URBAN-GEO BIG DATA

Project Meeting, 18 July 2018, Turin

POLIMI
Progress Report

Maria Antonia Brovelli, Candan Eylül Kilsedar, Monia Elisa Molinari, Daniele Oxoli

PRIN PROJECT: *URBAN GEOmatics for Bulk Information Generation, Data Assessment and Technology Awareness*

MINISTERO DELL'ISTRUZIONE DELL'UNIVERSITA' E DELLA RICERCA

POLIMI Tasks

- PART 1
  - Crowdsourced data: Twitter data collection and processing for investigating mobility perceptions and hubs

- PART 2
  - Land cover data collection; visualization of land use and land cover data (vector and raster), 3D buildings, displacement; raster processing

Crowdsourced data: Twitter data collection

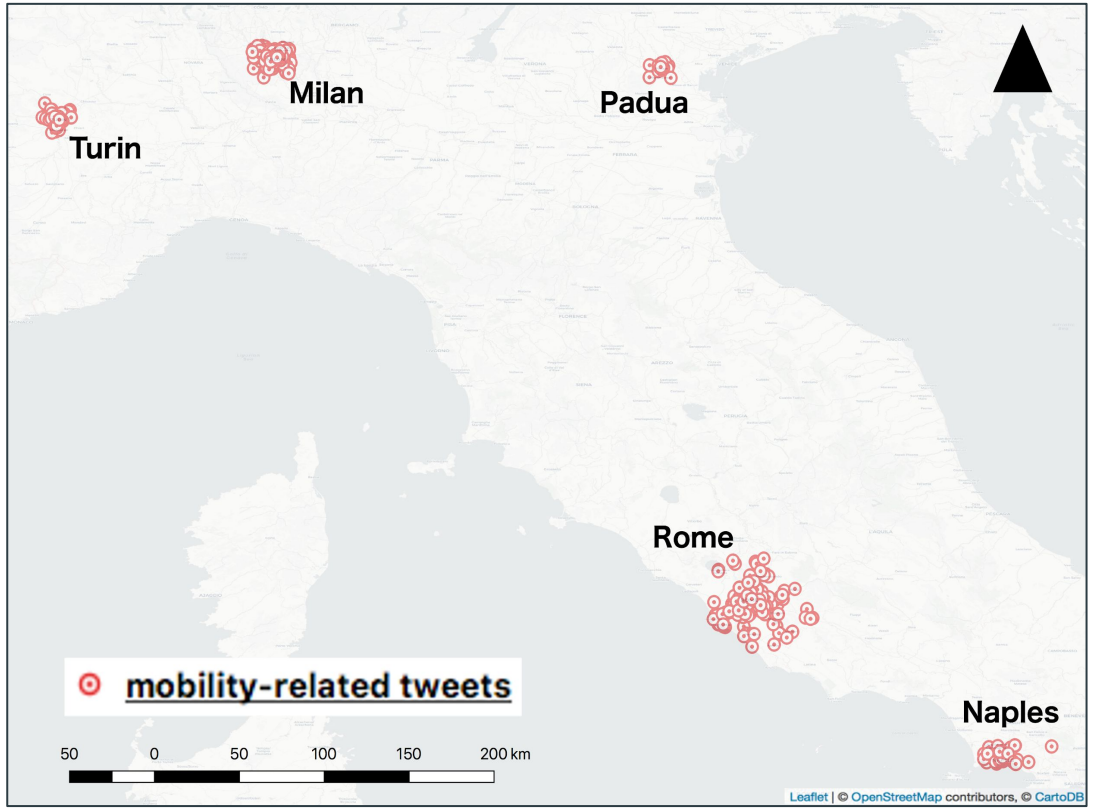
Collection (May 2017 - May 2018 on the 5 cities)	**N. of tweets (size in memory)**	**Language [%]**
test dataset (on mobility terminals in Milan only)	91893 (35.5 MB)	IT ~ 46% ENG ~ 32%
whole dataset	1450813 (5 GB)	IT ~ 58% ENG ~ 16%
whole dataset georeferenced	287521 (114.5 MB)	IT ~ 48%* ENG ~ 31%

*considered subset ~138k tweets (49% Rome, 33% Milan, 8.5% Naples, 7.5% Turin, 2% Padua)

Crowdsourced data: Twitter data processing (a)

- Extraction of mobility-related tweets* by means of text filtering based on keyword lists

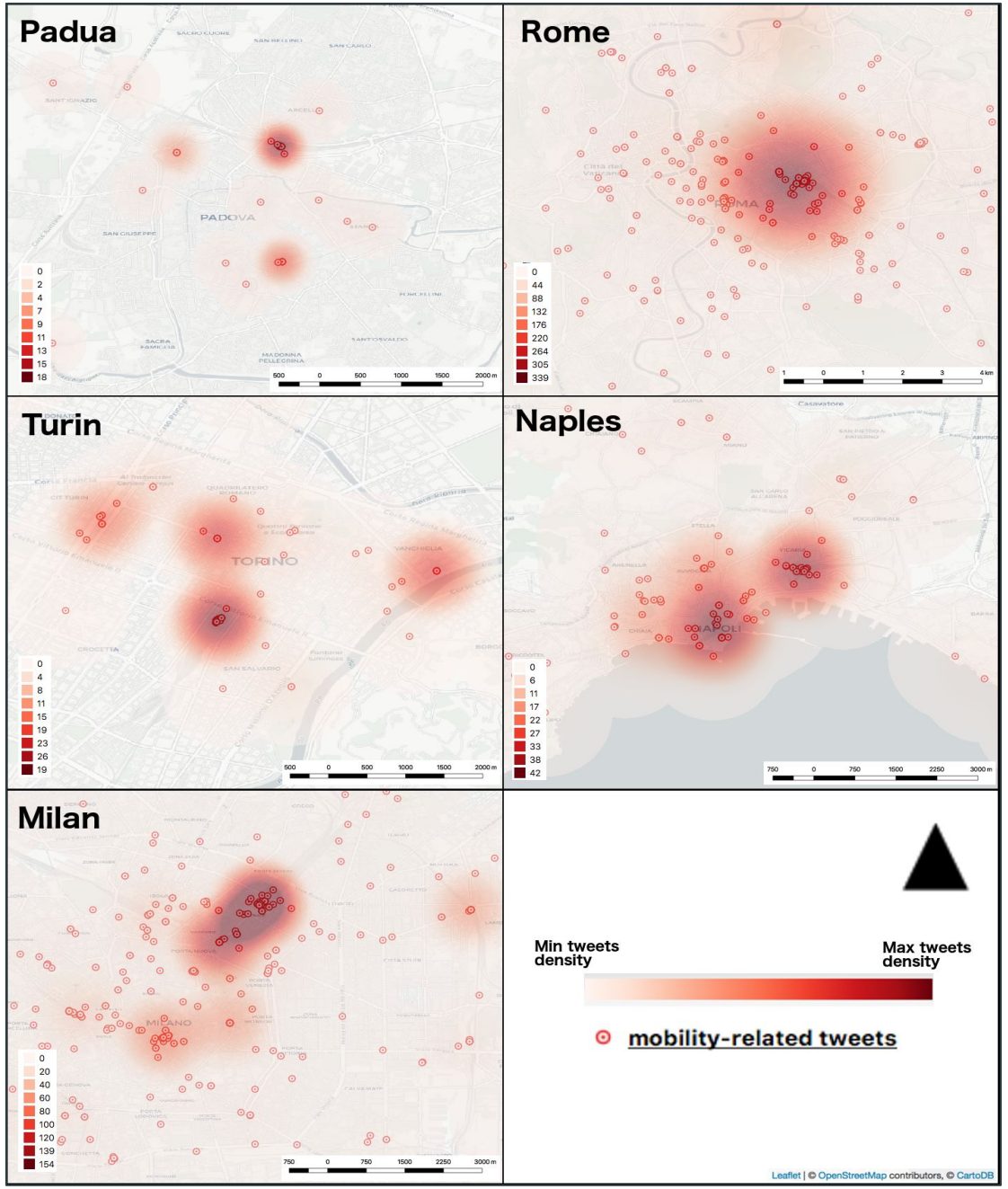
```
> list_means = ['auto','treno','trenitalia','autobus','tram','metropolitana'...]
> list_infrastructure = ['ferrovia','autostrada','fermata','stazione'...]
> list_action = ['sciopero'...]
```



