URBAN-GEO BIG DATA

POLITECNICO MILANO 1863

istituto per il rilevamento elettromagnetico dell'ambiente

# URBAN GEO BIG DATA Meeting
# Padua, 24 January 2019


- CNR
- A. Pepe
  - M. Bonano, M. Manzo, Riccardo Lanari (InSAR results)
  - Fabiana Calò, Pasquale Imperatore, Antonio Pepe (Amplitude SAR Images Experiment)
  - Gloria Bordogna, Luca Frigerio (Data handling and representation)



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

MINISTERO DELL'ISTRUZIONE DELL'UNIVERSITA' E DELLA RICERCA

# Project Roadmap: IREA-CNR Contribution

- Historical ERS-ENVISAT Displacement Time-Series Products

- Napoli and Milan Time-Series are available. **At time T0+24 (Padua), time T0+30 (Turin),** **Time T0+36 (Rome)**

- Sentinel-1 DInSAR Time-Series on **one/some** selected cities

- Experimental Results on Change Detection in Urban Areas using Multi-Temporal SAR Images. Research Activity. **Depending on the achieved results**, they will be pubblished and presented.

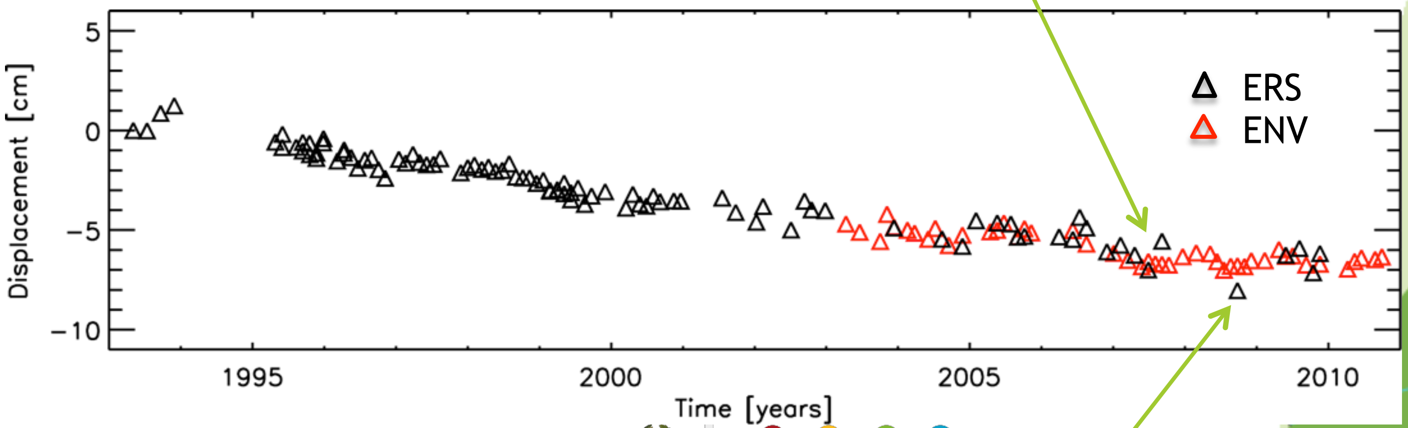
- Developing and publishing papers on the calibration of multi-polarization SAR images is in progress

# WP4.4 SAR and InSAR data generation

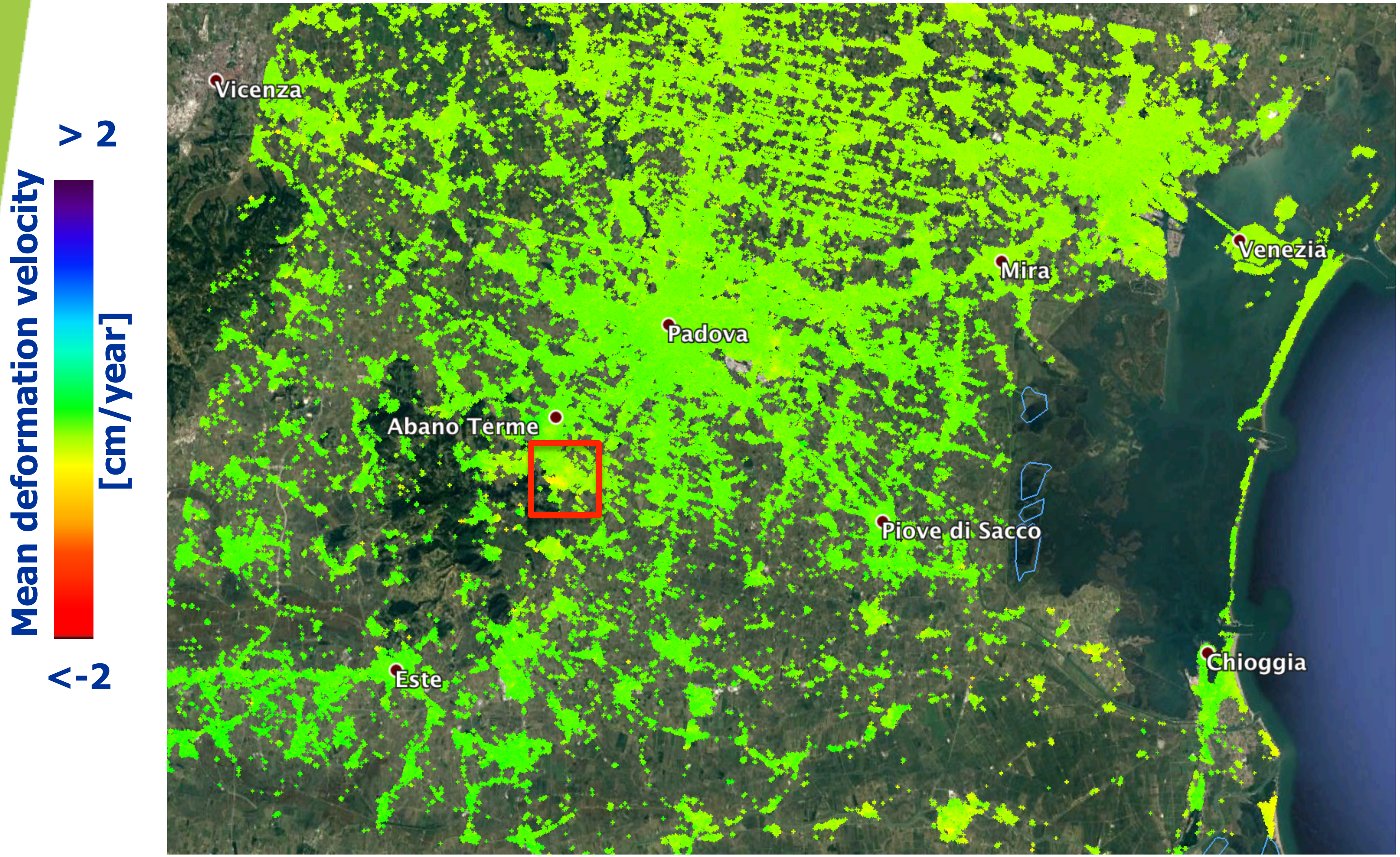
- Automatic generation of metadata for the correct ingestion, within the IREA-MI catalogue service (GET-IT), of the full resolution SBAS-DInSAR deformation time-series relevant to the Milan urban area (similarly to the Napoli case study)
- Generation of the low resolution SBAS-DInSAR deformation time-series relevant to the city of Padua

# WP4.4 Critical Issues

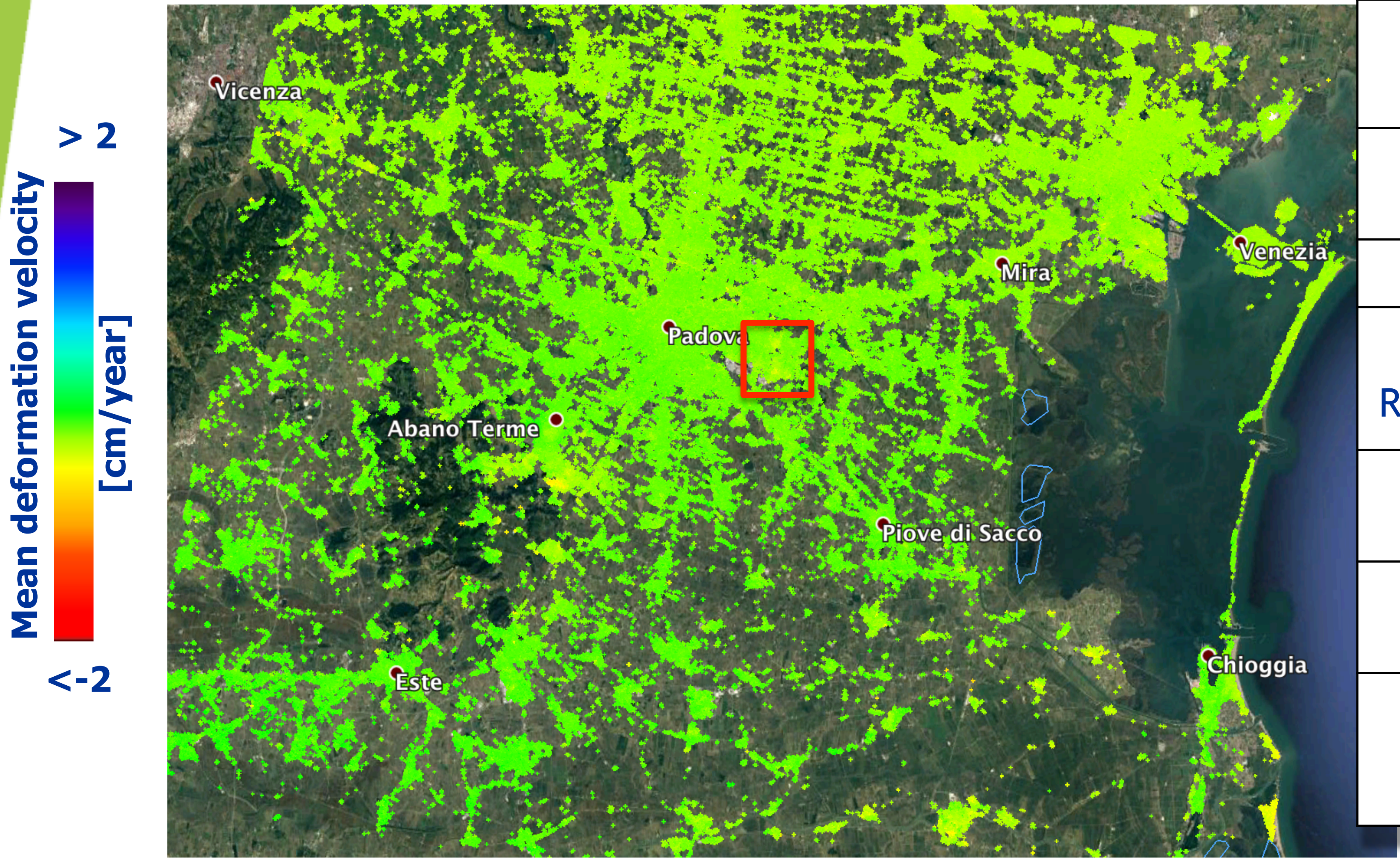
- Generation of LOS displacement time-series visualization
- Problems and critical issues
- Provided LOS displacement time-series are related to single pixel-wise targets on the ground
- Accuracy of the LOS time-series is on the order of 3-4 mm (on average)
- Accuracy of mean displacement time-series is of about 1 mm (or less) for linear deformations
- Deformation can be significantly non-linear
- Visualization problems and solutions adopted



# WP4.4 Low resolution SBAS-DInSAR results: the Montegrotto Terme area



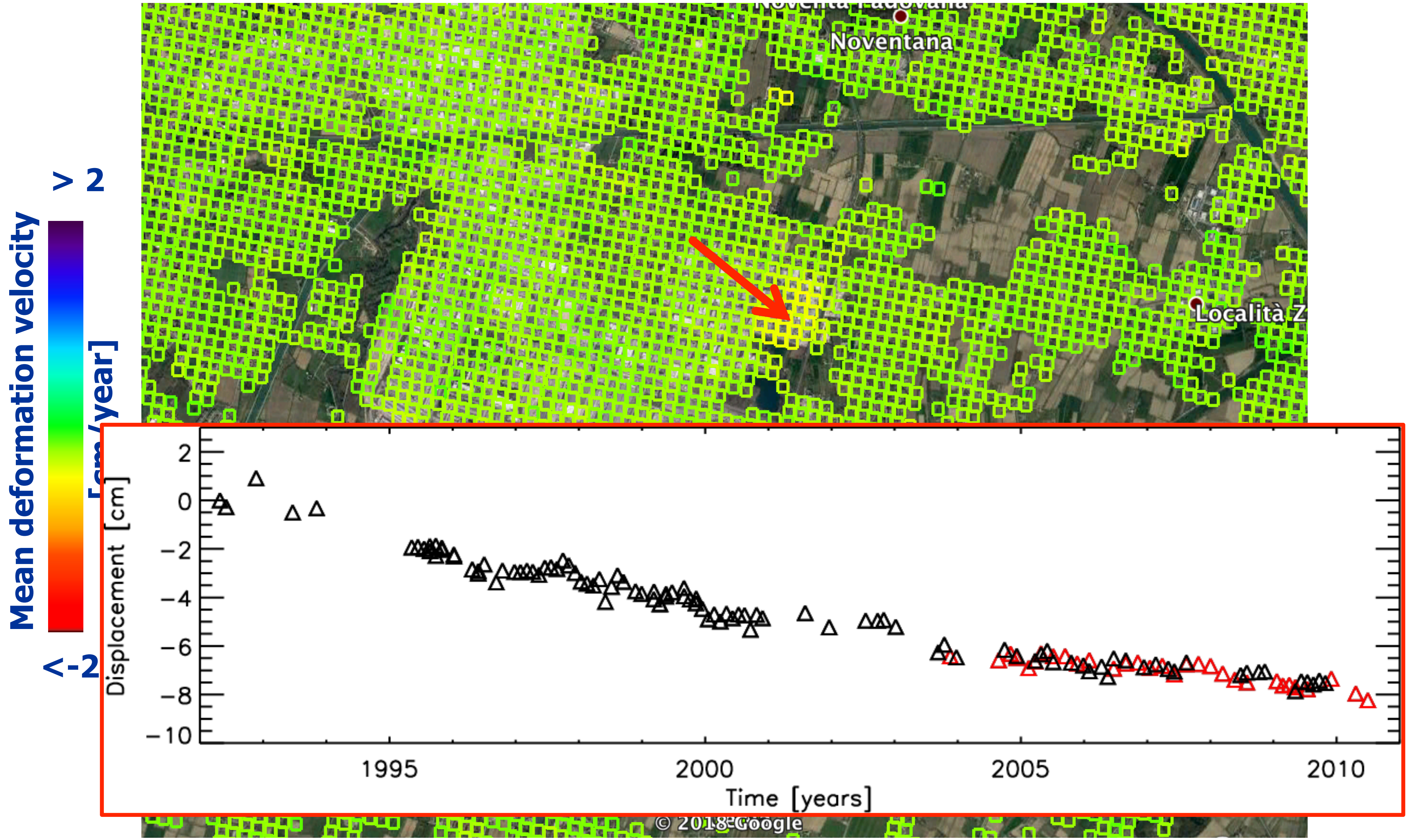
# WP4.4 Low resolution SBAS-DInSAR results: the Padua urban area



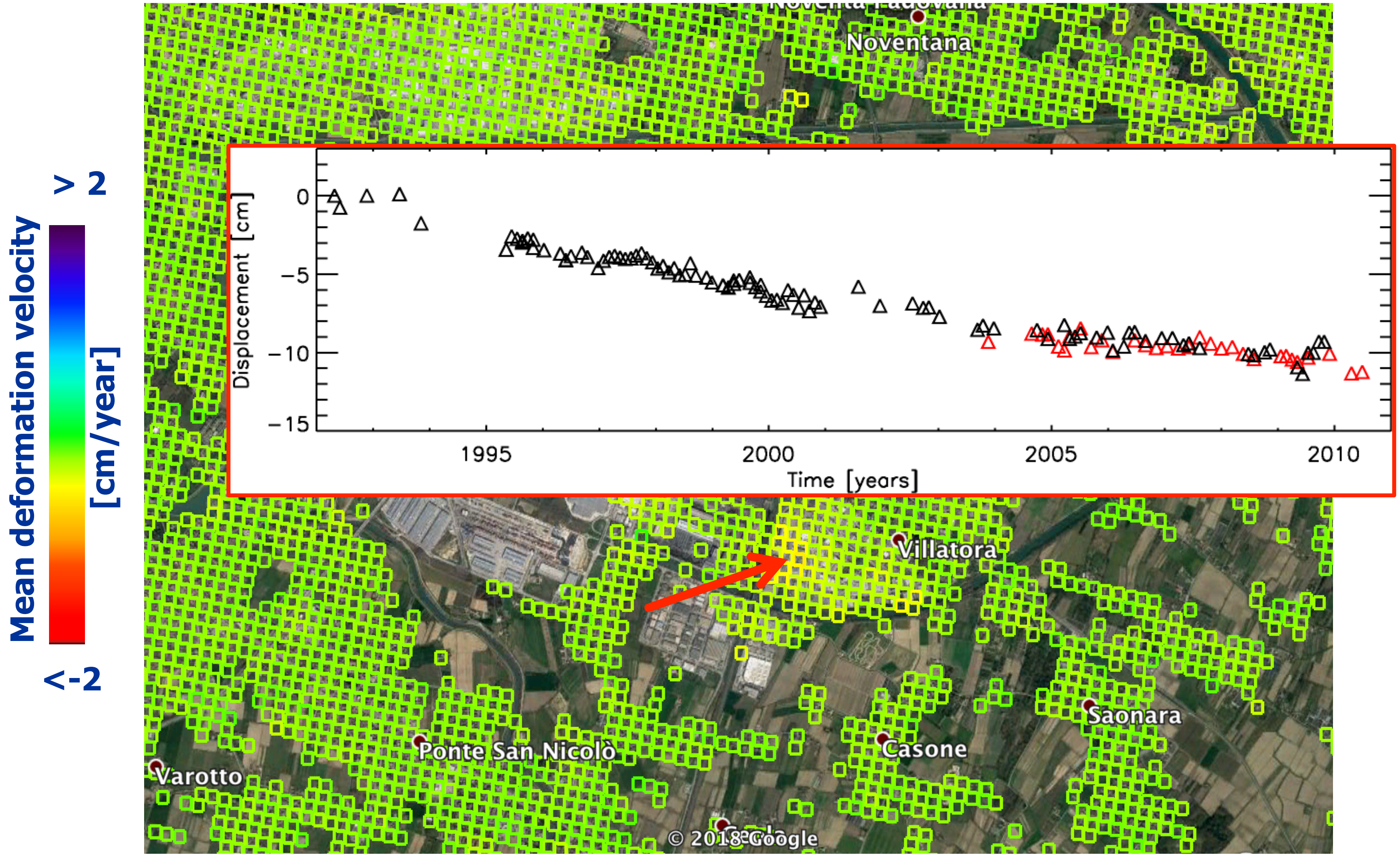

| Padua area (T 394, descending orbits) | |
| --- | --- |
| #Images | 134 (103 ERS,31 ENV) |
| Sensor | ERS/ENVISAT |
| Spatial Resolution [m] | 80x80 |
| Time span | 24/04/1992 – 28/06/2010 |
| #Pixels | 37600x 4800 ( 180 M) |
| #Interf | 395 (311 ERS,84 ENV) |

IT LOOKS STABLE BUT !!!!

