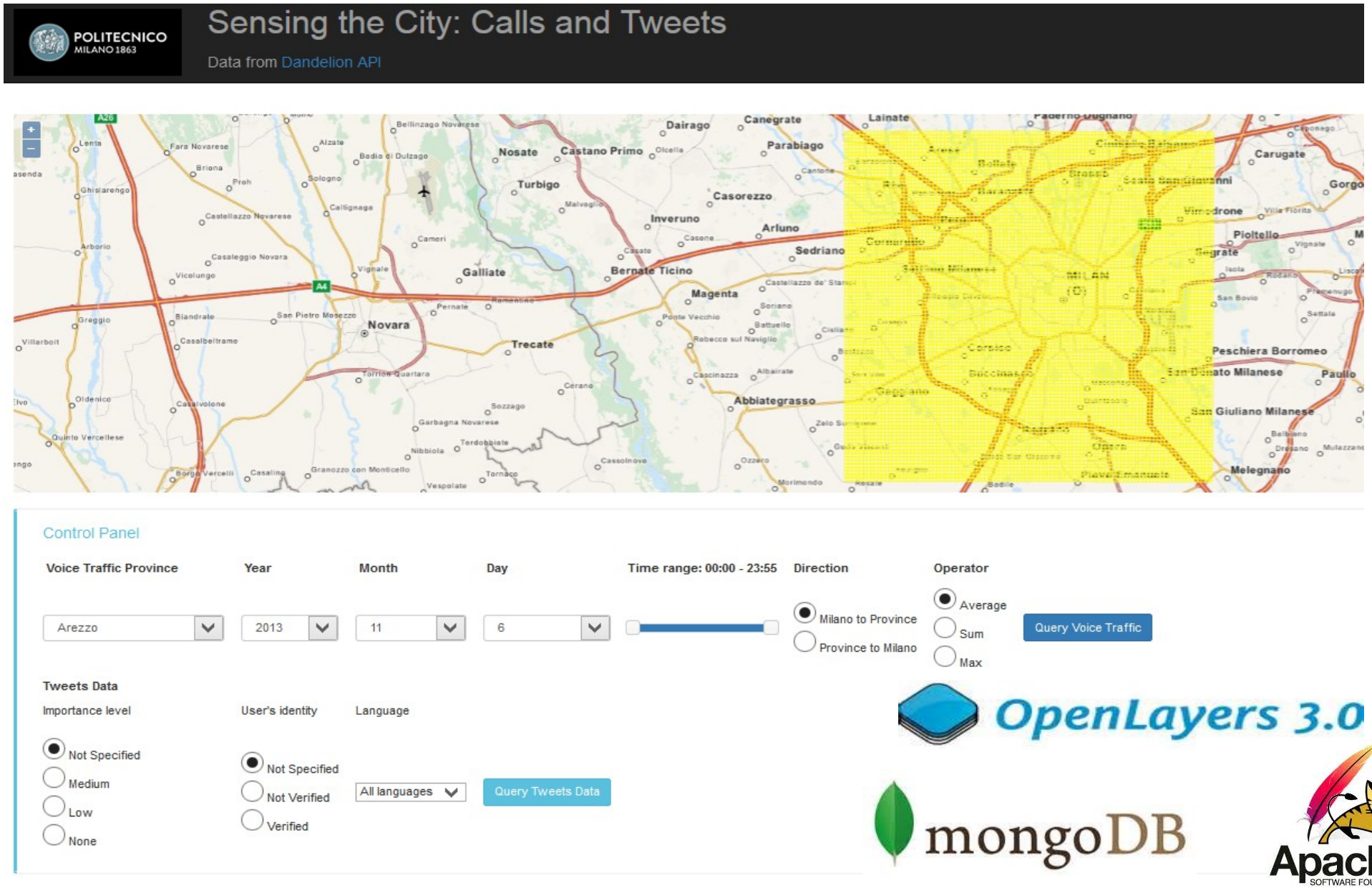
SPAZIODATI NODE TRENTO FONDAZIONE BRUNO KESSLER TELECOM ITALIA

Sensing the City - 2

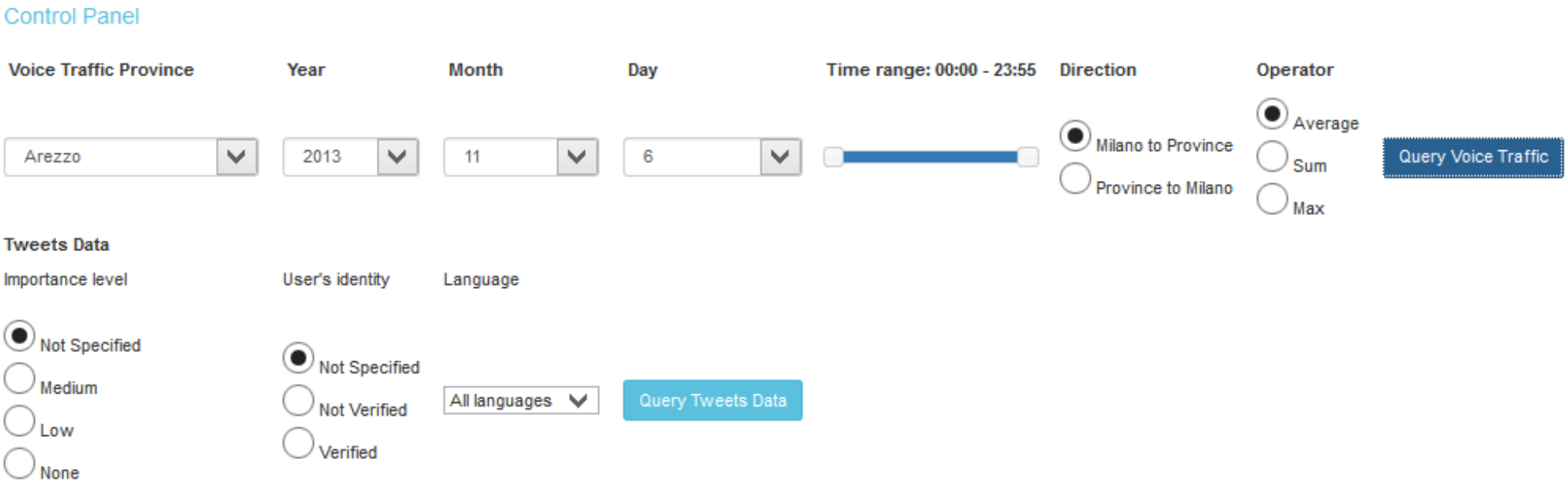
Geo Big Data: Milano GRID

- Received SMS: a Call Detail Record (CDR) is generated each time a user receives an SMS
- Sent SMS: a CDR is generated each time a user sends an SMS
- Incoming Calls: a CDR is generated each time a user receives a call
- Outgoing Calls: CDR is generated each time a user issues a call
- Internet: a CDR is generate each time :
  - a user starts an internet connection
  - a user ends an internet connection
  - during the same connection one of the following limits is reached:
  - 15 minutes from the last generated CDR
  - 5 MB from the last generated CDR
- Geolocalized Twetts (Anonymized twitter users)

Sensing the City - 3



Sensing the City - 4



http://landcover.como.polimi.it/BGDV/

Sensing the City - 5

POLITECNICO MILANO 1863

Sensing the City with Rasdaman

Demo

Choose your data

Variable: sms in

Select a variable (SMS, call, internet traffic)

Latitude 1: 45.3563

Longitude 1: 9.01149

Latitude 2: 45.5682

Longitude 2: 9.31269

Coverage

- [ ] Croplands
- [ ] Forest and Tundra
- [ ] Grass and Shrub
- [ ] Water and Wetlands
- [ ] Artificial cover
- [ ] Bare cover
- [ ] Permanent ice/snow