*Extracted tweets ~ 2300
(42% Milan, 40% Rome, 8% Naples, 7% Turin, 3% Padua)

Crowdsourced data: Results (a)

- Density hotspots in the tweets spatial distribution:
  - identified main urban mobility hubs: Central Stations (Milan, Padua and Naples), Termini Station (Rome), Porta Nuova Station (Turin) as well as city centres (Naples)



Crowdsourced data: Twitter data processing (b)

- Classification attempted on the mobility-related tweet text messages (sentiment analysis):
  - positive or neutral messages
  - negative and/or complain messages

- Method:
  - Naive Bayes Classifier (supervised)
  - training set ~ 400 manually tagged tweets

Natural Language Analysis with Python NLTK

Crowdsourced data: Results (b)

- Classification results:
  - positive or neutral messages ~ 2150 (93%)


```
{"_id":{"$numberLong":"9931668775757864 97"},
"loc":{"type":"Point","coordinates":[9.17168219,45.47641111]},
"text":"Ok alla fine per smaltire la brace ho preso la bici...
#pedalachetipassa… https://t.co/ITL2xB1lgy",
"id":"9931668775757865 00",
"source":"\u003ca href=\"http://instagram.com\"
rel=\"nofollow\"\u003eInstagram\u003c/a\u003e",
"lang":"it","userid":312732546,"timestamp_ms":"1525624410419"}
```


  - negative and/or complain messages ~ 150 (7%)


```
{"_id":{"$numberLong":"885065748028289025"},
"loc":{"type":"Point","coordinates":[9.19363774,45.45780795]},
"text":"#Milano ... traffico, afa, confusione. Una città congestionata dai
lavori della metro. Solo… https://t.co/yzWU5HQ3tE",
"id":"885065748028289000",
"source":"\u003ca href=\"http://instagram.com\"
rel=\"nofollow\"\u003eInstagram\u003c/a\u003e",
"lang":"it","userid":40030031, "timestamp_ms":"1525624410691"}
```

Crowdsourced data: Discussions

- Data features
  - **Prons**: easy to handle and process, interesting spatial character
  - **Cons**: collection, integration with other data sources, text classification automation and accuracy
- Amount of valuable information (georeferenced & Italian language) on mobility from Twitter ~ 0.15 % of the total stream
- Most of this is out-of-scope or not directly spendable within the project case studies
- To do:
  - Consider also English Language tweets (at least)
  - Include temporal dimension

...Any comment, suggestion or feedback?

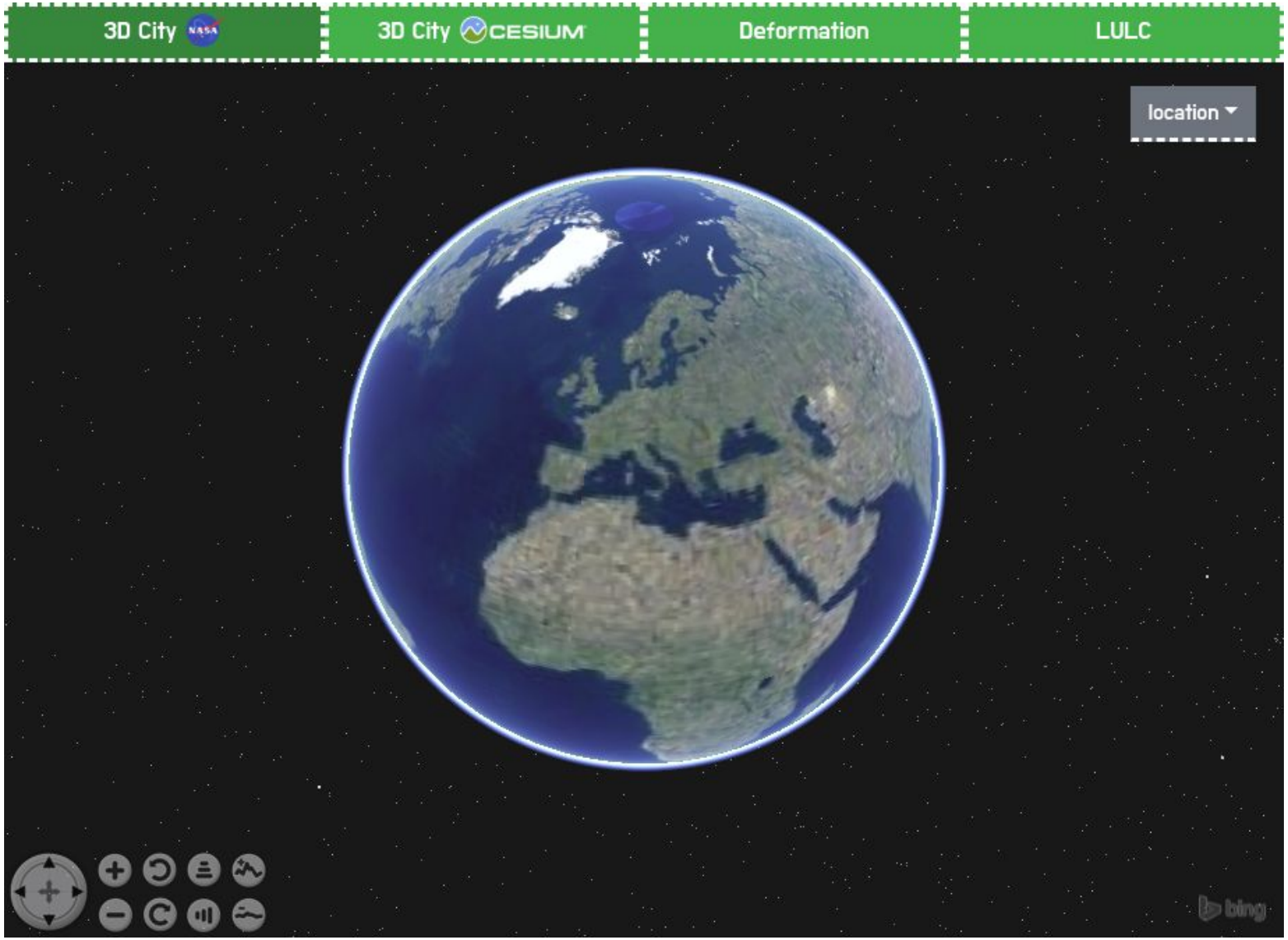
TC IV Symposium 2018
October 1 - 5, 2018, Delft, The Netherlands