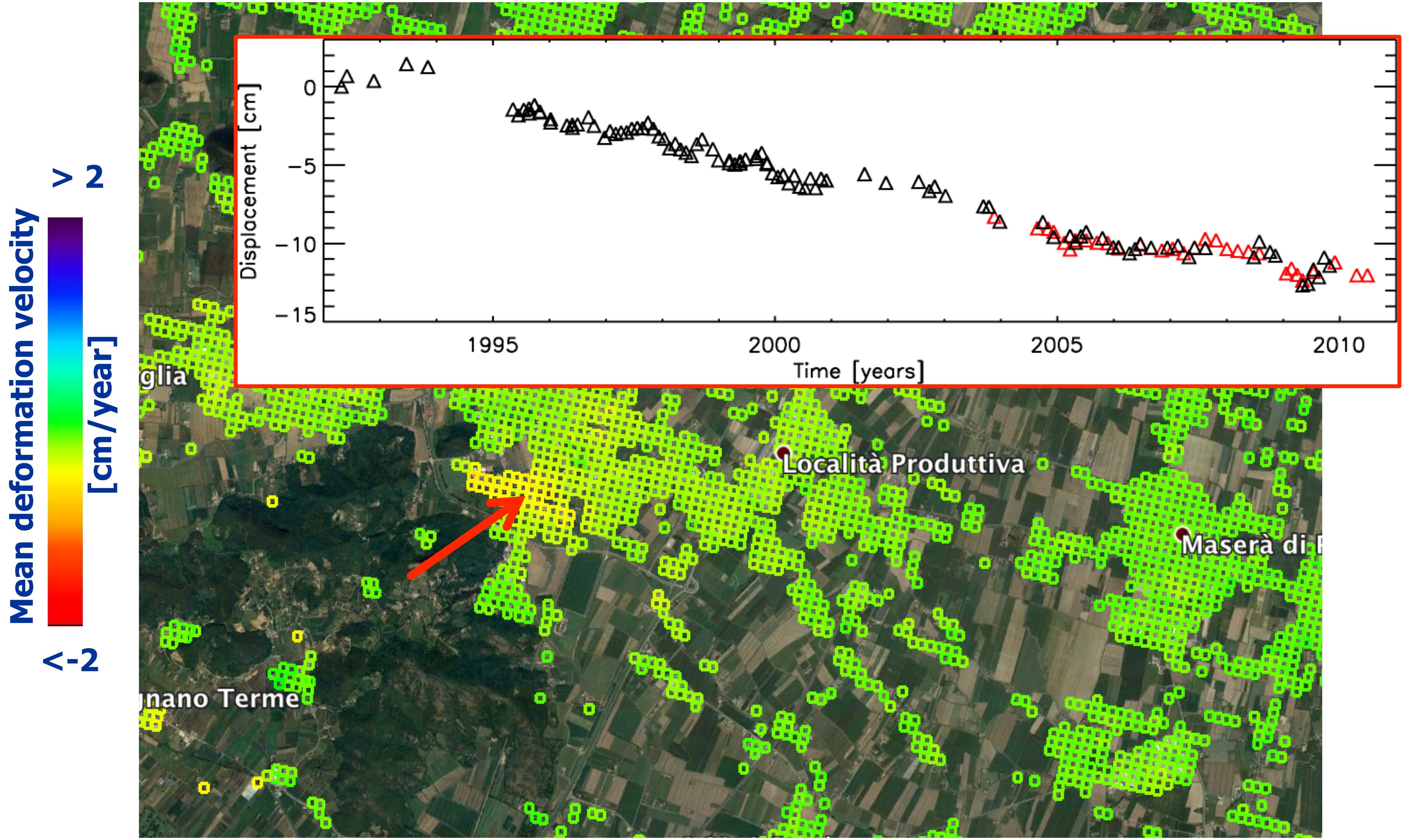
# WP4.4 Low resolution SBAS-DInSAR results: the Padua urban area



# WP4.4 Low resolution SBAS-DInSAR results: the Padua urban area



# WP4.4 Low resolution SBAS-DInSAR results: the Montegrotto Terme area



# Research products

**Published papers**

- A. Pepe, M. Bonano, G.Bordogna, M. Brovelli, F.Calò, P. Carrara, L. Congedo, L. Frigerio, P. Imperatore, R. Lanari, S. Lanucara, M. Manzo, M.Munafò The “Urban Geomatics for Bulk Information Generation, Data Assessment and Technology Awareness” Project: Detection, Representation and Analysis of the Urban Scenario Changes, short paper in proceeding of GARSS 2018 - 2018 IEEE Int. Geoscience and Remote Sensing Symposium, 2902-2905, Valencia, Spain.

- Geoinformatics and Citizen Science, Gloria Bordogna, Special issue of the ISPRS Int. J. of Geoinformation, 2018: http://www.mdpi.com/journal/ijgi/special_issues/citizen-science (10 papers included) Prefazione https://www.mdpi.com/2220-9964/7/12/474

**Organization activities**

- Workshop on Volunteered Geographic Information: Enabling VGI creation, management and sharing : AREA della Ricerca CNR Via Alfonso Corti 12 , 20133 MILANO April the 16th, 2018.

URBAN-GEO BIG DATA

# IREA Unit

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

MINISTERO DELL'ISTRUZIONE DELL'UNIVERSITA' E DELLA RICERCA

# Gantt

- WP3 architecture design and implementation : server side   due month 30

- Wp4.4: SAR and InSAR Data   - due month 24 : (delayed) March 2019 Padua, Turine and Roma T30.

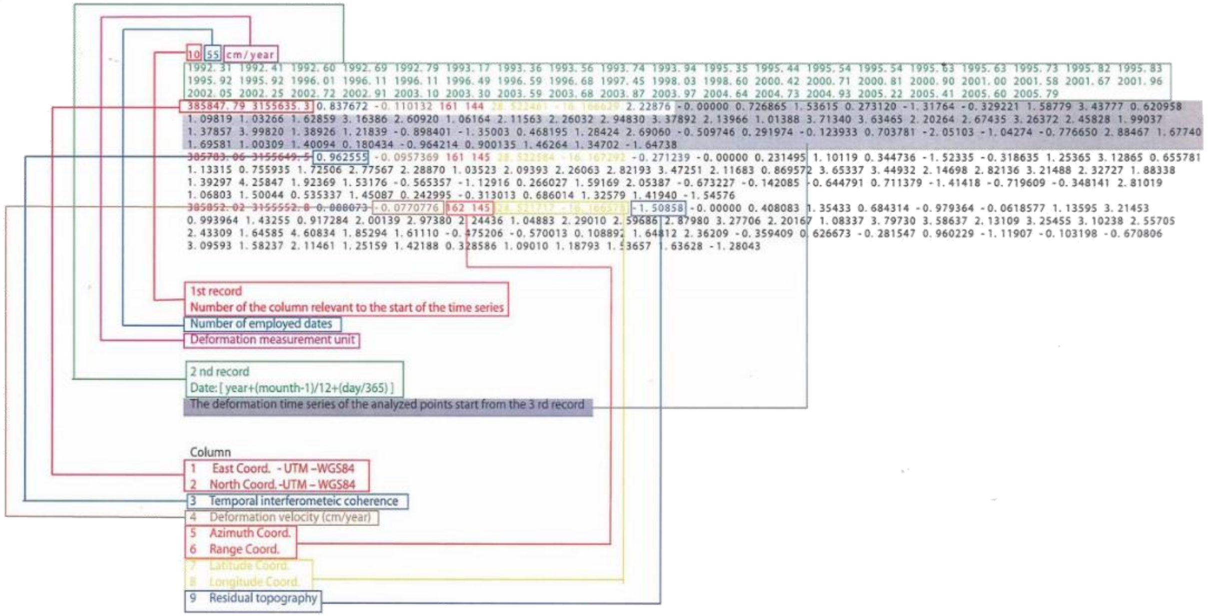
- Wp4.8: Data selection through quality evaluation due montyh 20 (delayed)

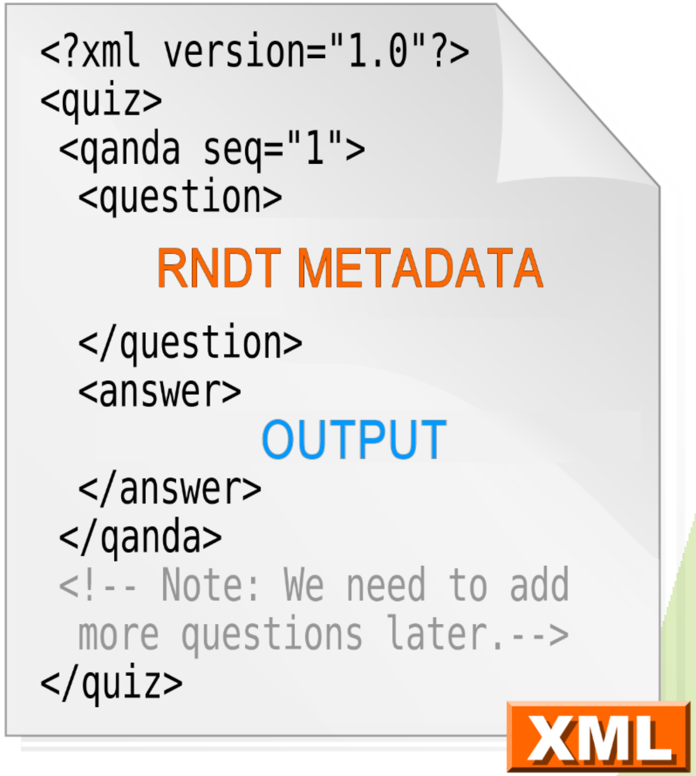
# WP3 and WP4.8 metadata generation and data and metadata publication

- Automatic generation of metadata from SAR and InSar data
- Automatic publication of metadata
- Automatic generation of WMS WFS and WMTS services from SAR and InSAr data on the IREA urban geo big data node
- Catalogue service inplementation

# Pilot experiment of AUTOMATIC RNDT METADATA CREATION from Data Header of Deformation maps

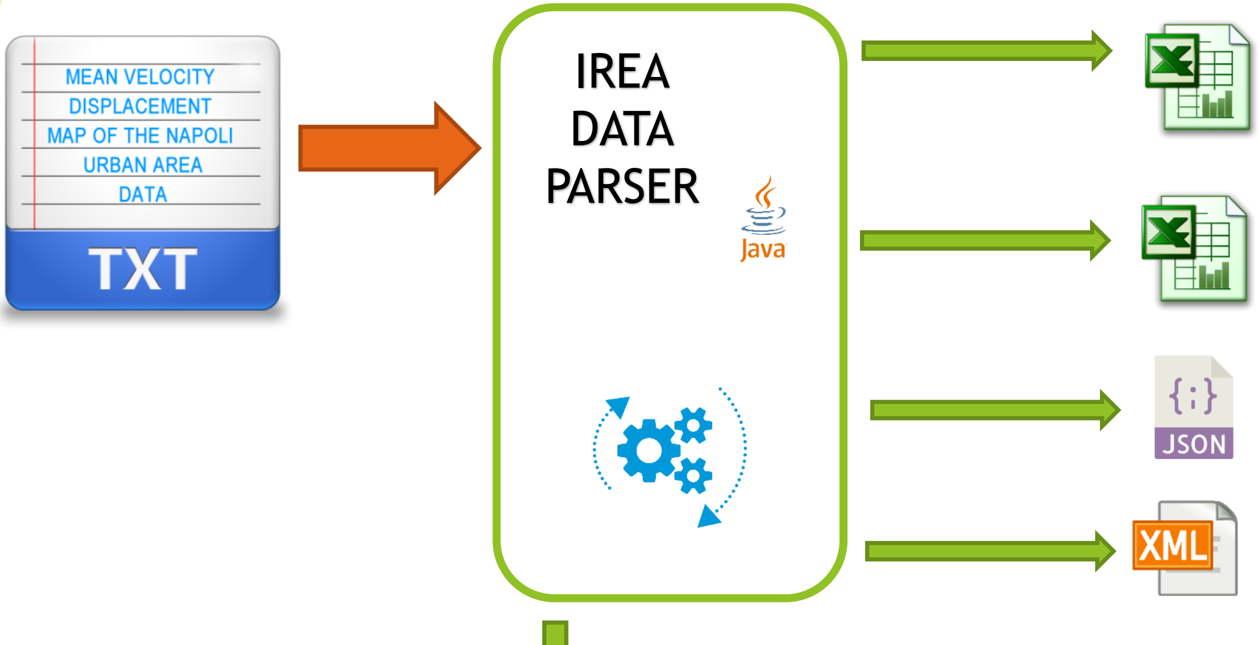
This is a small “sample” of the input data from which we generate RNDT metadata



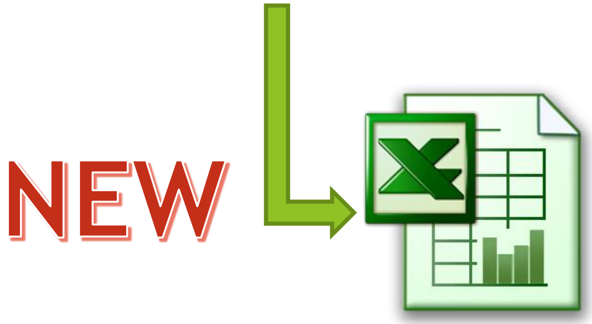




**Static metadata fields for deformation maps:**
Description, abstract, keywords, creator, etc,.

# WP3.2 GET IT node for IREA data node

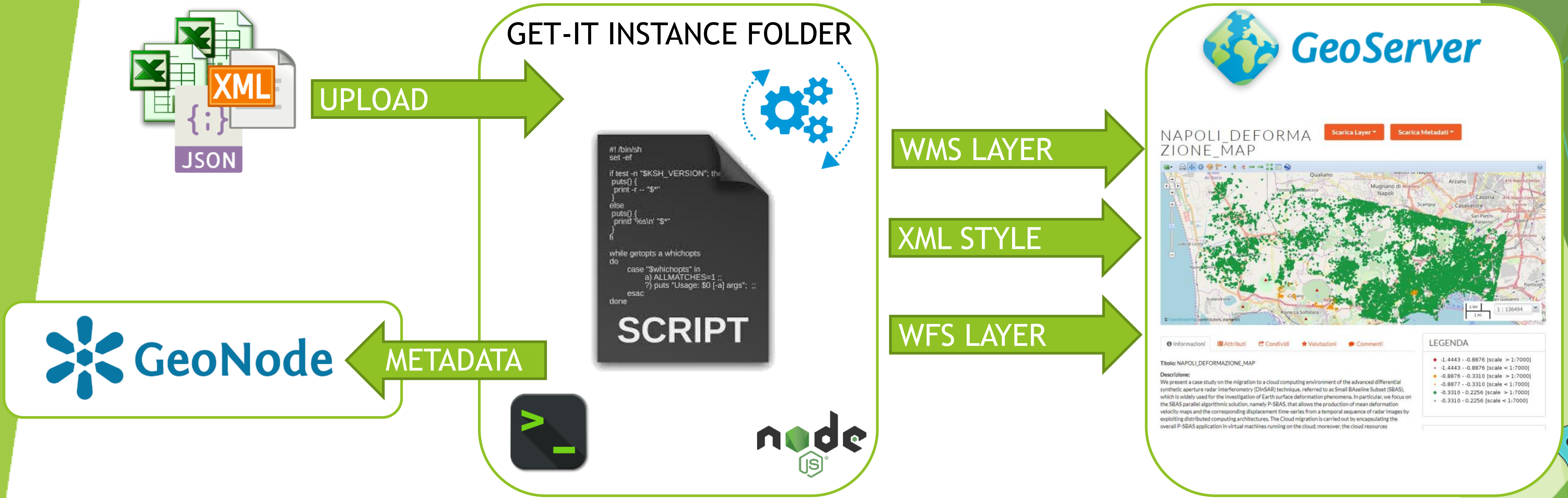



1. A csv file that contains all the information about the points described in the Napoli data related to the calculated average value. This file is needed in order to create a layer.
2. A massive csv file that contains the series values. This file is needed to configure the WFS server in order to retrieve series values related to a single point.
3. A simple json file that contains few metadata fields extracted from the original file.
4. An xml file that contains the style to visualize the map. For example, the color of the point based on the average value and the size of the point based on the map zoom.