Filtering with date and land coverage classes

Sensing the City - 6

debian

rasdaman
raster data management

PostgreSQL
pgAdmin3

Java

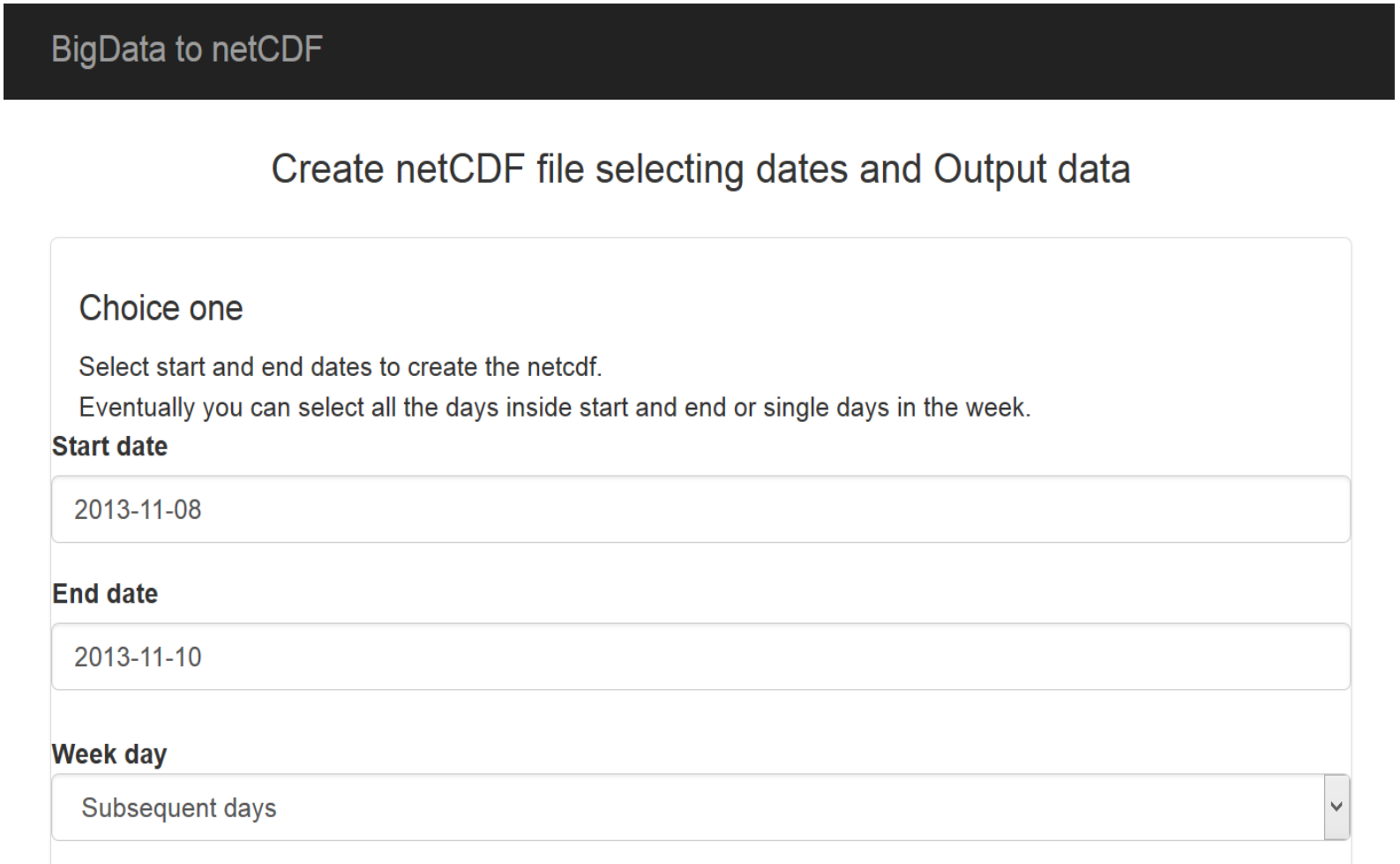
OpenLayers

eclipse
JUNO

+ Java, AJAX, Java Script, CSS, HTML

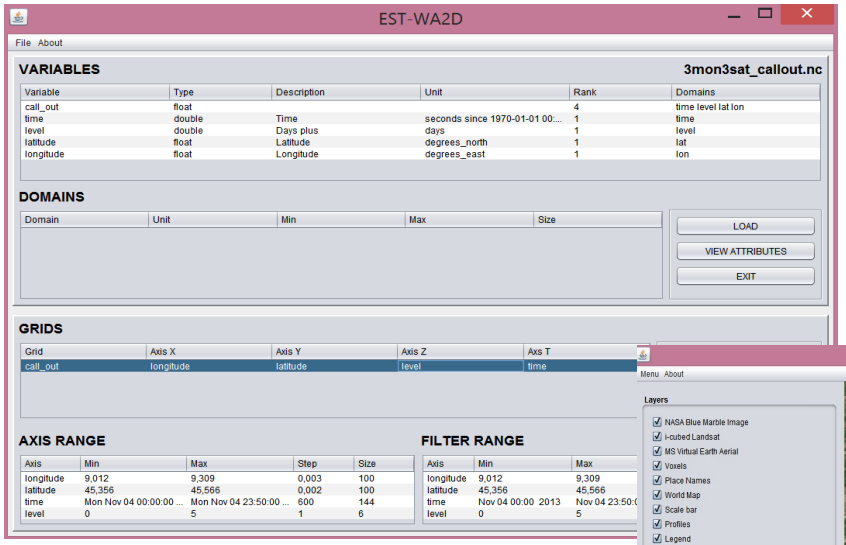
http://landcover.como.polimi.it/socialmedia_rasdaman/