POLIMI Tasks

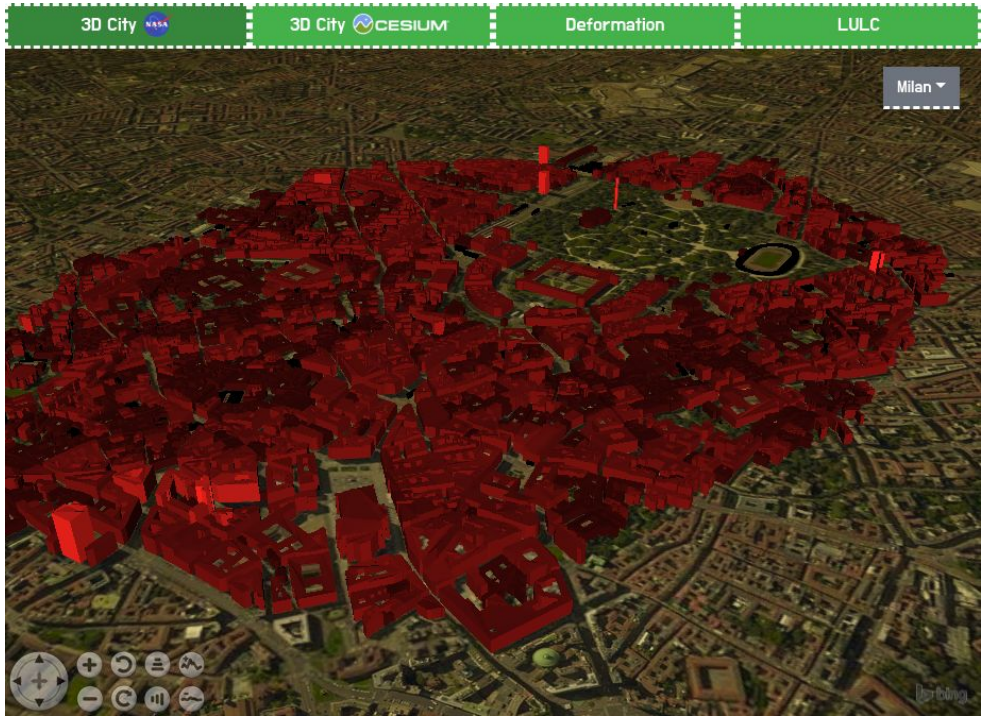
- PART 1
  - Crowdsourced data: Twitter data collection and processing for investigating mobility perceptions and hubs

- PART 2
  - Land cover data collection; visualization of land use and land cover data (vector and raster), 3D buildings, displacement; raster processing