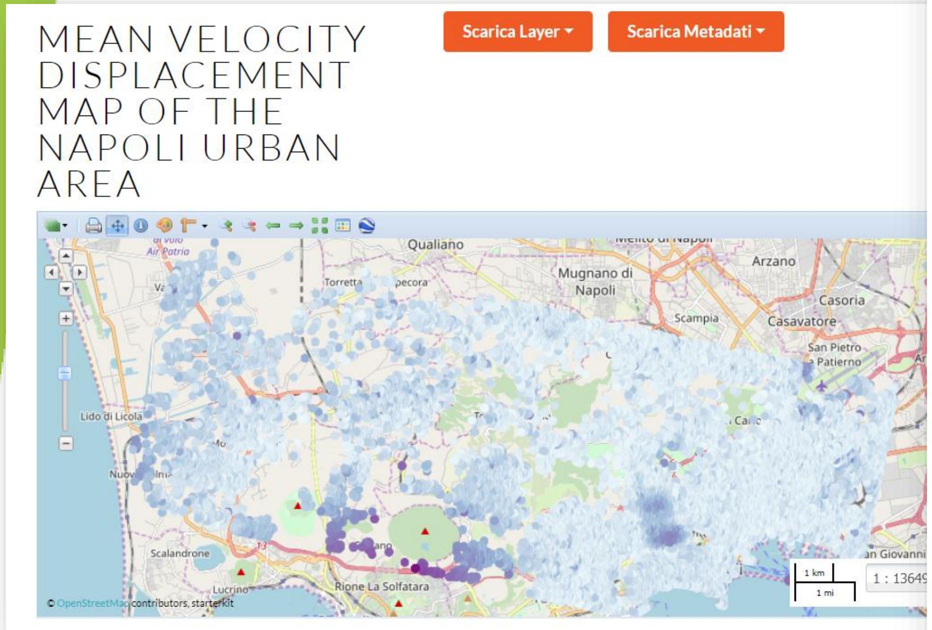

5. An xml file that contains all the information needed to create the GeoJson files.

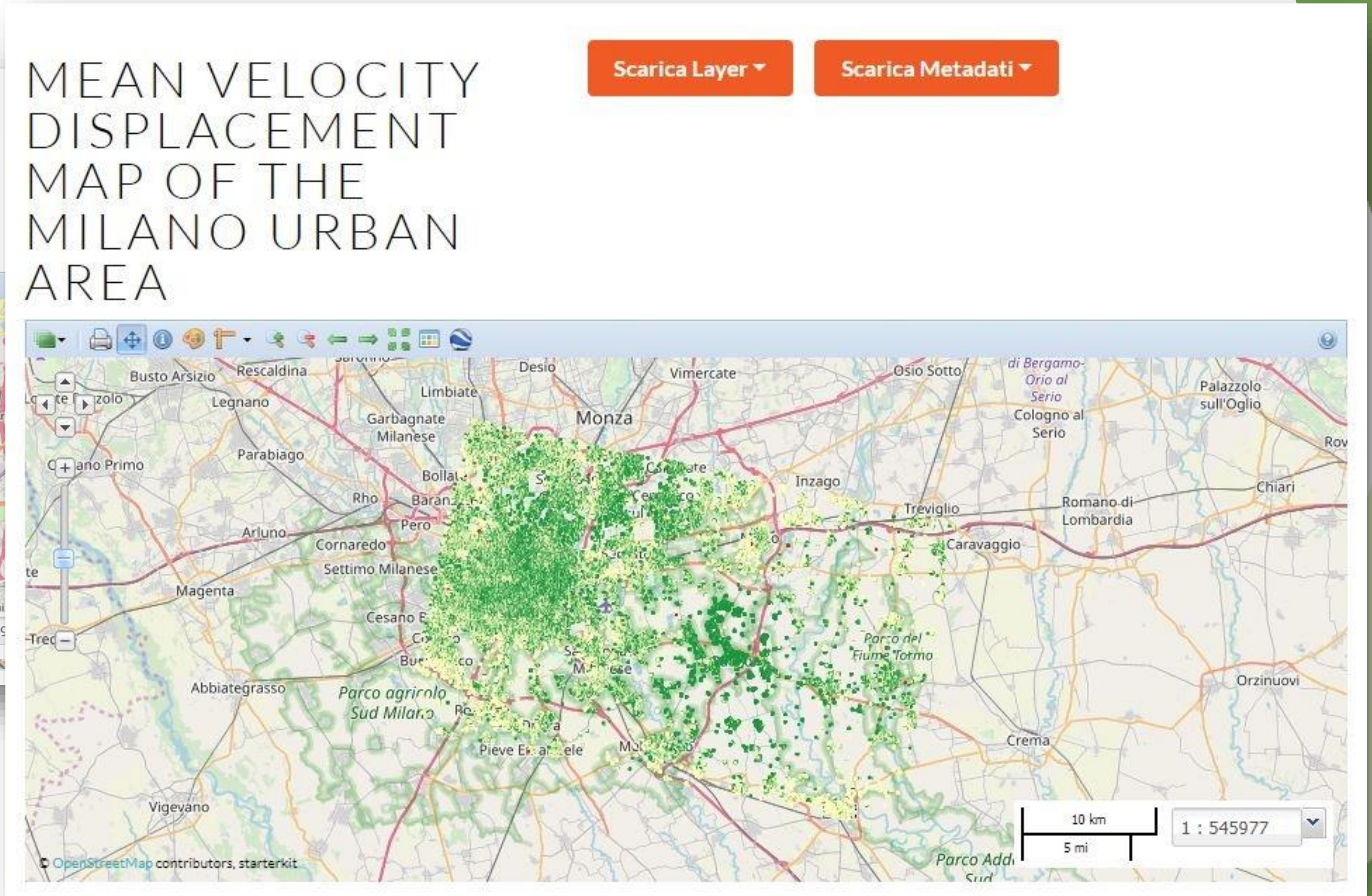
# WP3.2 GET IT node for IREA data node




All the layers https://ugbd.get-it.it/

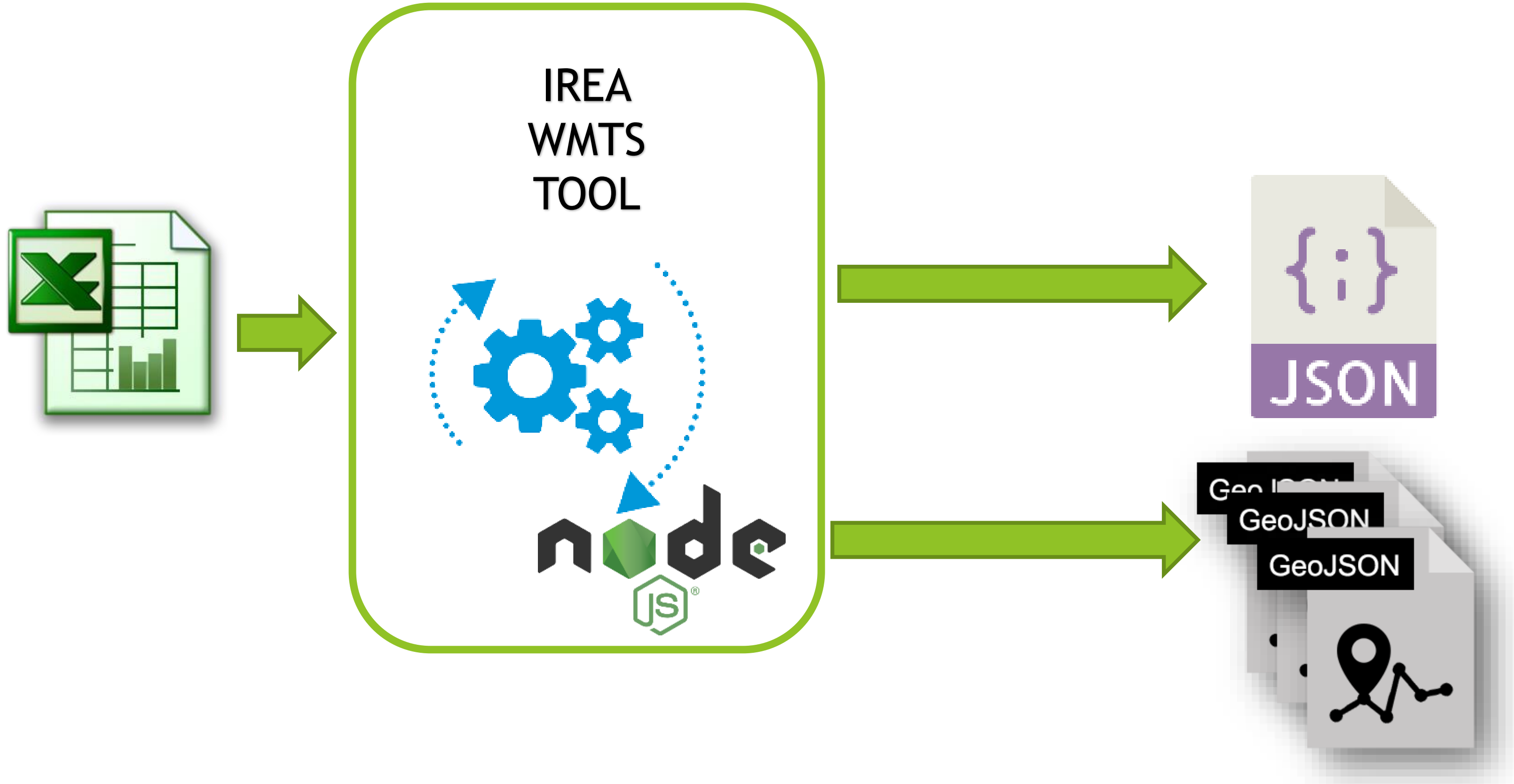
# Milano and Napoli Displacement maps now avaiable




All the layers https://ugbd.get-it.it/

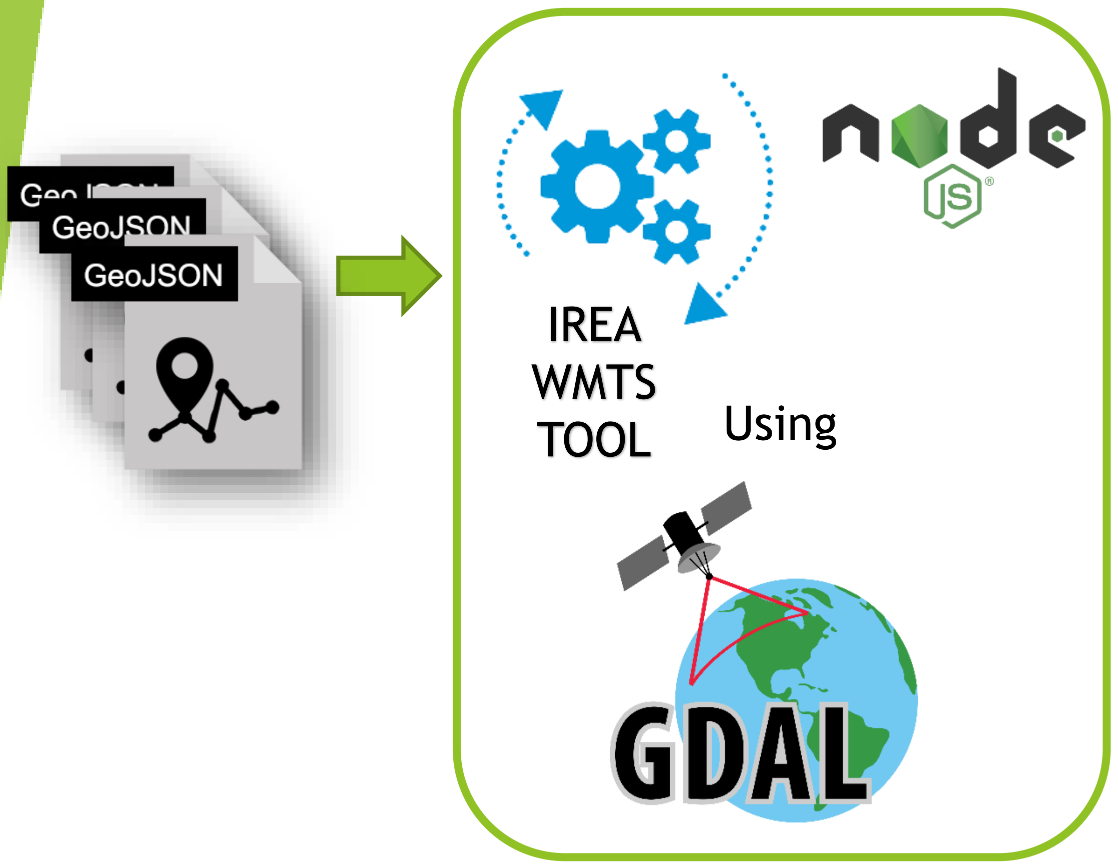
# WMTS multidimensional : first step




DATA ANALYSIS
- Max, Min values
- Distribution

GEO JSON CREATOR
It creates a GeoJson
for each interval of values
for each year

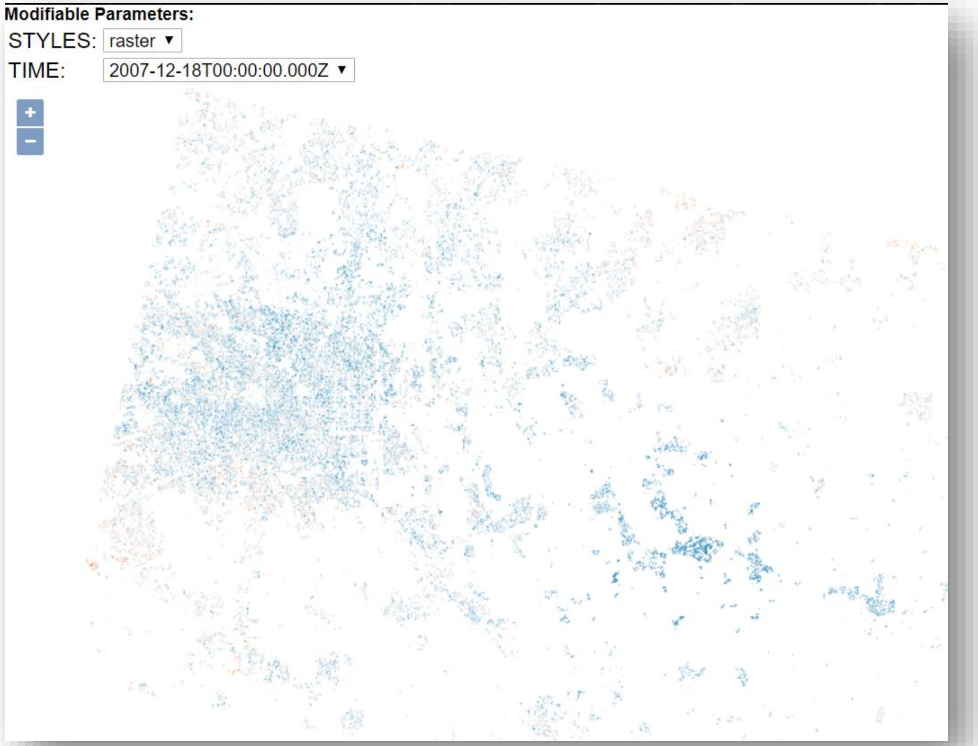
# WMST multidimensional : step two




GEO TIFF CREATOR
It creates a GeoTIFF for each year


http://ugbd-geoserver.get-it.it

# Milano and Napoli WMTS service


MILANO : http://ugbd-geoserver.get-it.it/geoserver/gwc/demo/DeformationTS:milanowmts?gridSet=EPSG:900913&format=image/png

NAPOLI : http://ugbd-geoserver.get-it.it/geoserver/gwc/demo/DeformationTS:imagenapolirealcolors?gridSet=EPSG:900913&format=image/png

# The Proxy

1. ADD C.O.R.S. TO QUERY AND IMAGES REQUESTS
2. HIDE THE SECOND GEOSERVER USED FOR THE WMTS

REDIRECTS HTTPS REQUESTO TO HTTP ONLY SERVER

EDI CLIENT

SSL TO NOT SSL PROBLEM SPARQL QUERIES

# Research products

**Published papers**

- A. Pepe, M. Bonano, G.Bordogna, M. Brovelli, F.Calò, P. Carrara, L. Congedo, L. Frigerio, P. Imperatore, R. Lanari, S. Lanucara, M. Manzo, M.Munafò The "Urban Geomatics for Bulk Information Generation, Data Assessment and Technology Awareness" Project: Detection, Representation and Analysis of the Urban Scenario Changes, short paper in proceeding of GARSS 2018 - 2018 IEEE Int. Geoscience and Remote Sensing Symposium, 2902-2905, Valencia, Spain.

- Geoinformatics and Citizen Science, Gloria Bordogna, Special issue of the ISPRS Int. J. of Geoinformation, 2018: http://www.mdpi.com/journal/ijgi/special_issues/citizen-science (10 papers included) Prefazione https://www.mdpi.com/2220-9964/7/12/474

**Organization activities**

- Workshop on Volunteered Geographic Information: Enabling VGI creation, management and sharing : AREA della Ricerca CNR Via Alfonso Corti 12 , 20133 MILANO April the 16th, 2018.