Sensing the City - 7

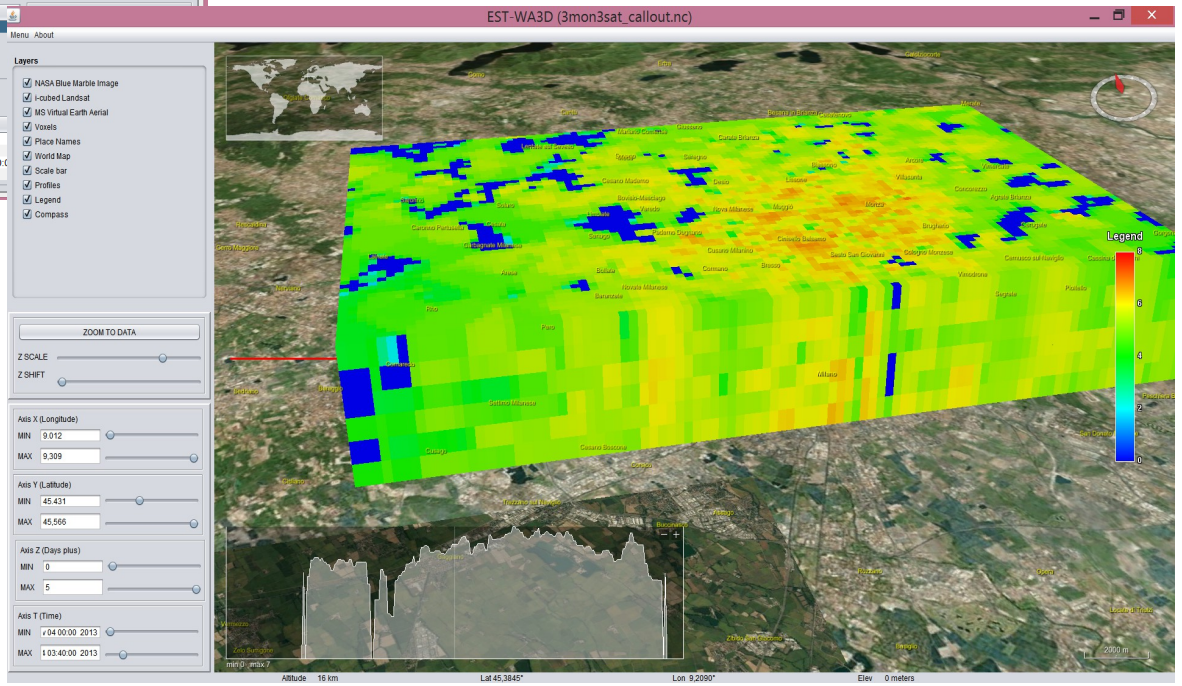


http://landcover.como.polimi.it/BigNetCDF/

Sensing the City - 8



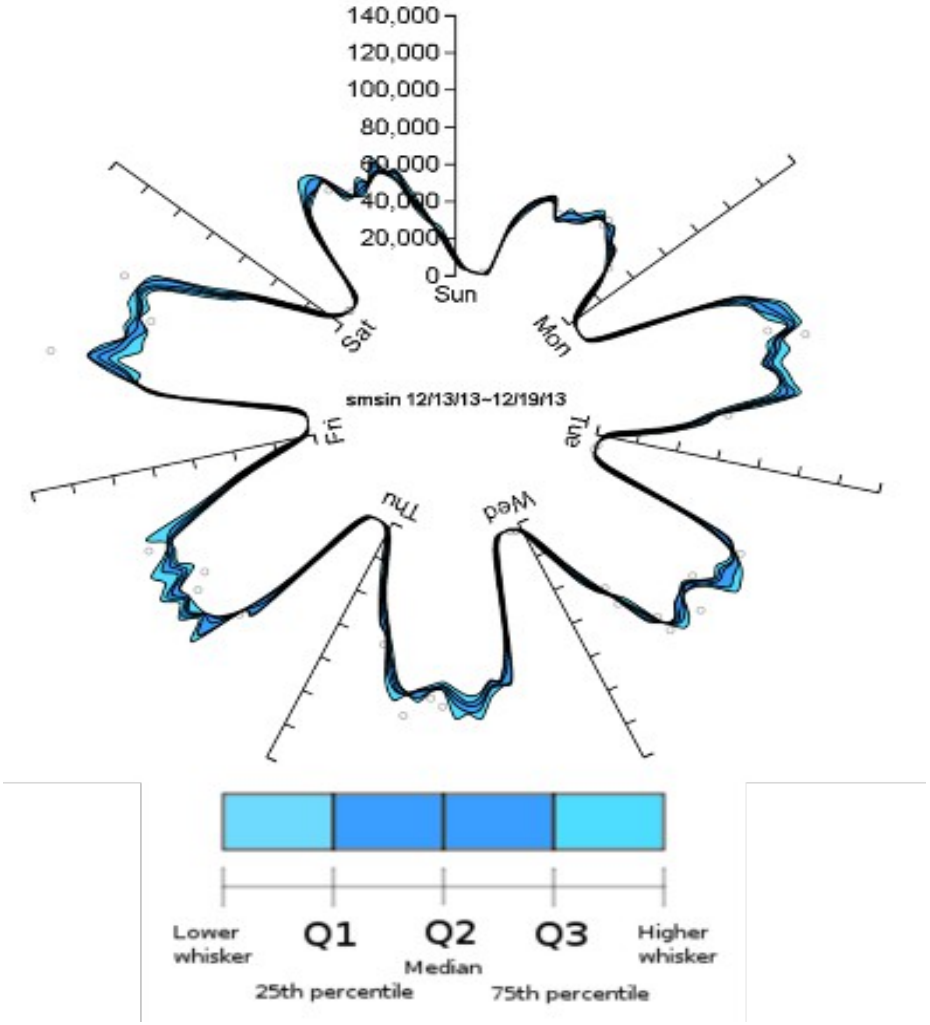
Filtering



Interactive multidimensional web visualisation - ESTWA

Web World Wind

Sensing the City - 9



Received SMS from Friday, December 13th to Thursday, December 19th for all Milano grid cells