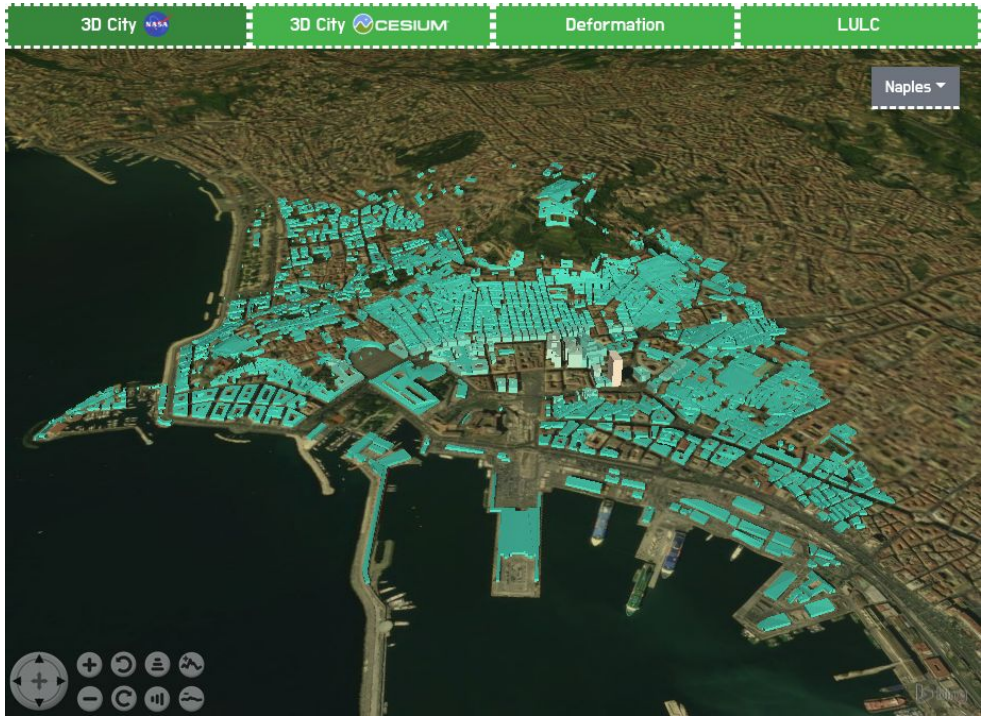
Web Application Interface



3D OSM Buildings



Milan



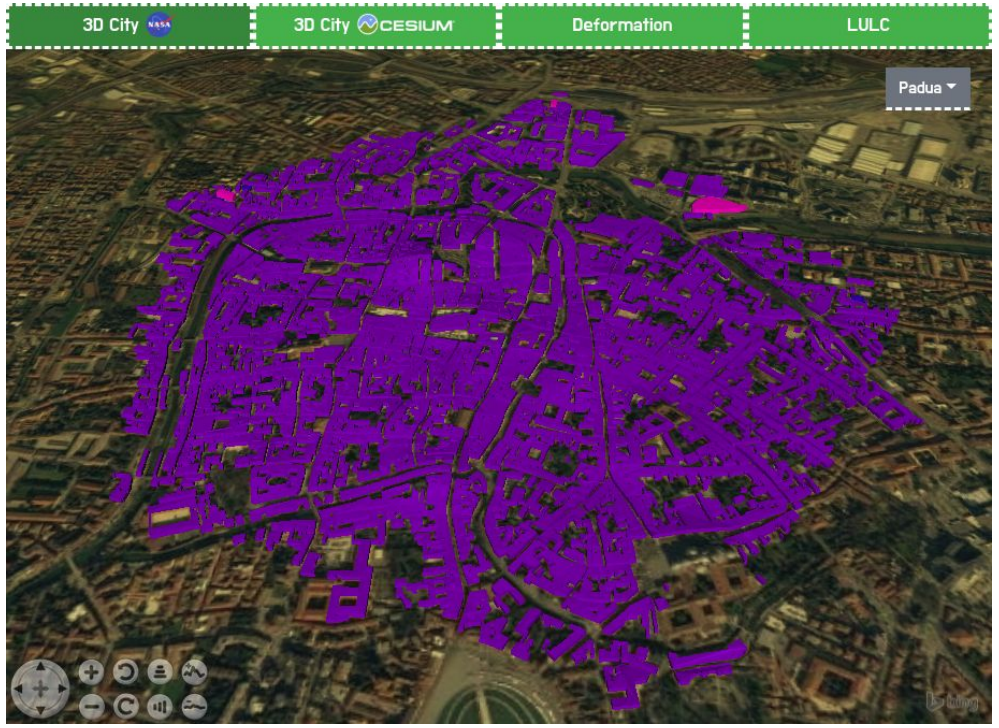
Naples

- The visualization is realized using 2017 GSoC project "3D OSM Plugin API" to NASA Web WorldWind API. Its source code is available at https://github.com/kilsedar/3dosm.
- Milan building heights are set using LiDAR and OSM GeoJSON in GRASS GIS.
- Naples building heights come from OSM.

PRIN PROJECT: *URBAN GEOmatics for Bulk Information Generation, Data Assessment and Technology Awareness*

3D OSM Buildings

Turin

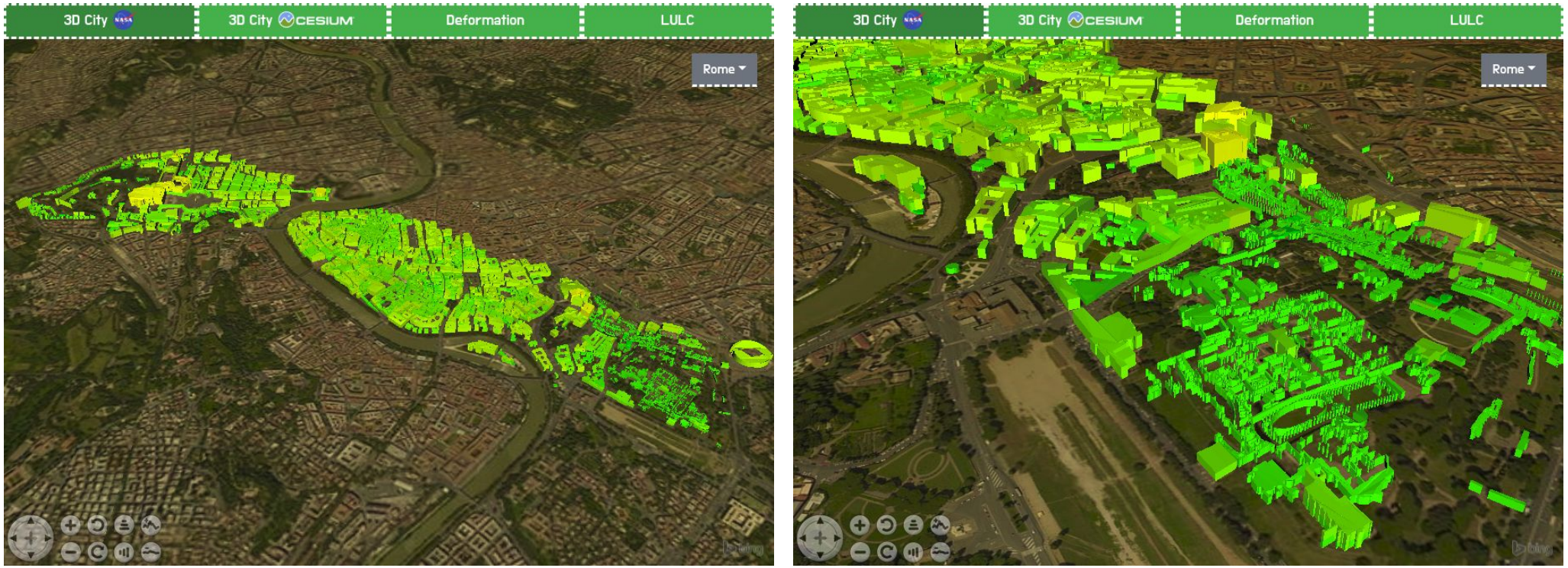


Padua

- The visualization is realized using 2017 GSoC project "3D OSM Plugin API" to NASA Web WorldWind API. Its source code is available at https://github.com/kilsedar/3dosm.
- Turin and Padua building heights come from OSM.

PRIN PROJECT: *URBAN GEOmatics for Bulk Information Generation, Data Assessment and Technology Awareness*