URBAN-GEO BIG DATA

POLITECNICO MILANO 1863

ISPRA
Istituto Superiore per la Protezione e la Ricerca Ambientale

# URBAN GEOmatics for Bulk Information Generation, Data Assessment and Technology Awareness

- Contribution by ISPRA – 24/01/2019

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

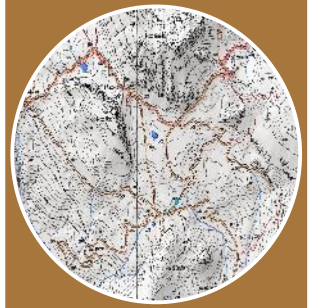
MINISTERO DELL'ISTRUZIONE DELL'UNIVERSITA' E DELLA RICERCA

# Available data

Satellite Images

Very High resolution Images

Commercial and in situ images



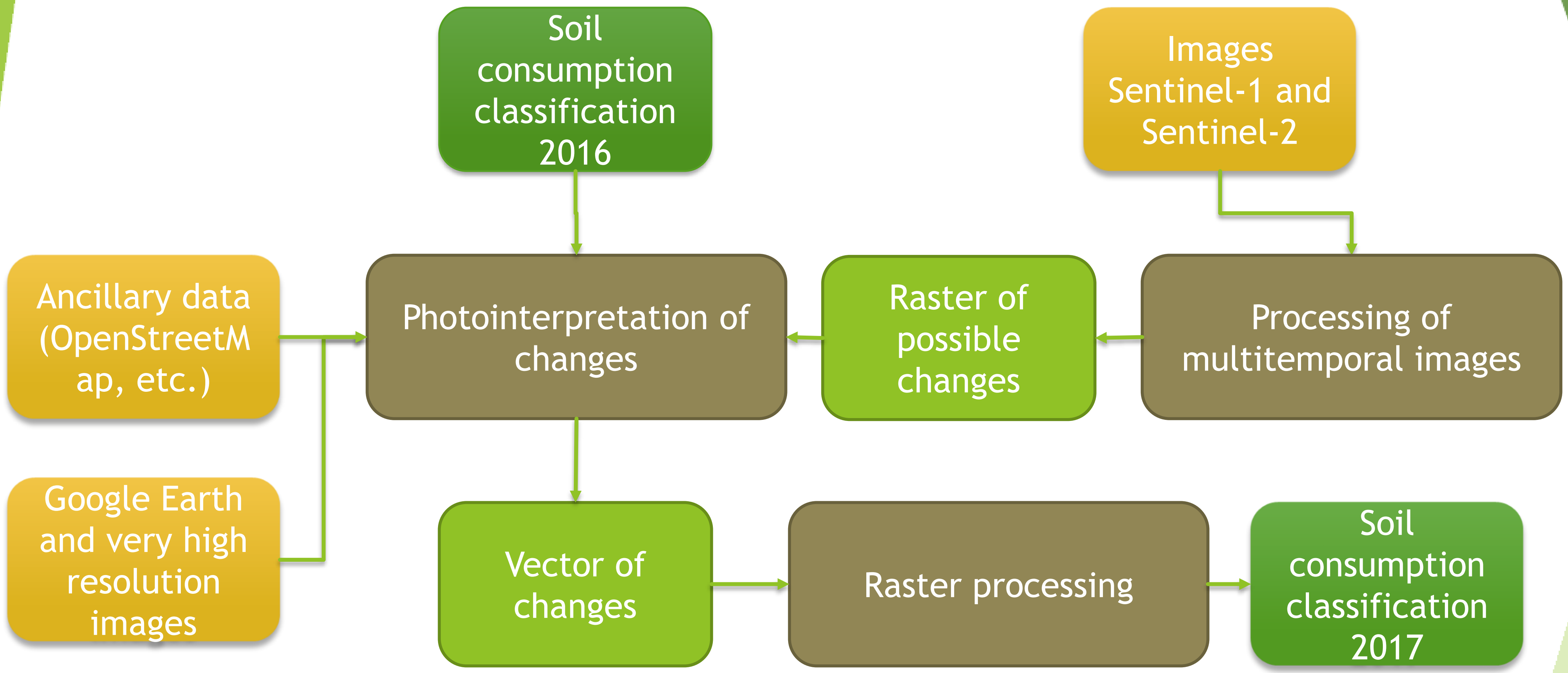
Topographic data

Thematic data from regional EIS

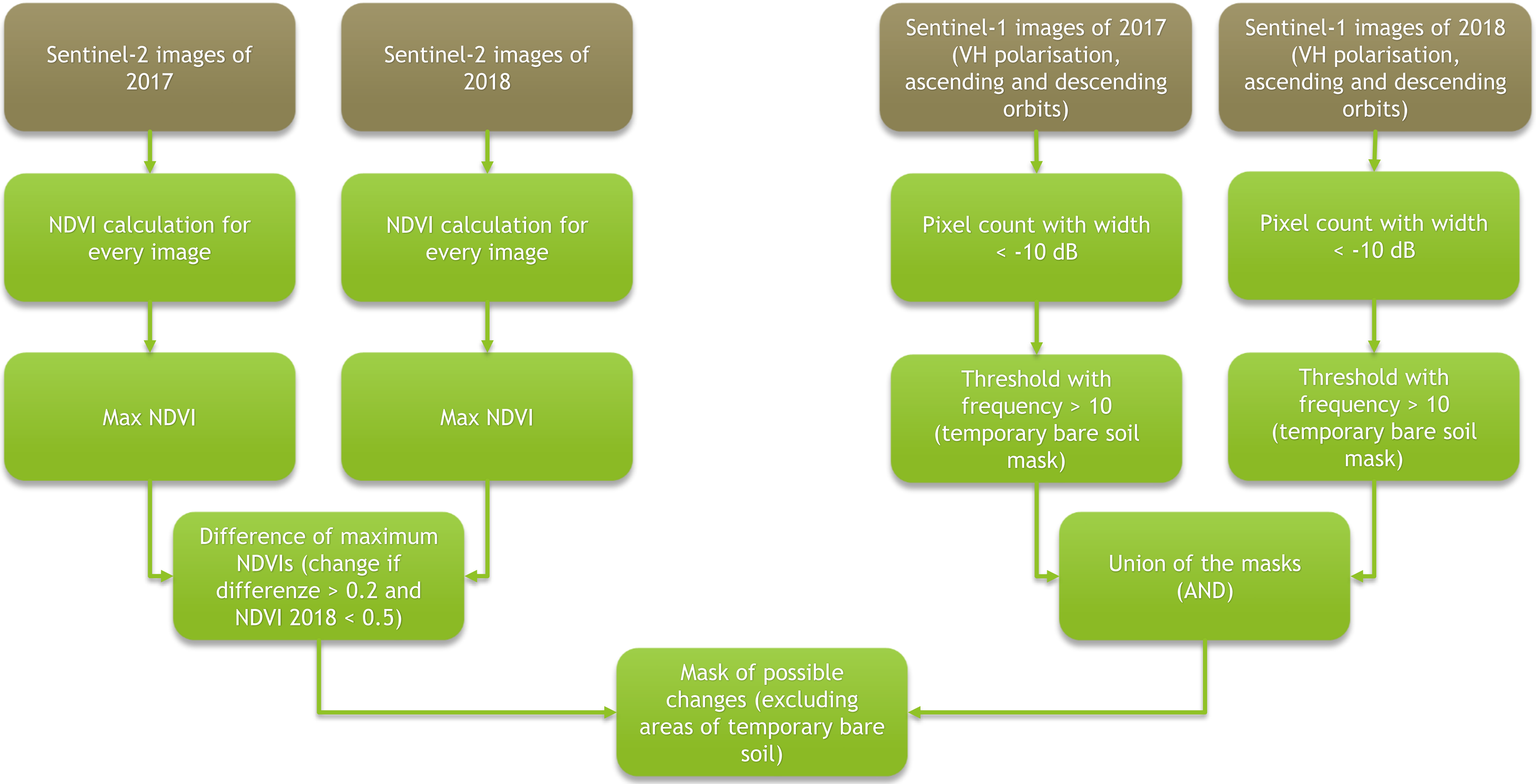
OSM data

**Land Consumption Result Map 2017**

# Processing and update of soil consumption map



# Processing and update of soil consumption map



# Change Map: Integration results

Real Change

Omission

Commission

# Change Map: Integration results

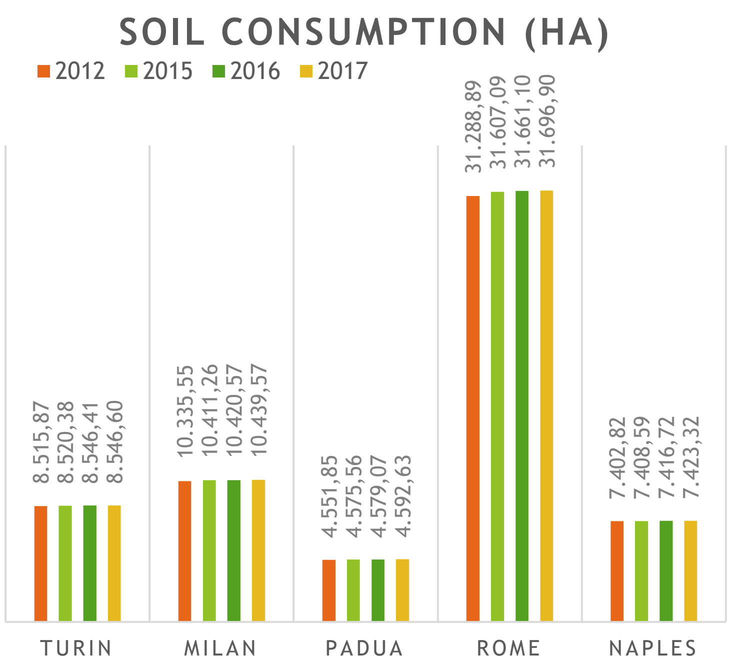
Integration of Multispectral and SAR data for change detection.

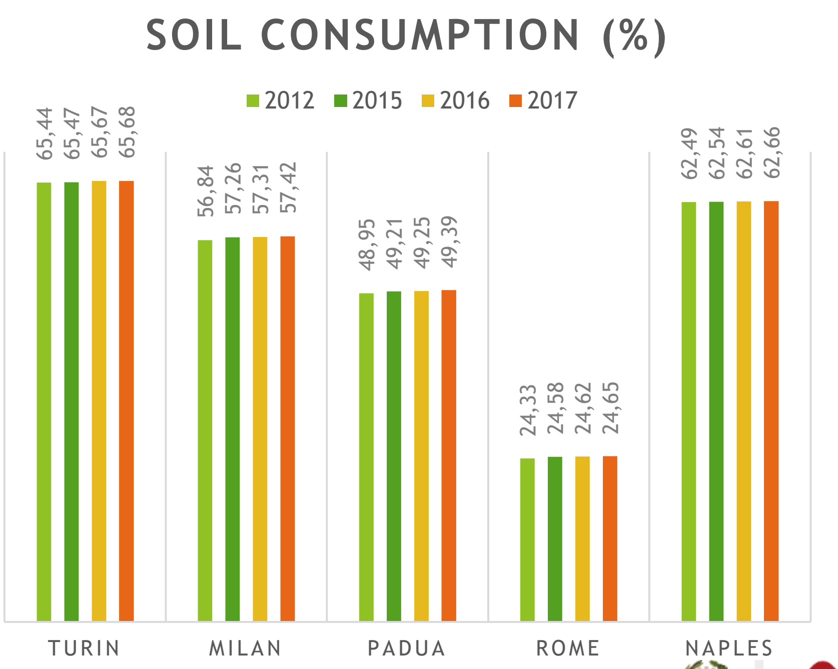

Comparison between mask and photointerpreted changes

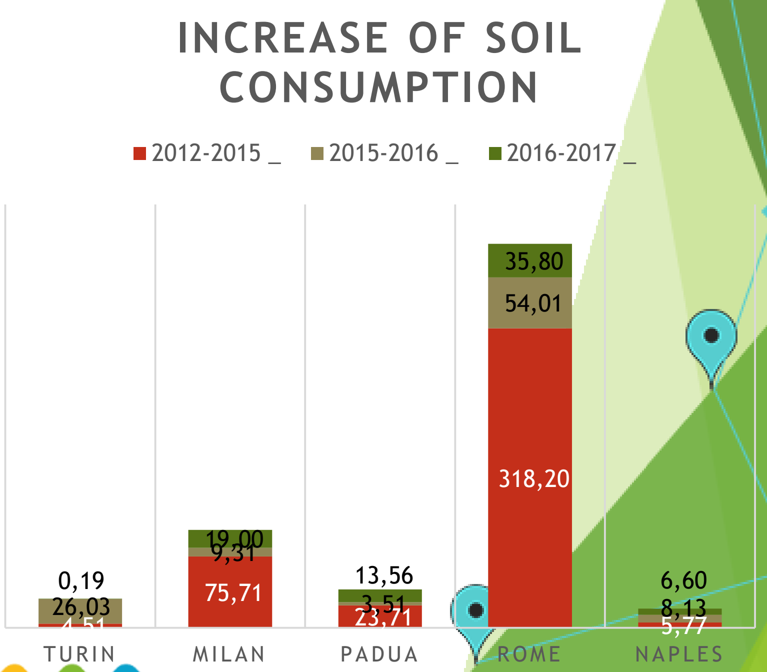
# Land cover classification results

- Evolution of soil consumption from 2012 to 2017

| Area | Consumed soil 2012 [ha] | Not consumed soil 2012 [ha] | Consumed soil 2012 [%] | Not consumed soil 2012 [%] | Consumed soil 2015 [ha] | Not consumed soil 2015 [ha] | Consumed soil 2015 [%] | Non consumed soil 2015 [%] | Consumed soil 2016 [ha] | Non consumed soil 2016 [ha] | Consumed soil 2016 [%] | Not consumed soil 2016 [%] | Consumed soil 2017 [ha] | Non consumed soil 2017 [ha] | Consumed soil 2017 [%] | Not consumed soil 2017 [%] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Turin | 8.515,87 | 4.497,38 | 65,44 | 34,56 | 8.520,38 | 4.492,87 | 65,47 | 34,53 | 8.546,41 | 4.466,84 | 65,67 | 34,33 | 8.546,60 | 4.466,65 | 65,68 | 34,32 |
| Milan | 10.335,55 | 7.846,79 | 56,84 | 43,16 | 10.411,26 | 7.771,08 | 57,26 | 42,74 | 10.420,57 | 7.761,77 | 57,31 | 42,69 | 10.439,57 | 7.742,77 | 57,42 | 42,58 |
| Padua | 4.551,85 | 4.746,64 | 48,95 | 51,05 | 4.575,56 | 4.722,93 | 49,21 | 50,79 | 4.579,07 | 4.719,42 | 49,25 | 50,75 | 4.592,63 | 4.705,86 | 49,39 | 50,61 |
| Rome | 31.288,89 | 97.294,65 | 24,33 | 75,67 | 31.607,09 | 96.976,45 | 24,58 | 75,42 | 31.661,10 | 96.922,44 | 24,62 | 75,38 | 31.696,90 | 96.886,64 | 24,65 | 75,35 |
| Naples | 7.402,82 | 4.443,36 | 62,49 | 37,51 | 7.408,59 | 4.437,59 | 62,54 | 37,46 | 7.416,72 | 4.429,46 | 62,61 | 37,39 | 7.423,32 | 4.422,86 | 62,66 | 37,34 |




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

# Landscape metrics

POP: Population (ISTAT)

Edclass: Edge density [m/ha]

DENSITY: Population density [ab/ha]

LCPI: Largest Class Patch Index in %

RMPS: Residual Mean Patch Size [ha]

Dispersion_index: Ratio of low density urban area on urban area in %

| Area | POP_2012 | POP_2015 | POP_2016 |
|---|---|---|---|
| Turin | 869.312 | 896.773 | 890.529 |
| Milan | 1.240.173 | 1.337.155 | 1.345.851 |
| Padua | 205.631 | 211.210 | 210.401 |
| Rome | 2.614.263 | 2.872.021 | 2.864.731 |
| Naples | 961.106 | 978.399 | 974.074 |

| Area | Edclass_2015 | Edclass_2016 |
|---|---|---|
| Turin | 182,85 | 181,95 |
| Milan | 467,10 | 466,84 |
| Padua | 502,17 | 501,94 |
| Rome | 759,22 | 758,64 |
| Naples | 366,71 | 366,50 |

| Area | DENSITY_2012 | DENSITY_2015 | DENSITY_2016 |
|---|---|---|---|
| Turin | 66,83 | 68,94 | 68,46 |
| Milan | 68,21 | 73,54 | 74,02 |
| Padua | 22,11 | 22,71 | 22,63 |
| Rome | 20,33 | 22,34 | 22,28 |
| Naples | 81,13 | 82,59 | 82,23 |

| Area | LCPI_2012 | LCPI_2015 | LCPI_2016 |
|---|---|---|---|
| Turin | - | 52,76 | 52,97 |
| Milan | - | 57,38 | 57,39 |
| Padua | - | 39,79 | 39,82 |
| Rome | - | 11,92 | 11,94 |
| Naples | - | 57,67 | 57,73 |

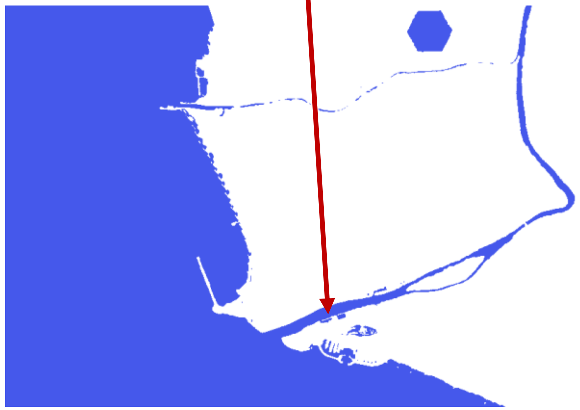
| Area | RMPS_2015 | RMPS_2016 |
|---|---|---|
| Turin | 15,31 | 15,40 |
| Milan | 12,75 | 12,90 |
| Padua | 7,82 | 7,80 |
| Rome | 7,65 | 7,68 |
| Naples | 10,33 | 10,50 |

| Area | Dispersion_index_2015 | Dispersion_index_2012 | Dispersion_index_2016 |
|---|---|---|---|
| Turin | 28,42 | 28,46 | 28,24 |
| Milan | 31,43 | 31,87 | 31,43 |
| Padua | 51,36 | 51,60 | 51,25 |
| Rome | 64,66 | 65,19 | 64,58 |
| Naples | 30,99 | 31,01 | 30,91 |

# Land cover classification

**EAGLE Classification system**

- Water and Snow
- Grass
- Forest and trees
- Bare soil (abiotic natural)
- Urban (abiotic artificial )



Multitemporal Image selection

Sentinel-1

Sentinel-2

Pre-processing

Decision Tree

Sentinel-1 Multitemporal Threshold setting on backscatter

Sentinel-2 Multitemporal Threshold setting on indexes

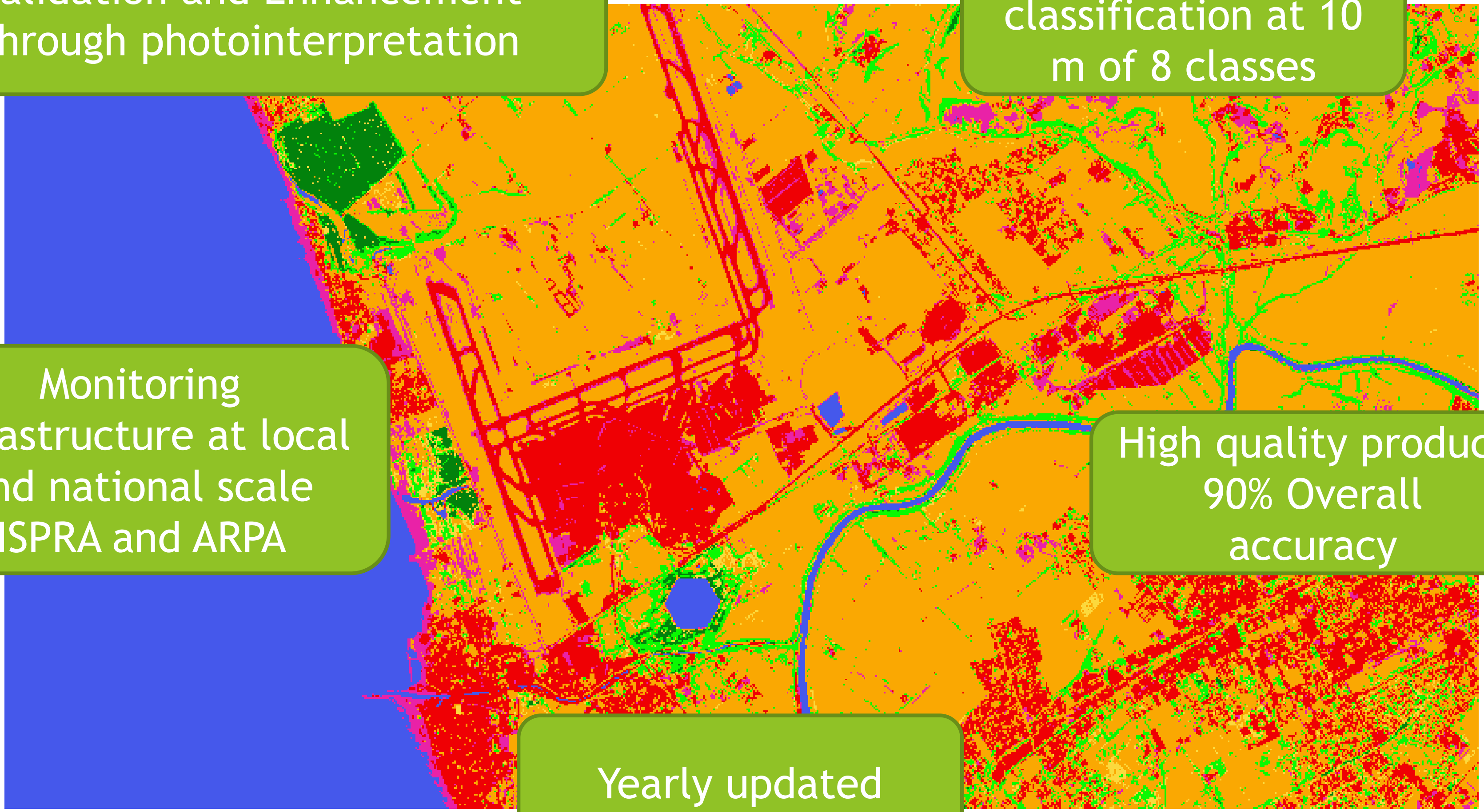
Binary mask
10 m resolution

- $NDVI = \frac{B8-B4}{B8+B4}$
- $NDSI = \frac{B11-B3}{B11+B3}$
- $NDWI = \frac{B3-B8}{B3+B8}$
- $NDCI = \frac{B8-B11}{B8+B11}$