http://landcover.como.polimi.it//BigNetCDF/cumulative.php

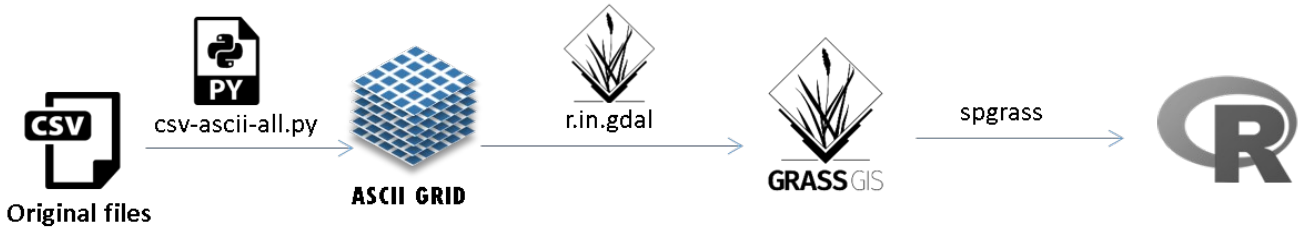
Sensing the City - 10

- Weather data comes from ARPA Lombardia's mesoscale meteorological network (http://www2.arpalombardia.it/siti/arpalombardia/meteo/osservazioniedati/datitemporeale/rilevazioni-in-tempo-reale/Pagine/Rilevazioni-in-tempo-reale.aspx)
- Land use data is being considered as well, taken from the Global Land Cover 30m (www.globallandcover.com)
- Data processing is being made with GIS Open tools such as :
  - GRASS GIS (https://grass.osgeo.org/) for preprocessing, basic statistics and filtering
  - QGIS (http://www.qgis.org/) for data visualization
  - R (https://www.r-project.org/) for advanced statistics analysis
  - Python Pandas, Scipy and Numpy libraries (https://www.python.org/) for advanced statistics analysis
- Data storage is being explored with MongoDB (www.mongodb.com) and RASDAMAN (http://www.rasdaman.com/

Sensing the City - 4

Data Processing

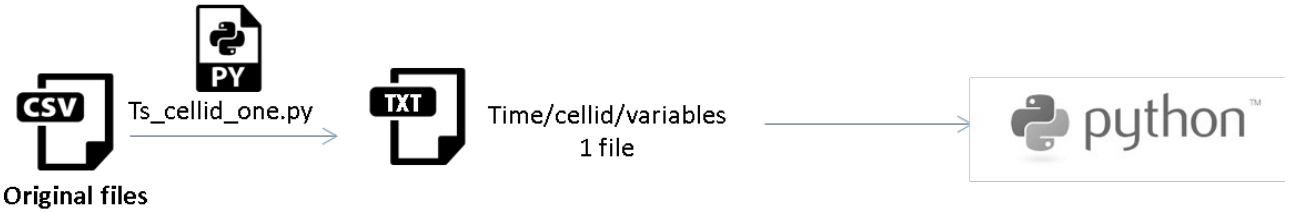
① Data processing with GRASS GIS



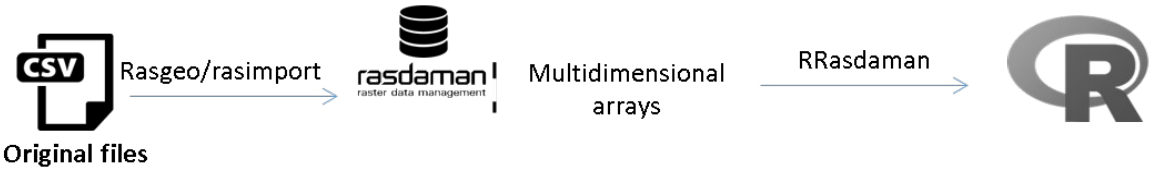
② Data processing with Python

Results next time!

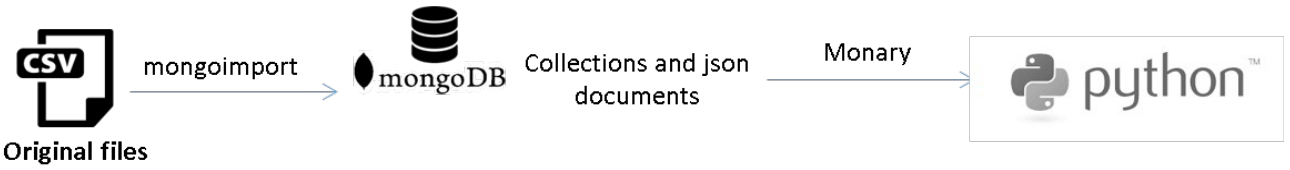
a)



b)



c)



Contacts

Thanks for your attention!

Thanks to all people of my team contributing on these topics: Carolina Arias, Eylul Kilsedar, Marco Minghini, Monia Molinari, Daniele Oxoli, Marco Pelucchi, Gabriele Prestifilippo, Giorgio Zamboni, Mayra Zurbaran

Politecnico di Milano

Laboratorio di Geomatica – Polo Territoriale di Como

Via Valleggio 11, 22100 Como (Italy)

maria.brovelli@polimi.it