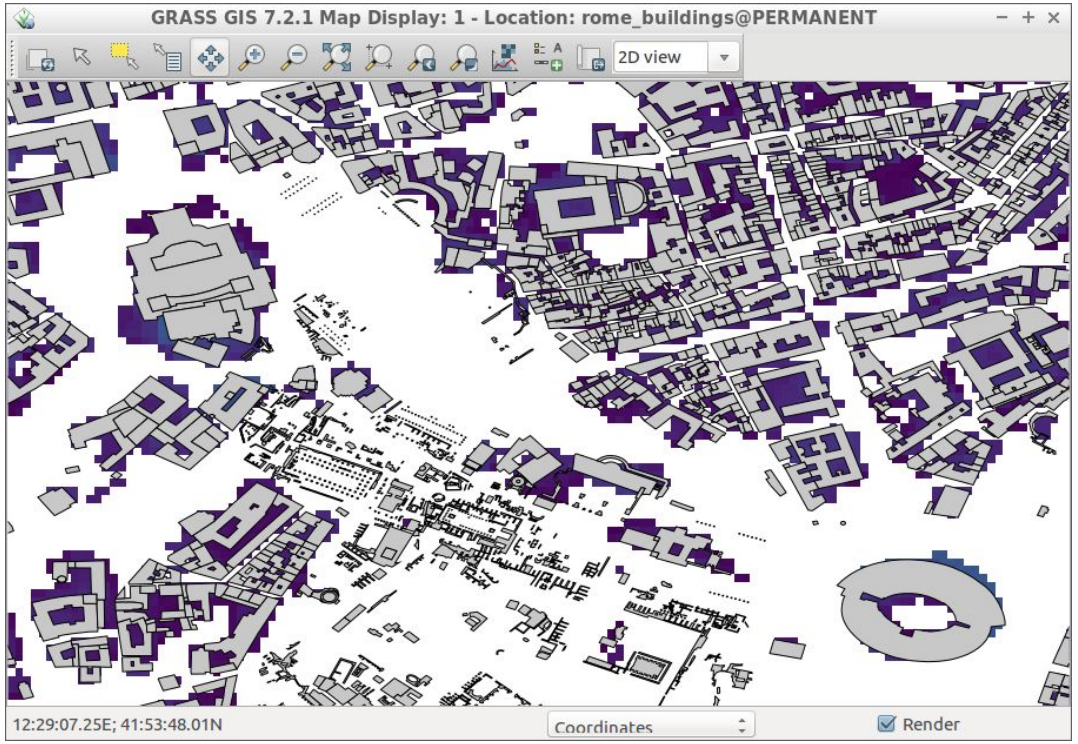
3D OSM Buildings



Rome

- The visualization is realized using 2017 GSoC project "3D OSM Plugin API" to NASA Web WorldWind API. Its source code is available at https://github.com/kilsedar/3dosm.
- Rome building heights are obtained using the Urban Atlas - Building Height 2012 data of the Copernicus programme and OSM GeoJSON in GRASS GIS.

3D OSM Buildings



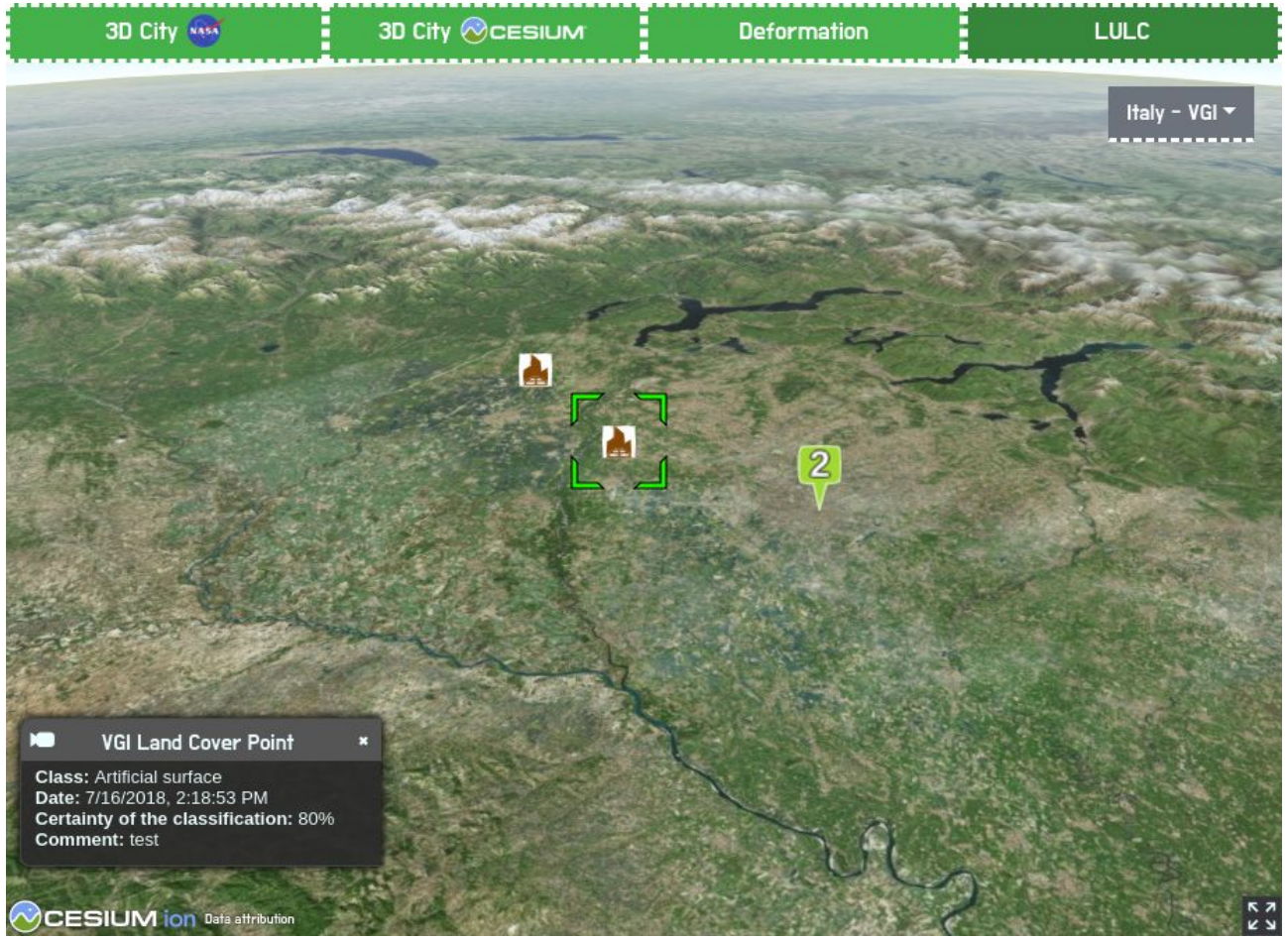
Overlap of raster data containing height information and the footprints of the buildings from OSM in GRASS GIS

- In GRASS GIS median of the pixel values within the footprint of each building is calculated and the value is assigned as a new attribute for that building, which is then an input for the visualization API.

Naples Deformation

- GeoServer and WMS are used for the visualization of displacement.
- Plot is drawn using Plotly.