# Land cover classification

Validation and Enhancement through photointerpretation

Land cover classification at 10 m of 8 classes

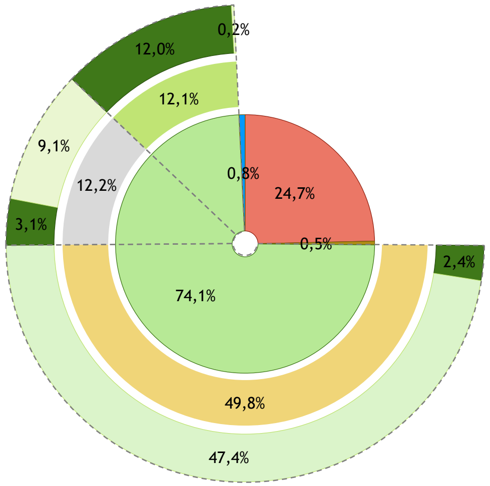


Monitoring infrastructure at local and national scale ISPRA and ARPA

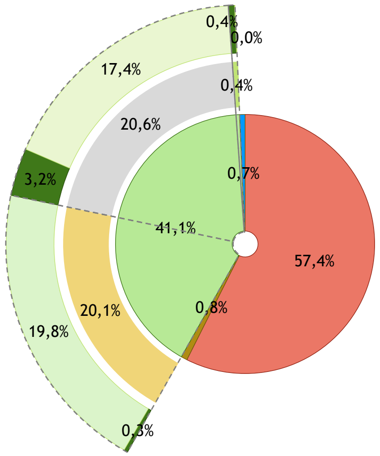
High quality product 90% Overall accuracy

Yearly updated

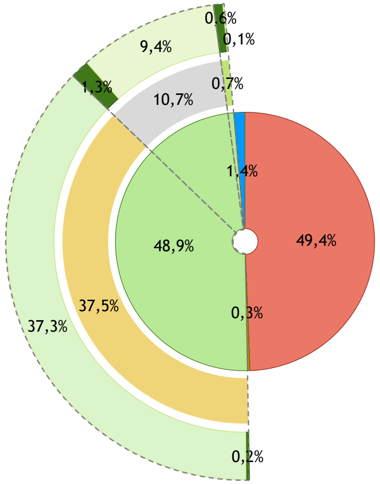
# Land cover classification results



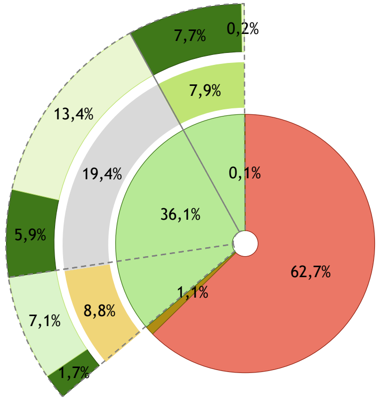

Rome (1.287 km$^2$)



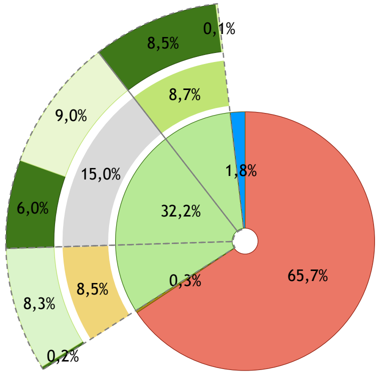

Milan (182 km$^2$)




Padua (93 km$^2$)




Naples (119 km$^2$)




Turin (130 km$^2$)

- Artificial Surfaces and Constructions
- Natural Material Surfaces
- Vegetated Surfaces
- Water

- Agricultural Areas
- Green Urban Areas
- Natural Areas
- Herbaceous Vegetation
- Woody Vegetation

URBAN-GEO BIG DATA

# Politecnico di Torino
# Progress Report on Mobility Applications

**Piero Boccardo, Emere Arco**

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

MINISTERO DELL'ISTRUZIONE DELL'UNIVERSITA' E DELLA RICERCA

# Goals and Objectives

- Identification and production of an **homogeneous national road network graph** for traffic monitoring, routing, asset inventory, meteo app…
  - Define an appropriate data model
  - Found accurate spatial data
- Data on **private traffic monitoring** (speeds, flow, travel times…) – appropriate modelling in order to set up the visualisation of both dynamic/real-time and historical data
- Data on **public transport**
  - Spatial data – bus routes and bus stops positions
  - Monitoring data - arrival predictions, real time vehicle positions

# Methodology
# Road Network Graph

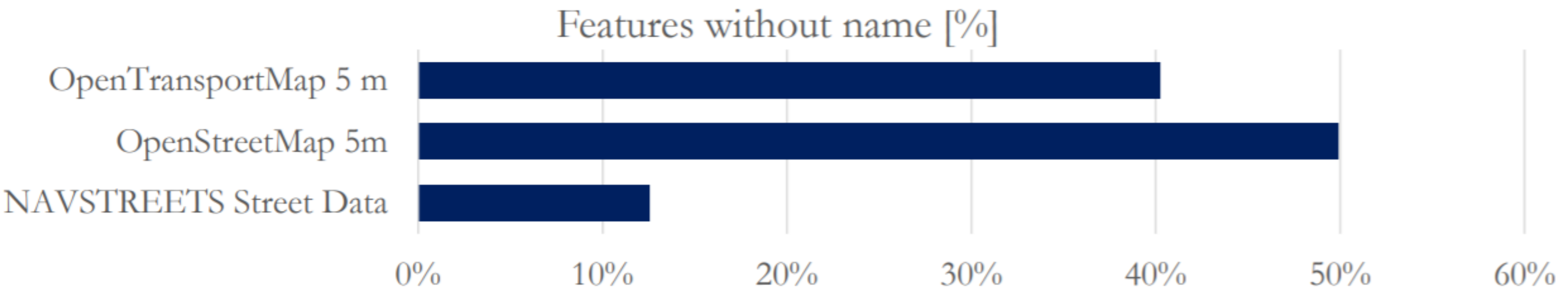
- Comparison between road network available datasets in term of:
  - Spatial quality
  - Attribute quality
- Considered datasources:
  - OpenStreetMap
  - OpenTransportMap
  - ANAS road network
  - Traffic Message Channel

# Methodology
## Road Network Graph

- The city of Turin used as a general example for the other cities: the commercial and reliable dataset from HERE has been used as TARGET for comparison
- Spatial accuracy and completness – evaluated through sets of buffers and intersections
- Attribute completness
- Network connection – evaluated through application of topology rules

# Methodology
## Road Network Graph

| | | NAVSTREETS Street Data | OpenStreetMap | OpenTransportMap |
|---|---|---|---|---|
| **Piedmont - all** | Total Kms | 58.924,85 | 101.324,40 | 128.300,77 |
| | Features count | 462.875 | 306.465 | 702.501 |
| **Piedmont - car traversable** | Total Kms | 48.935,19 | 82.051,99 | 105.459,93 |
| | N° of features | 406.291 | 255.984 | 609.952 |



# Results
# Road Network Graph

- Spatial quality:
  - OSM and OTM have an high quality both in term of spatial accuracy and completness;
  - The ANAS road network has several lacks in term of completness, but good quality in term of spatial accuracy;
  - TMC is only logical + lacks in completness (only major roads);
- Attribute completness (names, direction of flow, characteristics of the roads, hierarchy):
  - OSM is more complete than OTM (attribute reduction);
  - ANAS complete but only for name attribute;
  - TMC has a solid naming convention and direction of flow definition, but lacks of other attributes;
- Connectivity:
  - ANAS lacks to much in completness in order to be evaluated
  - OTM is better than OSM (topology correction already applied)
- **OpenTransportMap** has been choosen as reference road network datasets for the cities considered in the project

# Methodology
# Private Traffic Data

- Map matching and conflation algorithms applied in order to enanche and transfer useful informations between network data sources
- In particular mapping the Traffic Message Channel points on the OpenTransportMap graph
- Associate TMC location to a precise location on OTM allow then the location of traffic events and measures, usually spreads on TMC , directly on the OTM network

# Methodology
# Private Traffic Data

- One TMC point can be related to one or more OTM points:

1. Selection of OTM points representing a crossing (grade of the node > 2)
2. Association of the names of the roads to the OTM crossing nodes
3. Selection of the OTM nodes to be associated with TMC points through: proximity and similarity of the associated names (Levenstain index)

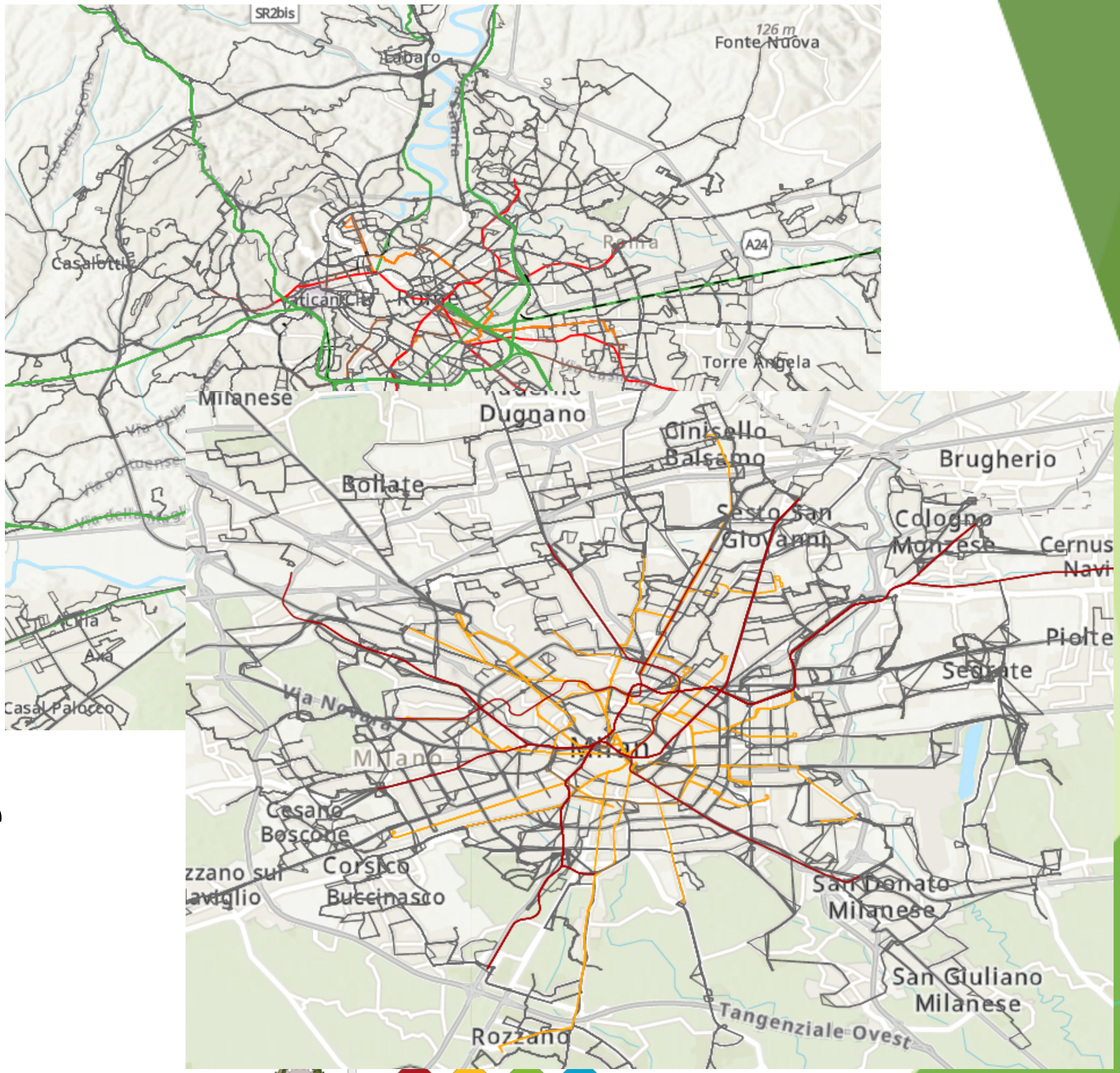
- At the end of this process we still don’t know if the OTM points selected are the right one – in particular we don’t know if the node is the one in the correct direction of flow (valid for double digitised roads)

# Methodology
## Private Traffic Data

- A routing algorithm using PostgisRouting has been set up (selection of preferential roads…)
- Find the route between two couples of OTM points following the rules of connection defined in the TMC
- Between multiple solutions only the ones with the minimum distance and minimum number of turns has been considered
- The resultant paths have been associated to the OTM network, adding an attribute which identify a TMC road in negative or positive direction)
- Results is correct for 83% of the considered TMC roads: in some cases the solution found was incorrect or no solution has been found by the routing algorithm → a manual revision is always needed!

# Methodology
# Public Transport Data

- GTFS is for now the only source found for public transport.
- GTFS is not available in Padua and Naples
- GTFS is generally published and updated every 2 weeks
- A comparison of differences between differents sets of GTFS is in progress, both in term of bus routes and stops (there is a change in time?) and in terms of predicted arrivals



# Further developments

- Enanchement of OTM network with TMC data also over the other cities.
- Waiting for the VIASAT data on private traffic monitoring – historical dataset which can be an homogeneous data over all the cities of the project
- Waiting for Rome data coming from the mobility platform (realtime travel times over main roads, real time bus stop arrivals)
- Integration of data of the other cites (viasat data, in particular) inside the traffic data visualisation platform
- Improvements of query performance and development of new functionalities in the traffic data visualisation platform
- Set up of a simple Geonode in order to share static data (tabular and spatial)
- Metadata definition is in progress…

# Next steps-activities

1. Implementation of the MIT Smart Roads Directive;
2. Asset security using mobility big data.

# Next steps-activities
# Implementation of the MIT Smart Roads Directive

By definition, Smart Road are a set of **road infrastuctures**, **technological platforms**, and **related services** aiming to accidents reduction, improve interoperabilty with next generation vehicles, ensure continuity with C-ITS European services, traffic decongestion, and road netwok sustainability, efficiency and resilience. The directive goal is then related to an **urgent and consistent improvement of the National road network** able to face the next technological and environmental challenges.

## Next steps-activities
## Implementation of the MIT Smart Roads Directive

Smart Roads can be implemented by an existent road network **digital transformation** step-by-step process, paying particular attention to **substainable activities costs**, balancing them, in the meantime, with **derived benefits**. With this view, the Intelligent Transportation Systems (ITS) promotion, should be considered as a must, considering that in this field, the return of the investiments is quite high.

# Next steps-activities
# Implementation of the MIT Smart Roads Directive

Regarding the directive functional specifications, below the list that will be implemented in the near future.

1. Road-side accessibility to a high speed communication data network
2. Routing and IoT connection Road network full coverage
3. Road-Side Unit for V2I communication
4. Hot-spot Wifi for personal device connectivity at least in service and parking areas
5. Traffic measurement and enforcement systems availibility
6. Availability of an archiving system for traffic measurement data

# Next steps-activities
# Implementation of the MIT Smart Roads Directive

7. Availability of a modelling system for traffic forecast
8. Availability of real time hydro/meteo systems
9. Infomobility service based on the actual traffic conditions (V2I and/or Web apps)
10. Real time traffic control centre availability
11. Parking and refuel (especially for electric vehicles) information to the users
12. Ability to deploy all the previous information in a V2X mode.

# Next steps-activities
# Implementation of the MIT Smart Roads Directive

Smart roads should be represented as an oriented and connected graph; theoretically this graph is not fully «cartographic», but implemented ad a transportation one, enabling by its attributes mathematical and data modelling procedures. Nevertheless, as stated in the directive transportation network should be associated to a mapping one, being many services based on geolocated positions. In addition, road network authorities (on the basis of comma 6 dell'art. 13 del Nuovo Codice della Strada, D.Lgs. n. 285, 30 aprile 1992,and of D.M. 1/6/2001 del Ministero dei Lavori Pubblici), should deliver the representation of their infrastructures in the Road Cadastre, by means of a graph with **G.D.F. specifications**.

**Unfortunately this data format does not ensure a fully interoperable approach and, at a national basis, is not homogeneous at all**.

# Next steps-activities
## Implementation of the MIT Smart Roads Directive

Smart road network, as stated in the directive, should be coinceived on the basis of three different levels:

1. **Connectivity layer**. Arcs (road centre line) connected by nodes only for primary roads hierarchy;
2. **Characteristics layer**. Arcs (as many as the number of road's roadways) connected by nodes for every single intersections;
3. **Segment layer** (compulsory for smart roads). As Characteristics layer with the addition of nodes (1 every 500 m) enabling infomobility homogeneity (for TMC purposes) and traffic and environmental procedure generalization.

## Next steps-activities
# Implementation of the MIT Smart Roads Directive

Test over the national available road network and procedures enabling Smart Roads Directive inplementation.

Concerning the other cities data, all the 5 mucipalities have been contacted and a data assessment will be produced in the next future. Anyway, the possibility to have homogeneous data, is related to FCD dataset acquired in the same time span.

Telecommunication data will be available with some pros and cons; peculiarities will be discussed later.