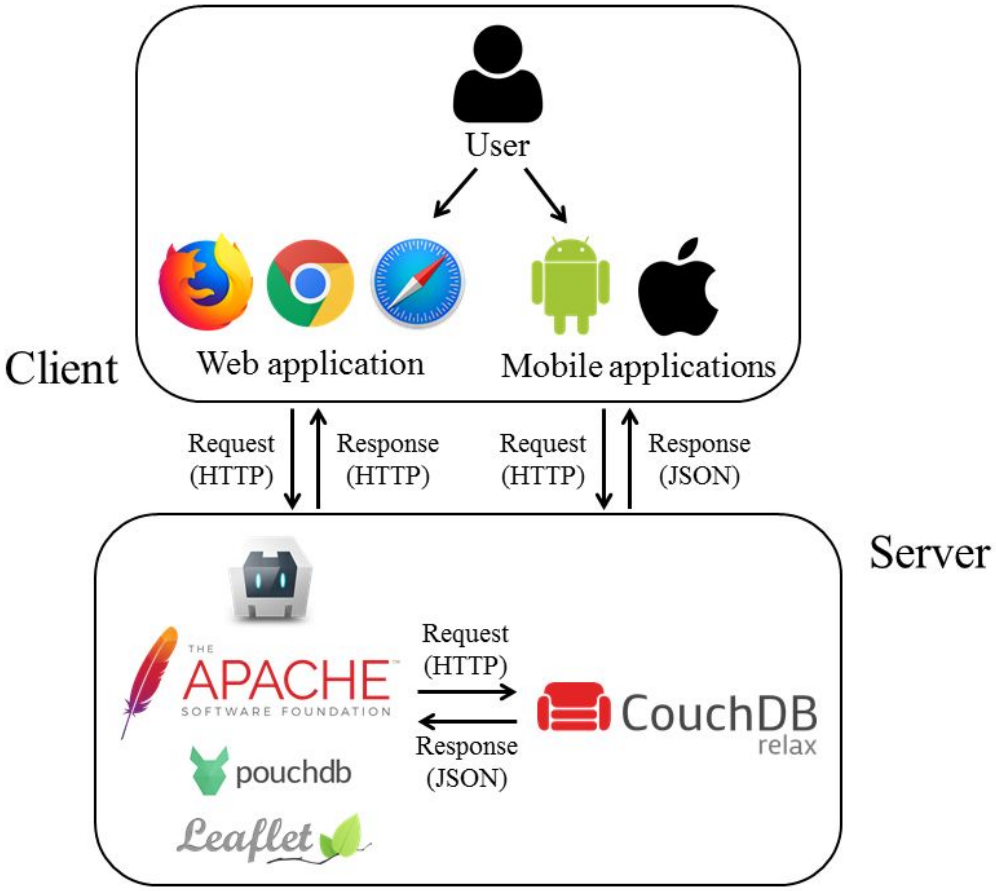
Land Cover VGI Visualization

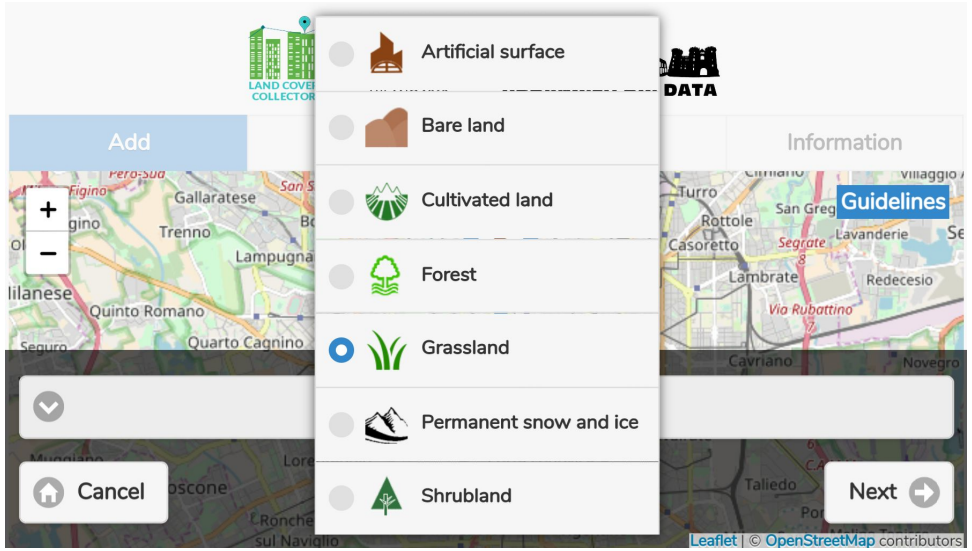


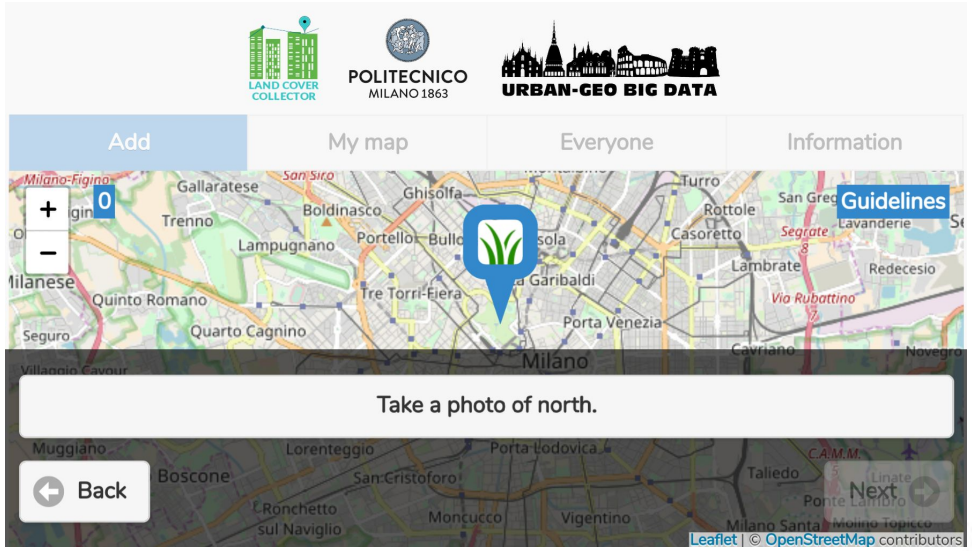
- Data collected using Land Cover Collector application can be visualized as “billboard” clusters on Cesium. Icons and cluster colors are adapted from the GL30. Each billboard can be queried. (Photos of north, east, south and west are to be added.)

Land Cover Collector

- An application to collect field data according to the land cover nomenclature of GL30.