URBAN-GEO BIG DATA

# Analysis of the Floating Car Data of Turin Public Transportation system: first results

Roberta Ravanelli, Mattia Crespi

University of Rome “La Sapienza”
Department of Civil, Constructional and Environmental Engineering
Geodesy and Geomatics Division

PRIN meeting
Padova, January 24th 2019

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

MINISTERO DELL'ISTRUZIONE DELL'UNIVERSITÀ E DELLA RICERCA

# Introduction

- The largest part of **movements** in an **urban environment** is constrained to the **road network**
- In the field of transportation, **GNSS data** collected from **vehicles** are frequently referred as **Floating Car Data** (**FCD**)
- **FCD** are **Urban Geo Big Data** and contain the key information for estimating traffic **impedance maps**, potentially in real-time

**Aim of the work**

To develop a reliable **methodology** able to perform the preliminary analyses needed for computing the **impedance maps from FCD**

- **management** and **visualization** of a **huge data amount**
- preliminary tests for **projecting** the **raw FCD** to the **route lines**

# Features of the analysed FCD

The analysed **FCD**:

- acquired in the month of April 2017 by the on board units installed on the vehicles of the **Turin Public Transportation** company (Gruppo Torinese Trasporti - GTT)
- include the pairs of WGS84 **geographical coordinates** (longitude, latitude) along with a **set of attributes** (vehicle code, line code, turn, timestamp, ecc.)
  - variable acquisition rate (from few to tens of seconds)
- provided in the **CSV** format
  - the original file is very **heavy** (2.19 GB)
  - converted in a **database** through a `Python` script based on the `sqlite3` and `pandas` libraries

# Database generation

About **30·000·000** records!

DB Browser for SQLite - G:\PRIN_GEO_BIG_DATA\Dati_floating_car_data_Boccardo\fcd\csv_database_formatted.db

File Edit View Help

New Database | Open Database | Write Changes | Revert Changes

Database Structure | Browse Data | Edit Pragmas | Execute SQL

Table: fcd_table | New Record | Delete Record

| | index | linea | turno(?) | date | mezzo | lat | lon |
|---|---|---|---|---|---|---|---|
| | Filter | Filter | Filter | Filter | Filter | Filter | Filter |
| 1 | 1 | 36 | 5 | 2017-04-28 21:05:09.000000 | 802 | 45.073677062... | 7.5964450836... |
| 2 | 2 | 64 | 1 | 2017-04-28 11:10:02.000000 | 3041 | 45.064193725... | 7.6750035285... |
| 3 | 3 | 51 | 2 | 2017-04-27 08:54:49.000000 | 977 | 45.119415283... | 7.7108950614... |
| 4 | 4 | 6 | 3 | 2017-04-26 13:41:13.000000 | 6027 | 45.073696136... | 7.6814966201... |
| 5 | 5 | 44 | 3 | 2017-04-13 13:47:58.000000 | 8017 | 45.066818237... | 7.5776481628... |
| 6 | 6 | 58 | 3 | 2017-04-10 07:18:26.000000 | 2620 | 45.038261413... | 7.6190347671... |
| 7 | 7 | 5 | 5 | 2017-04-09 08:49:13.000000 | 1039 | 45.028236389... | 7.6017150878... |
| 8 | 8 | 81 | 2 | 2017-04-08 09:20:20.000000 | 1254 | 44.994644165... | 7.7242064476... |
| 9 | 9 | 11 | 17 | 2017-04-06 11:56:30.000000 | 948 | 45.124114990... | 7.6440901756... |
| 10 | 10 | 16CS | 10 | 2017-04-24 19:16:25.000000 | 2857 | 45.072139739... | 7.6556334495... |
| 11 | 11 | 58SB | 22 | 2017-04-25 20:18:18.000000 | 2785 | 45.060665130... | 7.6614084243... |
| 12 | 12 | 63 | 6 | 2017-04-18 10:46:06.000000 | 2769 | 45.011482238... | 7.6365866661... |
| 13 | 13 | 72SB | 22 | 2017-04-15 19:34:18.000000 | 1000 | 45.095348358... | 7.6690135002... |
| 14 | 14 | 67 | 2 | 2017-04-13 11:31:45.000000 | 3007 | 45.004108428... | 7.6849350929... |
| 15 | 15 | 57 | 8 | 2017-04-12 22:27:30.000000 | 855 | 45.067314147... | 7.6714982986... |
| 16 | 16 | 3 | 23 | 2017-04-07 08:51:16.000000 | 5012 | 45.099411010... | 7.6486001014... |
| 17 | 17 | 13 | 8 | 2017-04-06 09:30:10.000000 | 2857 | 45.076423645... | 7.6698732376... |
| 18 | 18 | 5 | 8 | 2017-04-05 17:12:22.000000 | 800 | 45.056266784 | 7.6644783020 |

1 - 18 of 30015120 | Go to: 1

# Velocity analysis

- The **FCD** were **organized** for **lines**, then for **vehicles** and finally they were **chronologically ordered**

- For every line of the transportation network:
  - the *Vincenty* formula was used to compute the **planimetric displacement** $\Delta s$ between **two positions** of the specific vehicle in two **consecutive time moments**
  - the **velocities** were computed as $v = \dfrac{\Delta s}{\Delta t}$

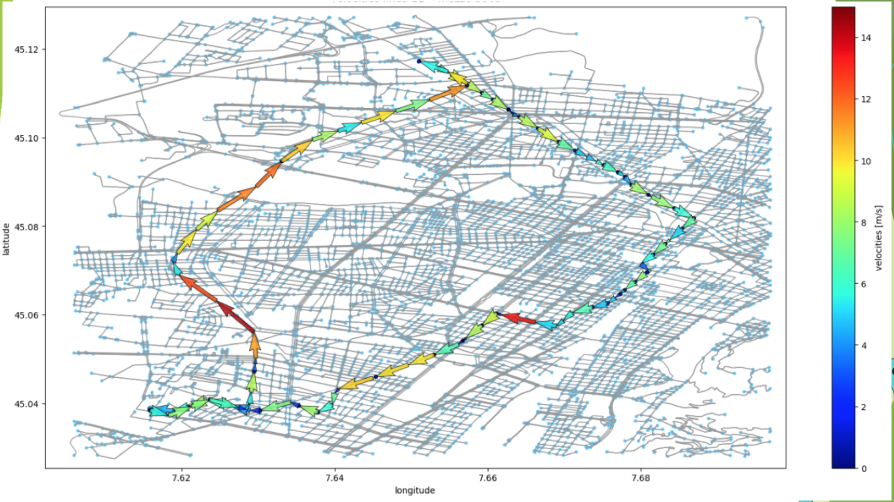
# Velocity analysis

- The computed **velocities** were represented as **arrows** and plotted on top of the Turin drive network graph, automatically downloaded from Open Street Map through the `OSMnx` Python library

```
def boh(drive_network_graph, start_lon, start_lat, delta_lon, delta_lat, velocities, sogli
    fig_from_function, ax_from_function = ox.plot_graph(drive_network_graph, close = False
    nz = mcolors.Normalize(vmin = soglia_inf, vmax = soglia_sup)
    #plt.gca().set_aspect('equal', adjustable='box')
    plt.quiver( start_lon, # start x
                start_lat, # start y
                delta_lon, # delta x
                delta_lat, # delta y
                angles='xy', # 'xy': arrows point from (x,y) to (x+dx, y+dy). Use this for
                scale=1, # più è grande, + le frecce sono corte Number of data units per a
                scale_units='xy', # usando le scale units, non è più necessario alterare l
                color=cm.jet(nz(velocities)),# color = velocities
                zorder = 5, #più è alto, più il plot è in primo piano
                edgecolor='k', # colore bordo freccia
                linewidth=.7,
                alpha=0.8) # trasparenza
```

# Example of computed velocities

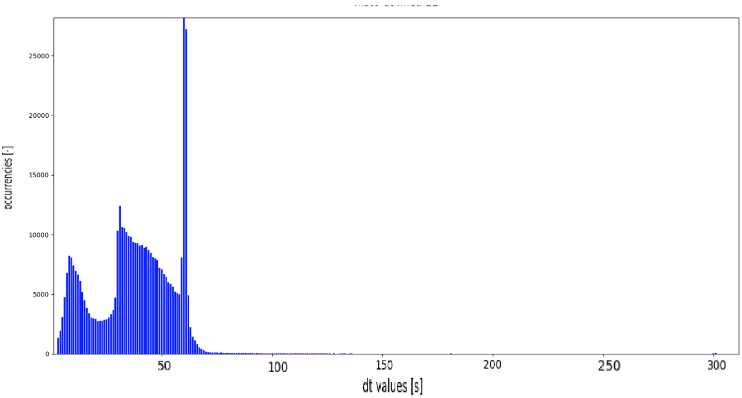
Line 11, vehicle 3063



# Outlier removal

Before proceeding with the time analysis, the **outliers** were **removed** by eliminating all the records:

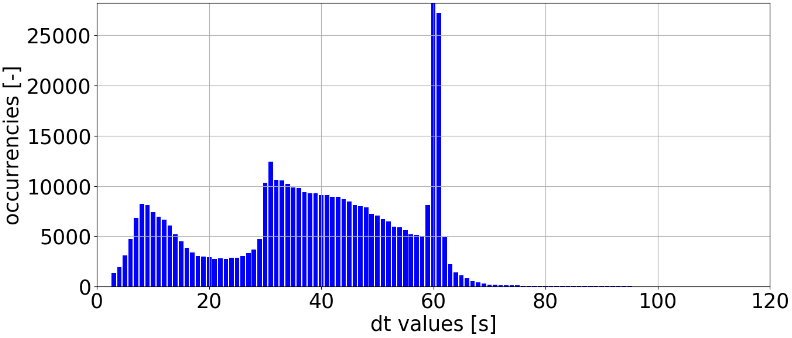
1. whose $\Delta t$ are higher than $99.5^{th}$ percentile and lower than $0.5^{th}$ (statistically not significant)
2. characterized by a velocity higher than 5 times the mean



# Outlier removal

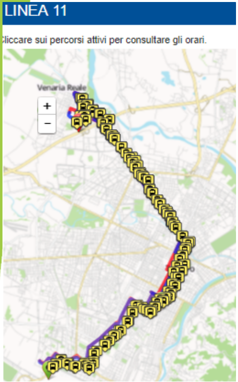
Before proceeding with the time analysis, the **outliers** were **removed** by eliminating all the records:

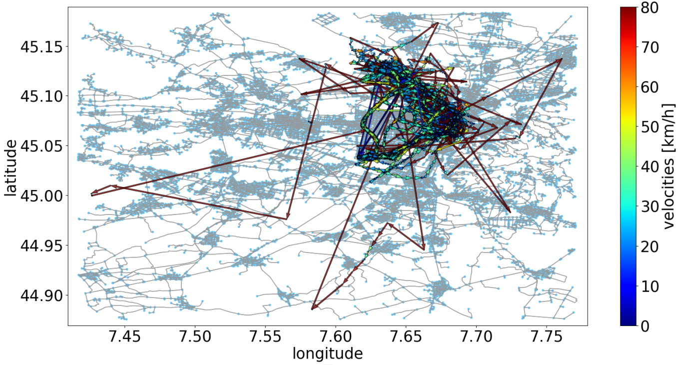
1. whose $\Delta t$ are higher than $99.5^{th}$ percentile and lower than $0.5^{th}$ (statistically not significant)
2. characterized by a velocity higher than 5 times the mean



# Line 11: velocities

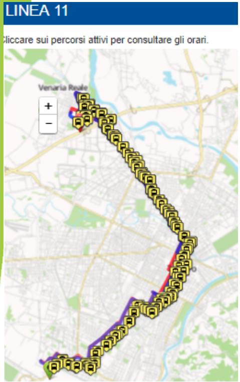
After the **outlier removal**, the reconstructed path follows more closely the actual line route: the **longest arrows**, probably due to the bus routes from and to the depot, are **eliminated**

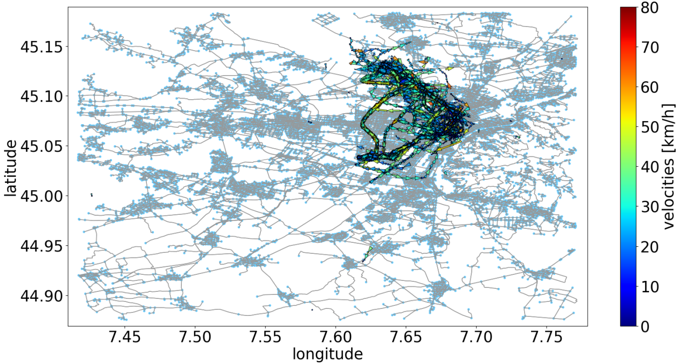


# Line 11: velocities

After the **outlier removal**, the reconstructed path follows more closely the actual line route: the **longest arrows**, probably due to the bus routes from and to the depot, are **eliminated**

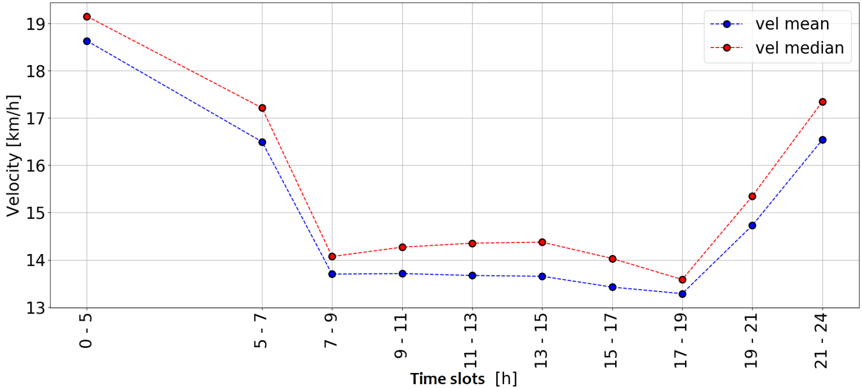


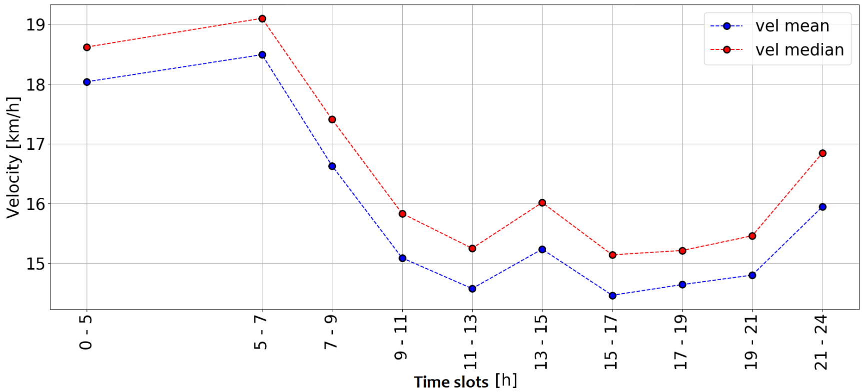
# Temporal analysis

- Once the outliers were removed, a **temporal analysis** was performed
- The data were divided into **working** and **weekend days**, considering the following **time intervals** during the day:
  - 0 - 5
  - 5 - 7
  - 7 - 9
  - 9 - 11
  - 11 - 13
  - 13 - 15
  - 15 - 17
  - 17 - 19
  - 19 - 21
  - 21 - 24

# Line 11: time slot velocities in working days



# Line 11: time slot velocities in weekend days



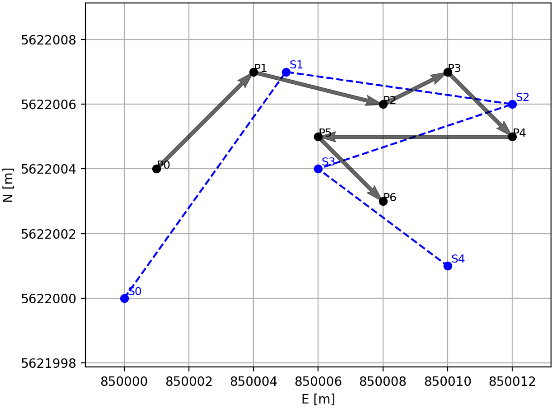
# Considerations

- The **highest velocities** occur at **night** and in **late evening**, with a **local peak** shortly after the **lunch hour**
- The **lowest velocities** occur during the **peak hours**, in correspondence of the **office entrance** and **exit hours**
- The differences between **working** and **weekend days** are more **evident** in the **peak hour** time slots
- During the **0-5** and **21-24** time slots, the **difference** is **small**, since in these hours the **traffic** level is low also in the **working days**

# FCD projection to line networks

A preliminary strategy was implemented to **assign** the **velocities to** the **line network topology**:

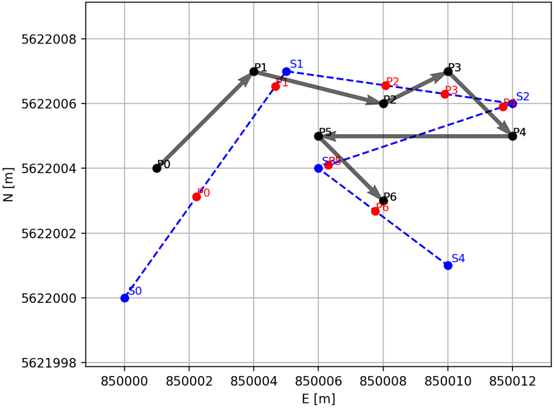
- for **every FCD point**, the **closest tree** of the specific line network is selected



# FCD projection to line networks

A preliminary strategy was implemented to **assign** the **velocities to** the **line network topology**:

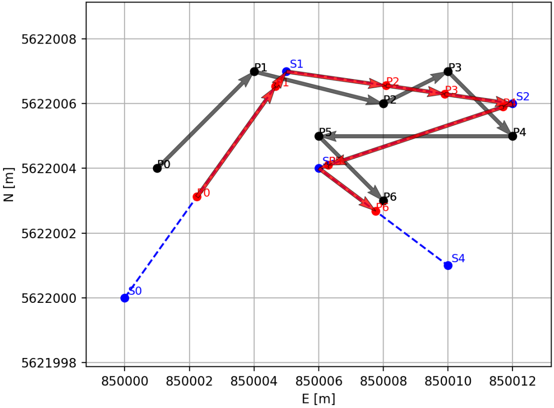
- for **every FCD point**, the **closest tree** of the specific line network is selected



# FCD projection to line networks

A preliminary strategy was implemented to **assign** the **velocities to** the **line network topology**:

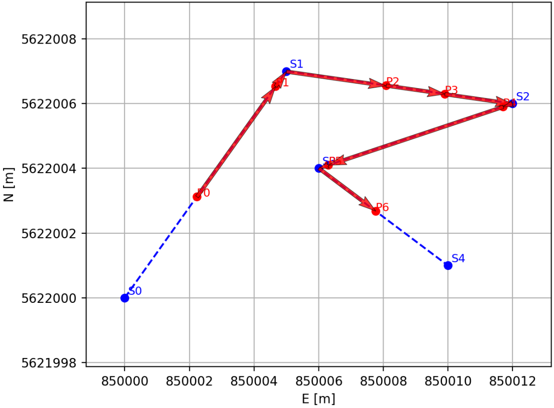
- for **every FCD point**, the **closest tree** of the specific line network is selected



# FCD projection to line networks

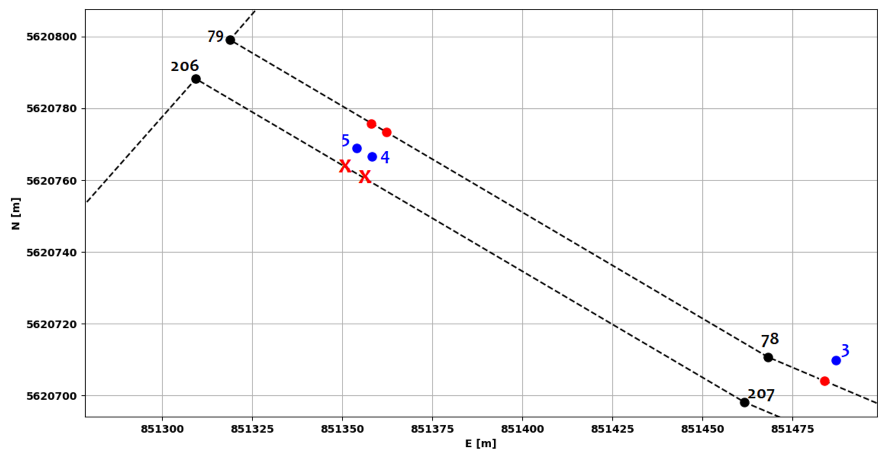
A preliminary strategy was implemented to **assign** the **velocities to** the **line network topology**:

- for **every FCD point**, the **closest tree** of the specific line network is selected



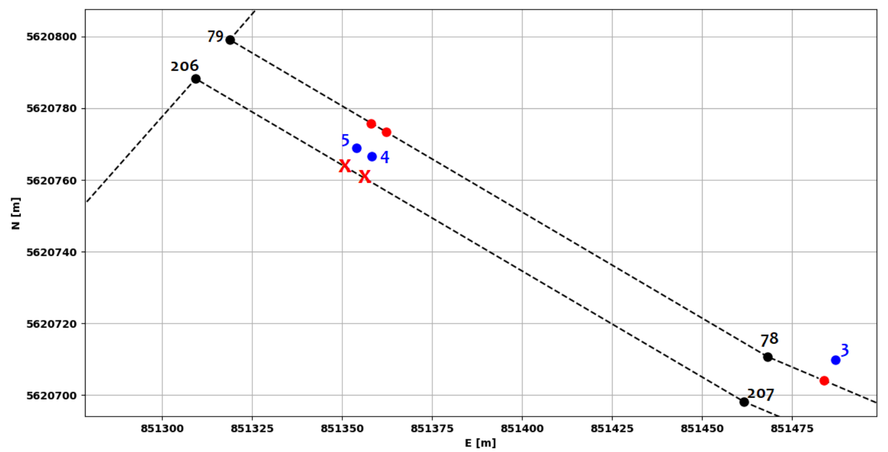
# Topological issues

**Topological issues** occur when the FCD point is located in a segment in which the **distance between two** (or more) **arcs** is **comparable** to the **GNSS measurement errors**



# Topological issues

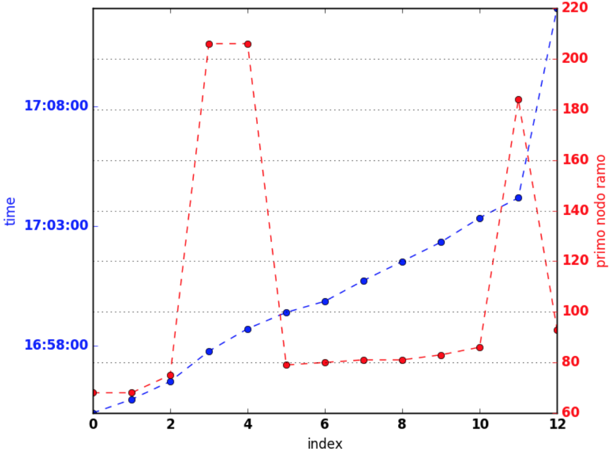
It is rather improbable that the FCD **point 4** and **point 5** may be assigned to the **tree 206-207** of the network, since the vehicle was located in the **tree 77-78** few moments before



# Topological issues

A possible solution is to consider:

- the **temporal information** contained in the **FCD**
- the **cardinality information** contained in the **line network**




Select the segment closest to the previous selected tree

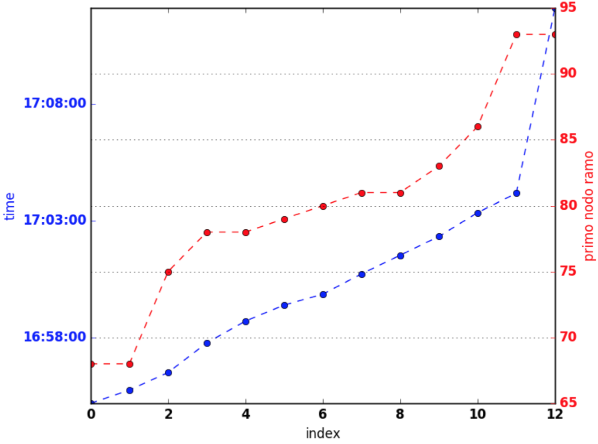
# Projection algorithm

The **assignment errors** can be identified considering that:

1. the bus cannot travel back in time $t_{i+1} > t_i$
2. the bus cannot travel big distances in a short time interval $nodo_{t+1} < nodo_t + 20$
3. the bus cannot move in the wrong direction: $node_{t+1} > node_t$ (possible problem when a new lap begins)

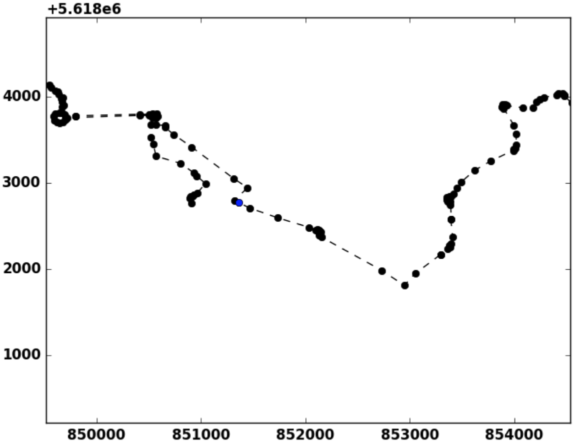
# Projection algorithm

The **temporal trend** of the **node IDs** must be **constant** or **increasing** with small slopes (constant or positive derivative)



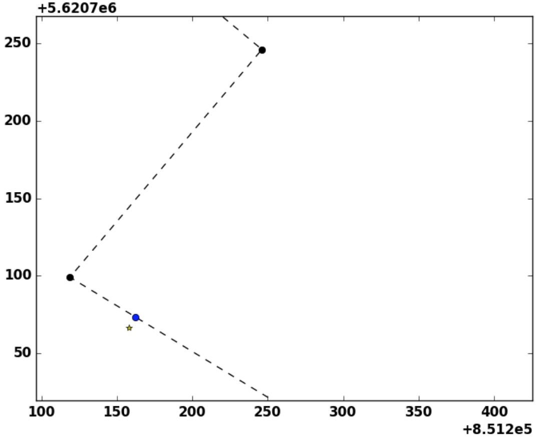
# Projection algorithm: error removal

For every projected point, if the **ipothesis number 2** is **not verified**, the **arc** incorrectly selected is **removed** from the network together with the **following arcs** and the projection is newly performed



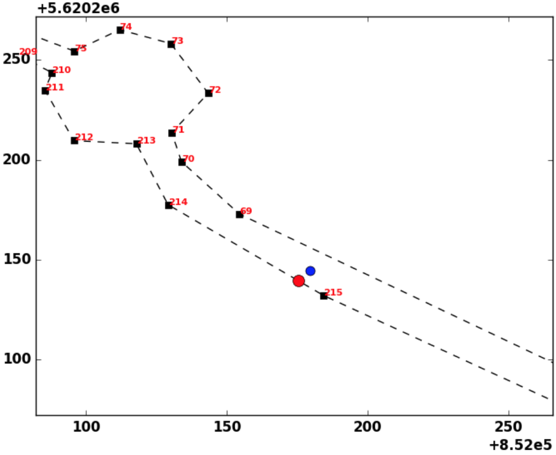
# Projection algorithm: error removal

For every projected point, if the **ipothesis number 2** is **not verified**, the **arc** incorrectly selected is **removed** from the network together with the **following arcs** and the projection is newly performed



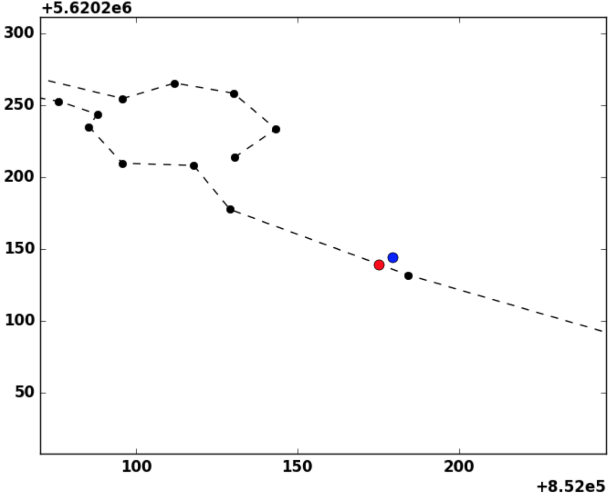
# Projection algorithm: error removal

For every projected point, if the **ipothesis number 3** is not verified, the **arc** incorrectly selected is **removed** from the network together with the **previous arcs** and the projection is newly performed



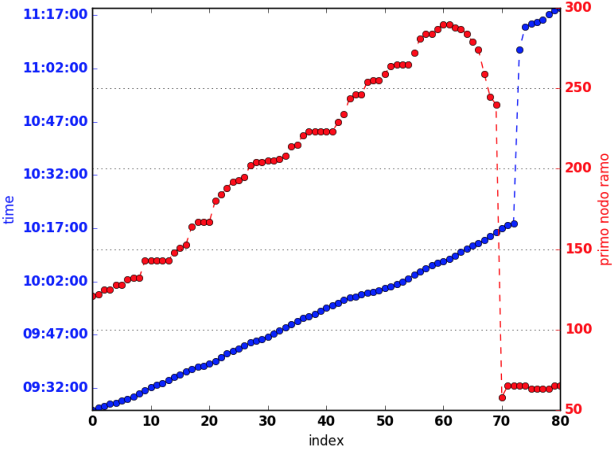
# Projection algorithm: error removal

For every projected point, if the **ipothesis number 3** is not verified, the **arc** incorrectly selected is **removed** from the network together with the **previous arcs** and the projection is newly performed



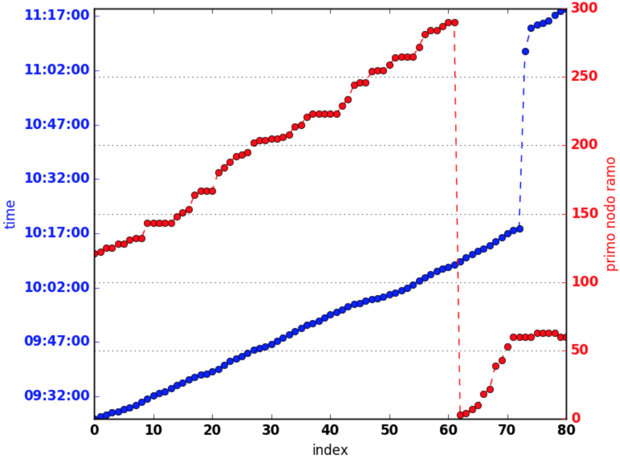
# Identification of a new lap

- A **new lap** can be identified on the basis of a **peak** in the trend of the **node IDs**
- Once the last point of the lap is identified, the following point are newly projected forcing the algorithm to consider only nodes with low values of IDs



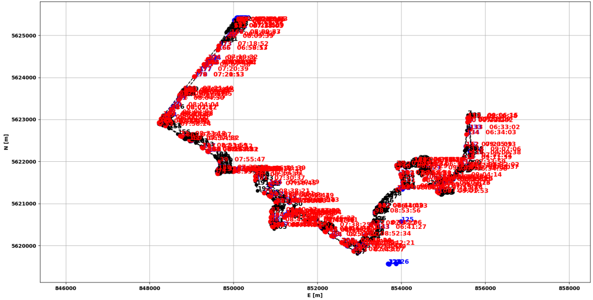
# Identification of a new lap

- A **new lap** can be identified on the basis of a **peak** in the trend of the **node IDs**
- Once the last point of the lap is identified, the following point are newly projected forcing the algorithm to consider only nodes with low values of IDs



# Results

- The designed and implemented algorithm is quite effective
- Few assignement errors still remain, nevertheless a solution has already been designed and is under implementation




Problem in visually validating such huge amount of data

# Conclusions

- A **first strategy** to analyse the **FCD** of the Turin Public Transportation system was implemented, in view of an automatic and possible real-time impedance map generation
- A **huge amount** of **FCD** were processed to compute the **vehicles velocities**
- A **visualization** approach based on Osmnx library was adopted
- A preliminary **temporal analysis** was carried out
- A method to **assign** the **velocities** to the line **network topology** was developed and successfully tested

# Further developments

- To refine the outlier removal process in order to all the **velocities not referable** to the **actual path** of the **lines**
- To **test** the developed **topological procedure** on **all** the **velocity data**, by checking the **effective reliability** and **real-time feasibility** of the designed methodology
- To **compute** the **impedence maps** and deliver the corresponding **metadata**
- To extend the developed methodology to other cities

# Thank you for your kind attention!

This work was supported by **URBAN-GEO BIG DATA**, a Project of National Interest (**PRIN**) funded by the Italian Ministry of Education, University and Research (**MIUR**) id. 20159CNLW8