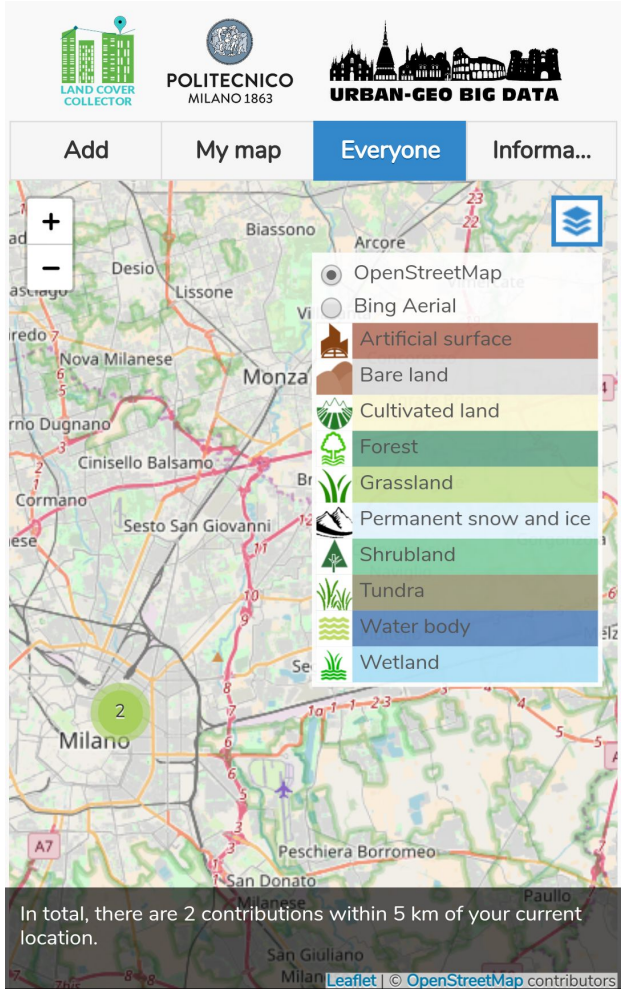




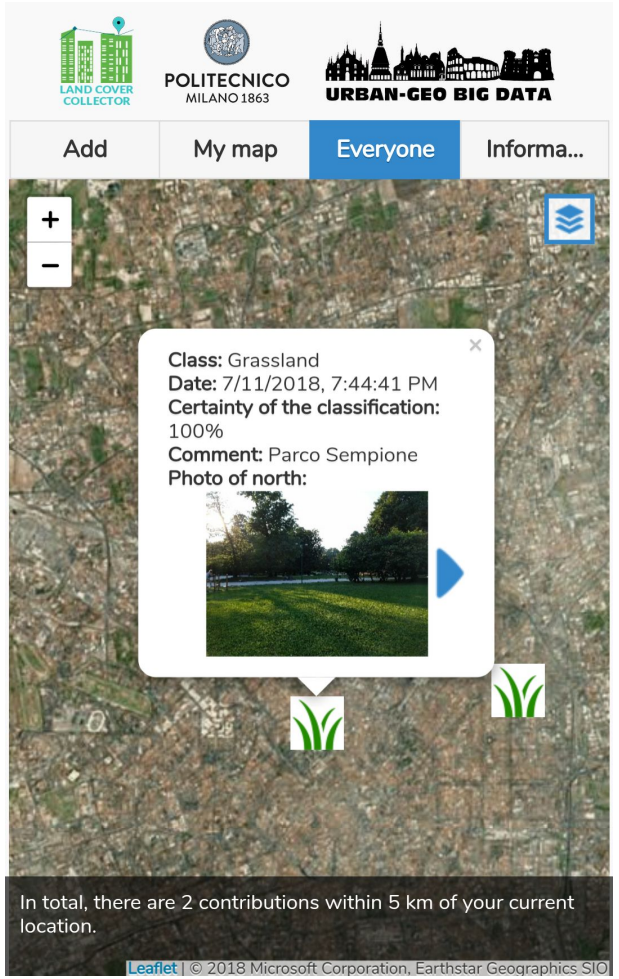
Land Cover Collector

- Its source code is available at https://github.com/kilsedar/land-cover-collector.
- It is available as a mobile application for Android and iOS as well as a web application at https://landcover.como.polimi.it/collector.
- Possible to collect data offline.
- The application is currently available in eight languages: English, Italian, Arabic, Russian, Chinese, Portuguese, French and Spanish.
- Collected data are released under the Open Database License and can be downloaded in JavaScript Object Notation (JSON) format.
- When collecting a land cover observation, in addition to enabling their GPS, users are required to indicate the land cover class, the degree of certainty; to take four photos in the north, east, south and west directions; and optionally to add a comment.
- Guidelines are provided.

Land Cover Collector



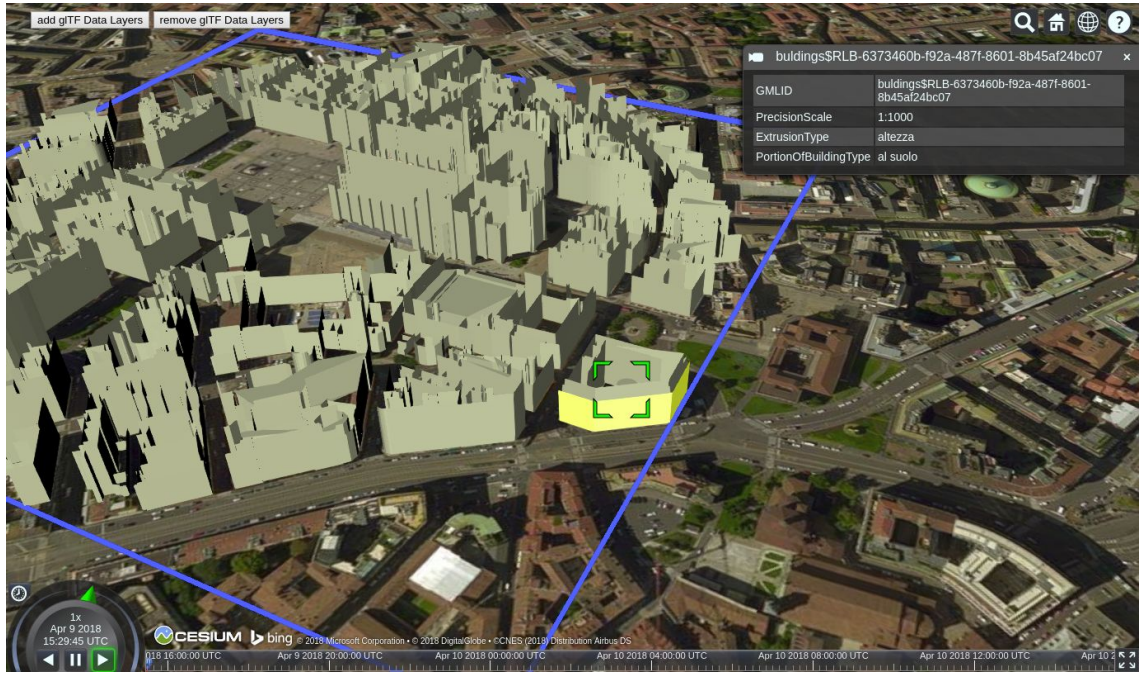
The collected land cover data can be visualized as clusters, aggregated based on the declared land cover class.



The details of each point can be queried by clicking it on the map.

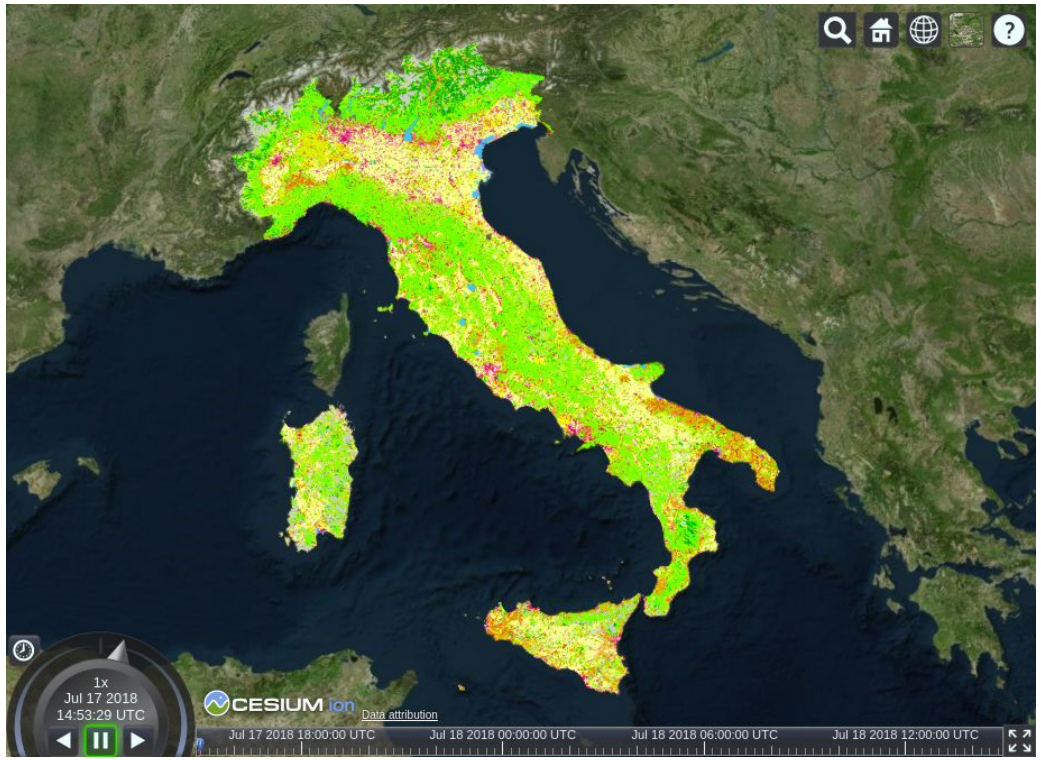
PRIN PROJECT: *URBAN GEOmatics for Bulk Information Generation, Data Assessment and Technology Awareness*

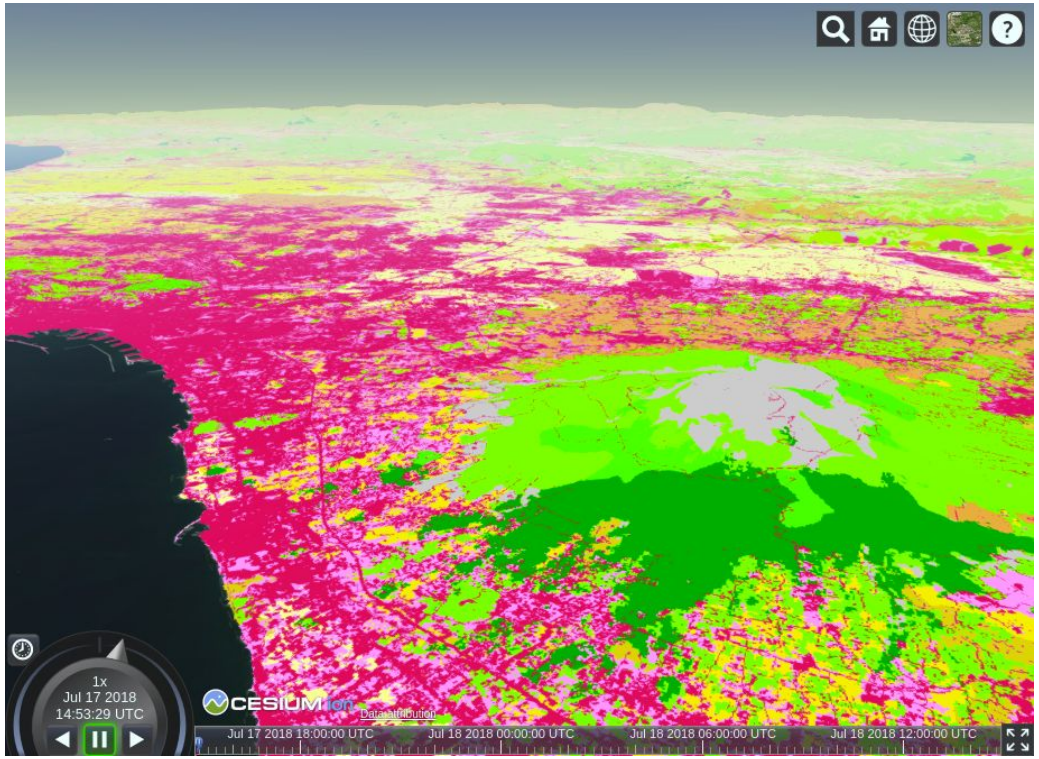
3D Buildings Visualization and Query (in progress)



- shapefile → CityGML (by University of Padova, Francesca Fissore and Francesco Pirotti)
- CityGML → KML/gltF → 3DCityDB-Web-Map-Client and Cesium virtual globe
- Recently generated CityGML data has to be integrated to the web client.

Big Raster Visualization (in progress)





10m resolution land cover map of Italy from ISPRA

- WMTS in GeoServer is used.
- Together with other datasets, has to be integrated to the web client.

Big Raster Visualization (in progress)

- Binary soil consumption data from ISPRA, 10m, [2012, 2015, 2016]
- Land cover data from ISPRA, 10m, [2012]
- Europe Settlement Map, 10m, [2016]
- Global Human Settlement Map, 40m, [1975, 1990, 2000, 2014]
- Global Human Footprint, 12m
- ...
- (Urban Atlas for 5 cities, vector, [2016, 2012])

- Visualizing this data with a time slider, placed under VGI land cover data to enable quick understanding of discrepancies.

WMTS with Time (to do)

- WMTS with Time is implemented in GeoServer.
- Click on Play button or move the Cursor to animate the imagery through years.
- To be implemented for the Naples Displacement Map. It may be implemented also for the binary soil consumption data from ISPRA or Global Human Settlement Map.

Processing Raster Data (to do)

- Using rasdaman and WCPS, for the listed LULC data, allow on the web interface to do the following operations:
  - click on a pixel, select a date, return the value for a certain dataset
  - click on a pixel, return the soil consumption/land cover change over time for a certain data set
  - for one of the cities, return the difference of soil consumption for two selected years
  - for one of the cities, set the color for a classification (rasdaman returns a new raster and client adds it on top of the client)
  - ...

URBAN-GEO BIG DATA

PRIN PROJECT: *URBAN GEOmatics for Bulk Information Generation, Data Assessment and Technology Awareness*

Thank You!

Maria Antonia Brovelli, Candan Eylül Kilsedar, Monia Elisa Molinari, Daniele Oxoli

Dept. of Civil and Environmental Engineering | GEOlab, Politecnico di Milano

for any question: maria.brovelli, candaneylul.kilsedar, moniaelisa. molinari, daniele.oxoli@polimi.it