URBAN-GEO BIG DATA

POLITECNICO MILANO 1863

ISPRA
Istituto Superiore per la Protezione e la Ricerca Ambientale

istituto per il rilevamento elettromagnetico dell'ambiente

24 January 2019, Padua

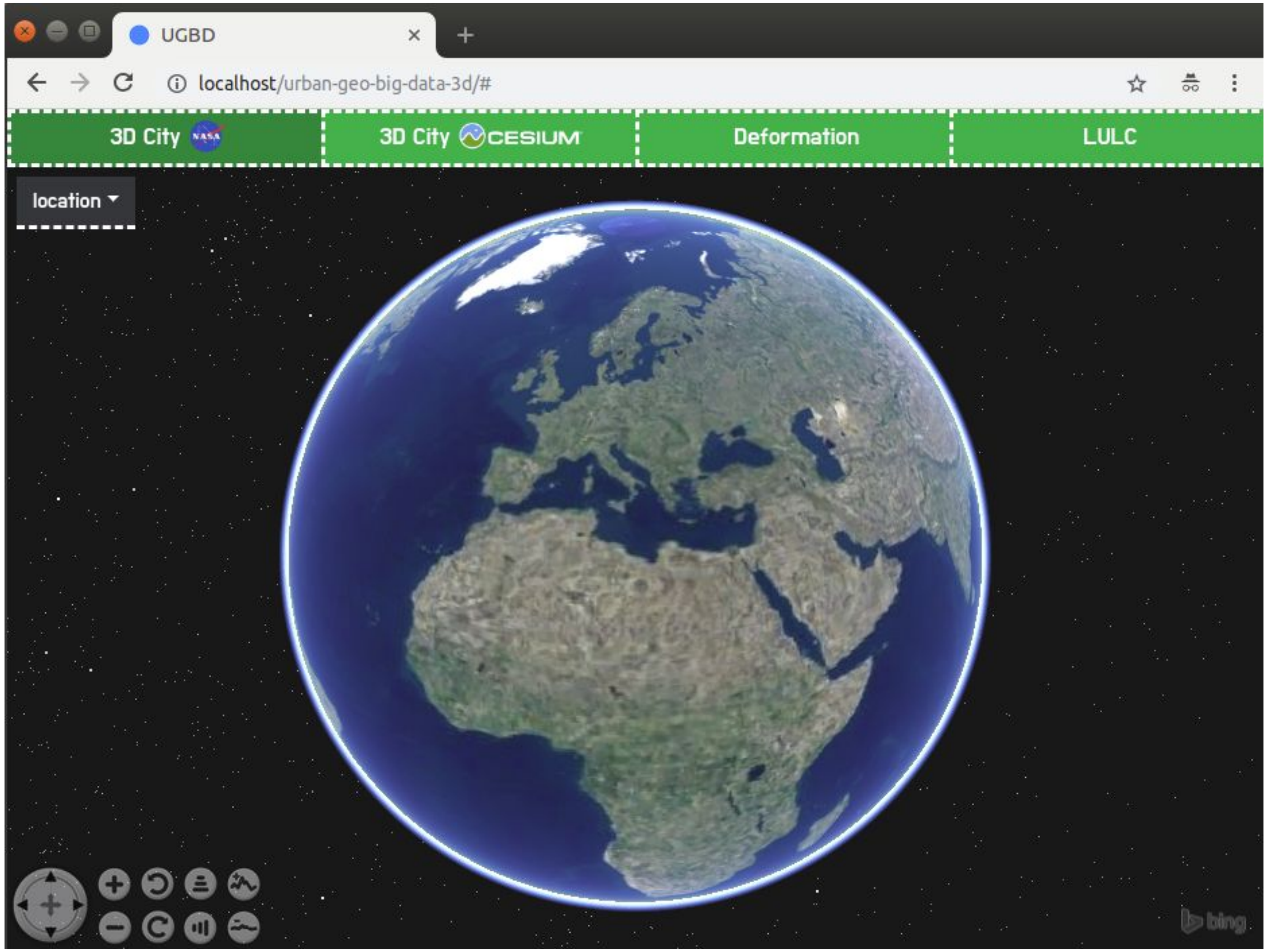
# PoliMi
# Progress Report

Maria Antonia Brovelli, Candan Eylül Kilsedar,
Daniele Oxoli, Gorica Bratic

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

MINISTERO DELL'ISTRUZIONE DELL'UNIVERSITA' E DELLA RICERCA

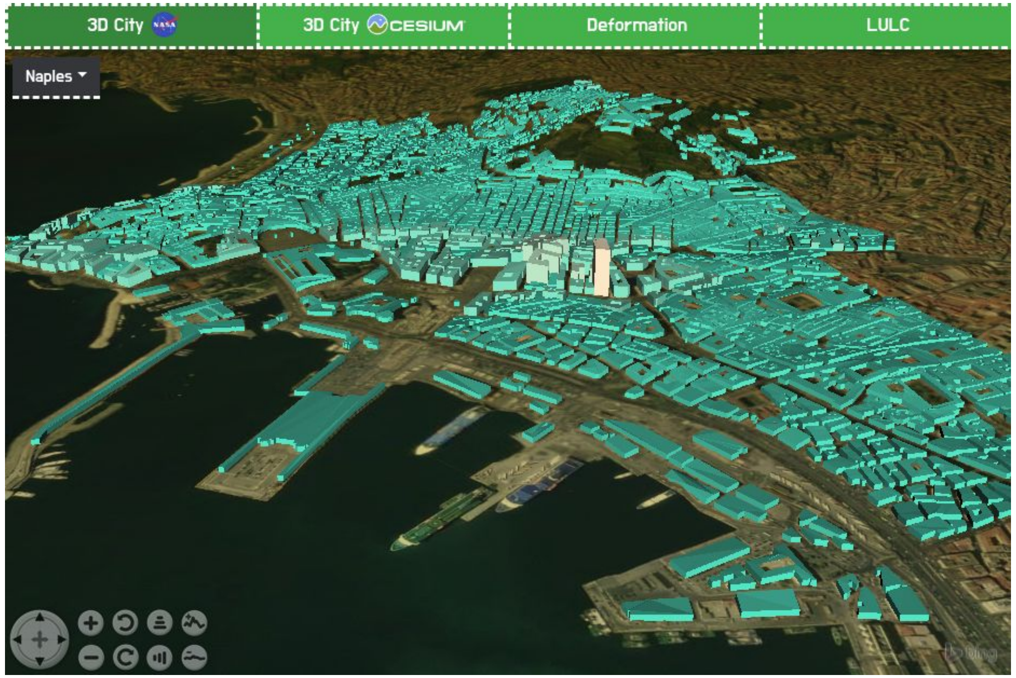
# Web GIS Interface

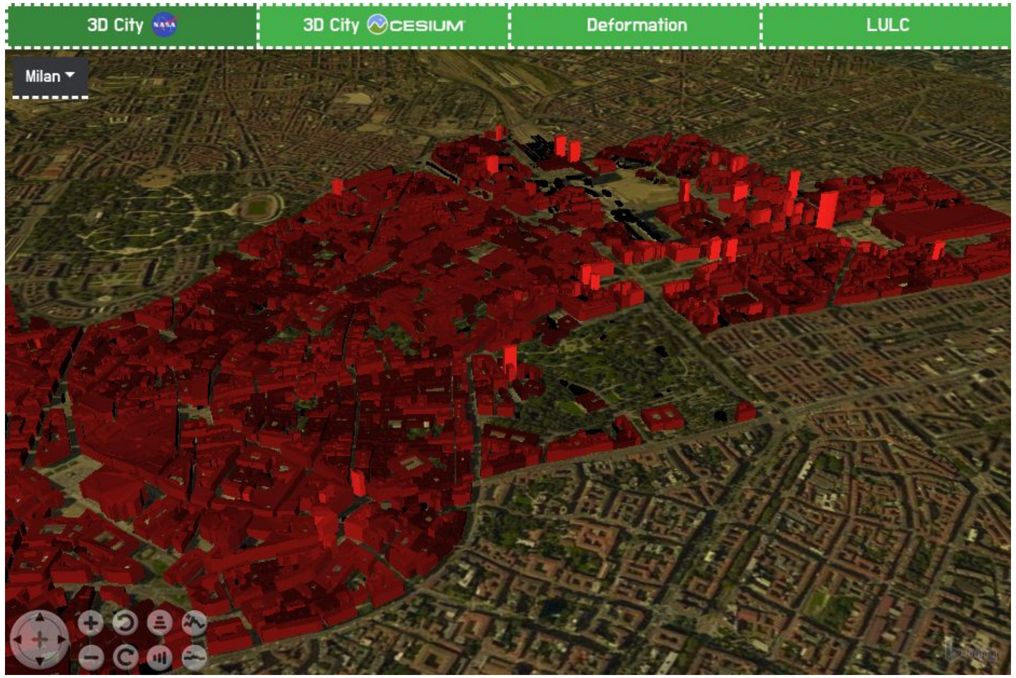



https://github.com/kilsedar/urban-geo-big-data-3d

# 3D City Visualization

- 3D OSM buildings visualization is performed using NASA Web WorldWind API and my 2017 GSoC project 3D OSM Plugin API. Its source code is available at https://github.com/kilsedar/3dosm.
- LiDAR and OSM GeoJSON are used in GRASS GIS to set **Milan** buildings' height.
- Urban Atlas Building Height 2012 data of the Copernicus programme and OSM GeoJSON are used in GRASS GIS to set **Rome** buildings' height.
- OSM attributes are used to set the buildings' height in **Naples**, **Padua**, and **Turin**.



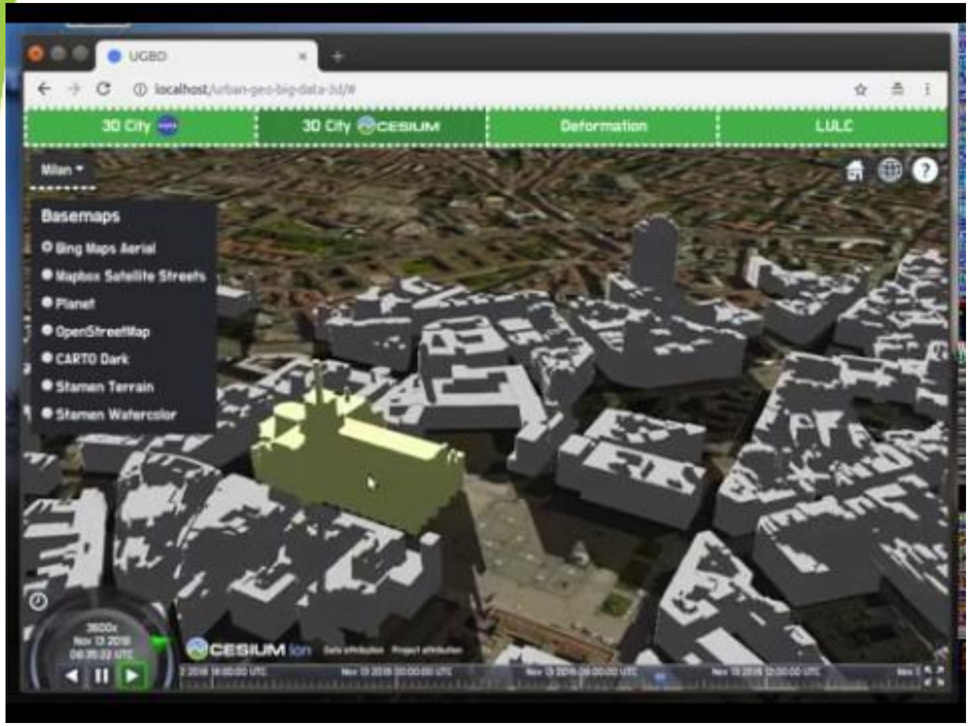
# CesiumJS Basemaps and Terrain

- Added 7 basemaps to the CesiumJS-based virtual globe:
  - Bing Maps Aerial
  - Mapbox Satellite Streets
  - Planet
  - OpenStreetMap
  - CARTO Dark
  - Stamen Terrain
  - Stamen Watercolor


- VR-TheWorld Server is used to construct terrain. It provides global elevation data with 90-meter resolution (DTED level 1) for the entire globe, including bathymetry.

# 3D City Visualization

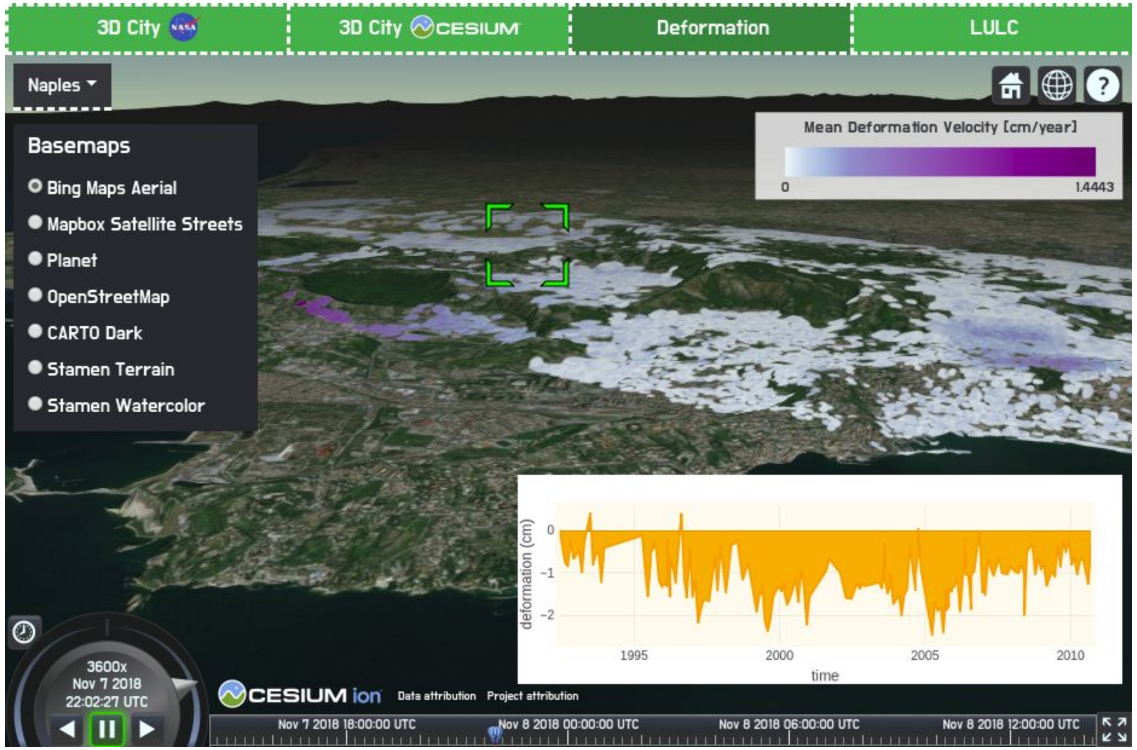
- 3D buildings visualization is performed using CityGML data and CesiumJS.
- Conversion from shapefile to CityGML is developed by the University of Padua (Francesca Fissore and Francesco Pirotti).
- CityGML dataset is converted to KML/gltF using [3D City Database Importer/Exporter].
- Data is visualized and queried using CesiumJS and [3DCityDB-Web-Map-Client].




# Deformation Visualization and Query

- Data is stored on GeoServer and used through Web Map Service (WMS) on a virtual globe built with CesiumJS. Each target on the ground can be queried to display the deformation time series plot. Plotted the time series using Plotly.
- Selected the colors using ColorBrewer for sequential data:

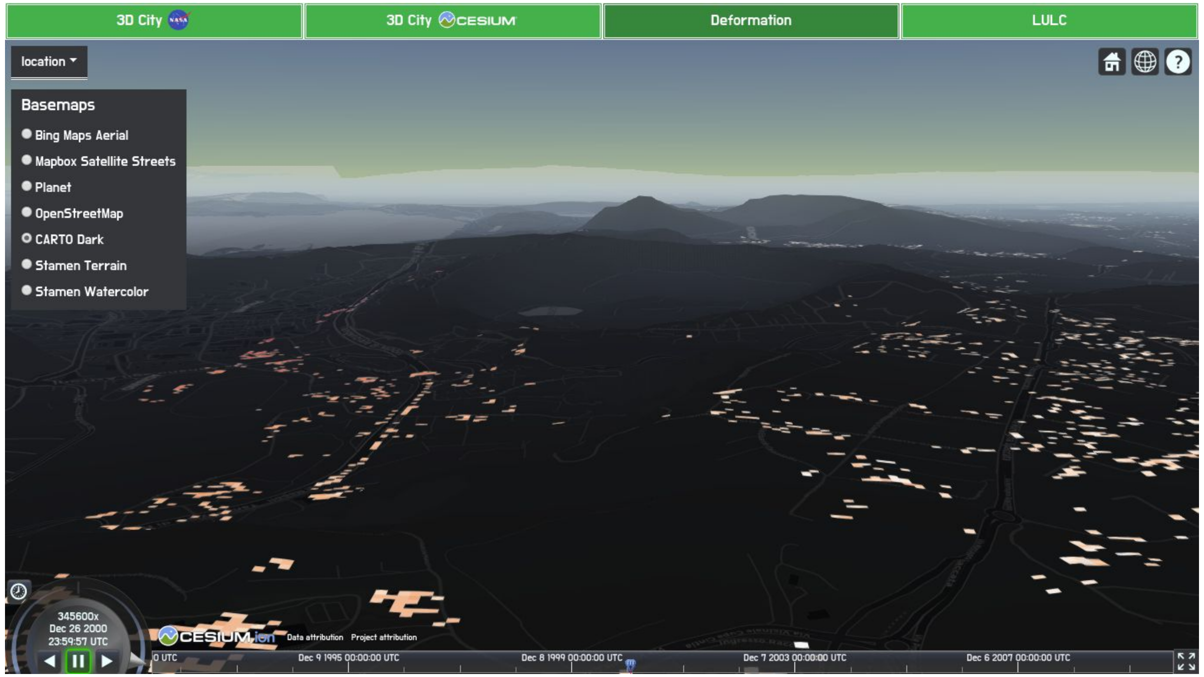
  http://colorbrewer2.org/#type=sequential&scheme=BuPu&n=7.



# Deformation Animation

- Deformation is also visualized in cm/year for 16 years as animation using Web Map Tile Service (WMTS) and ImageMosaic through GeoServer and timeline and animation widgets of CesiumJS to demonstrate the movement of the land for Naples. The same will also be implemented also for Milan, Rome, Turin, and Padua.
- Selected the color scheme using ColorBrewer for diverging data:

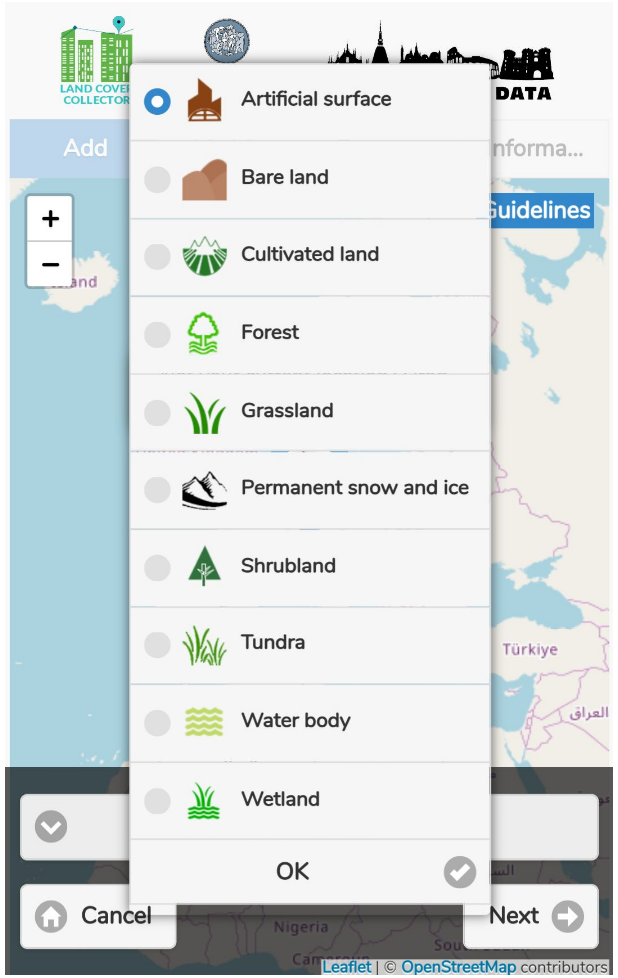
  http://colorbrewer2.org/#type=diverging&scheme=RdBu&n=11.



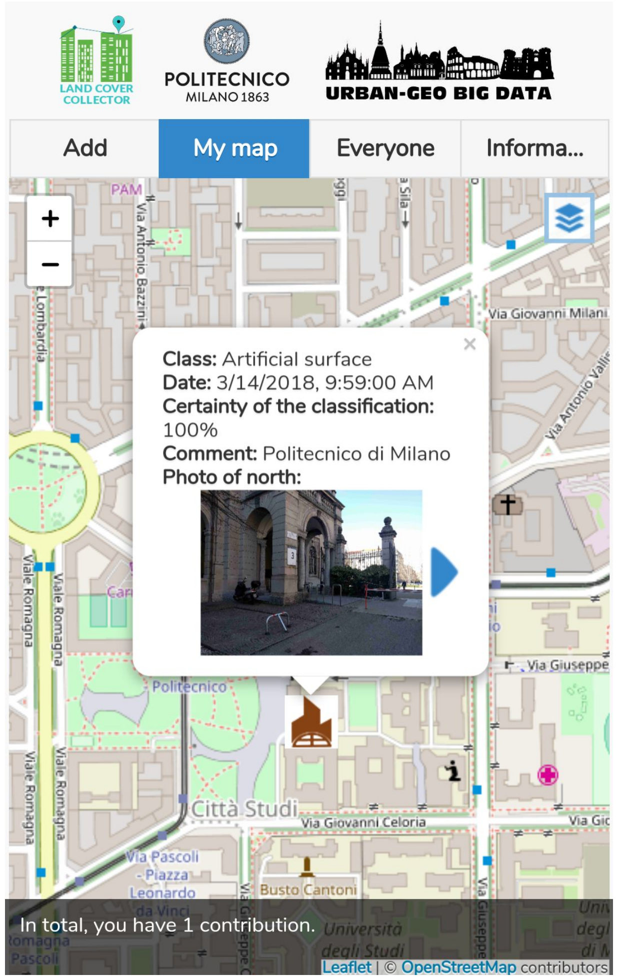
# Land Cover Collector

- Land Cover Collector application aims to enable data collection on land cover classification using the nomenclature of GlobeLand30 (GL30).
- 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.
- It is possible to collect data online and offline.
- It is cross-platform and available as Android and iOS applications and on Web.
- Collected data are released under the Open Database License (Odbl) v1.0 and can be downloaded in JavaScript Object Notation (JSON) format.
- Guidelines are provided.
- It is available in eight languages: English, Italian, Arabic, Chinese, Spanish, Russian, Portuguese and French.
- It is an open source project, available on GitHub at https://github.com/kilsedar/land-cover-collector. The project is protected under the GNU GPLv3 license. The application is partially based on and extends the project EmoMap.

# Land Cover Collector



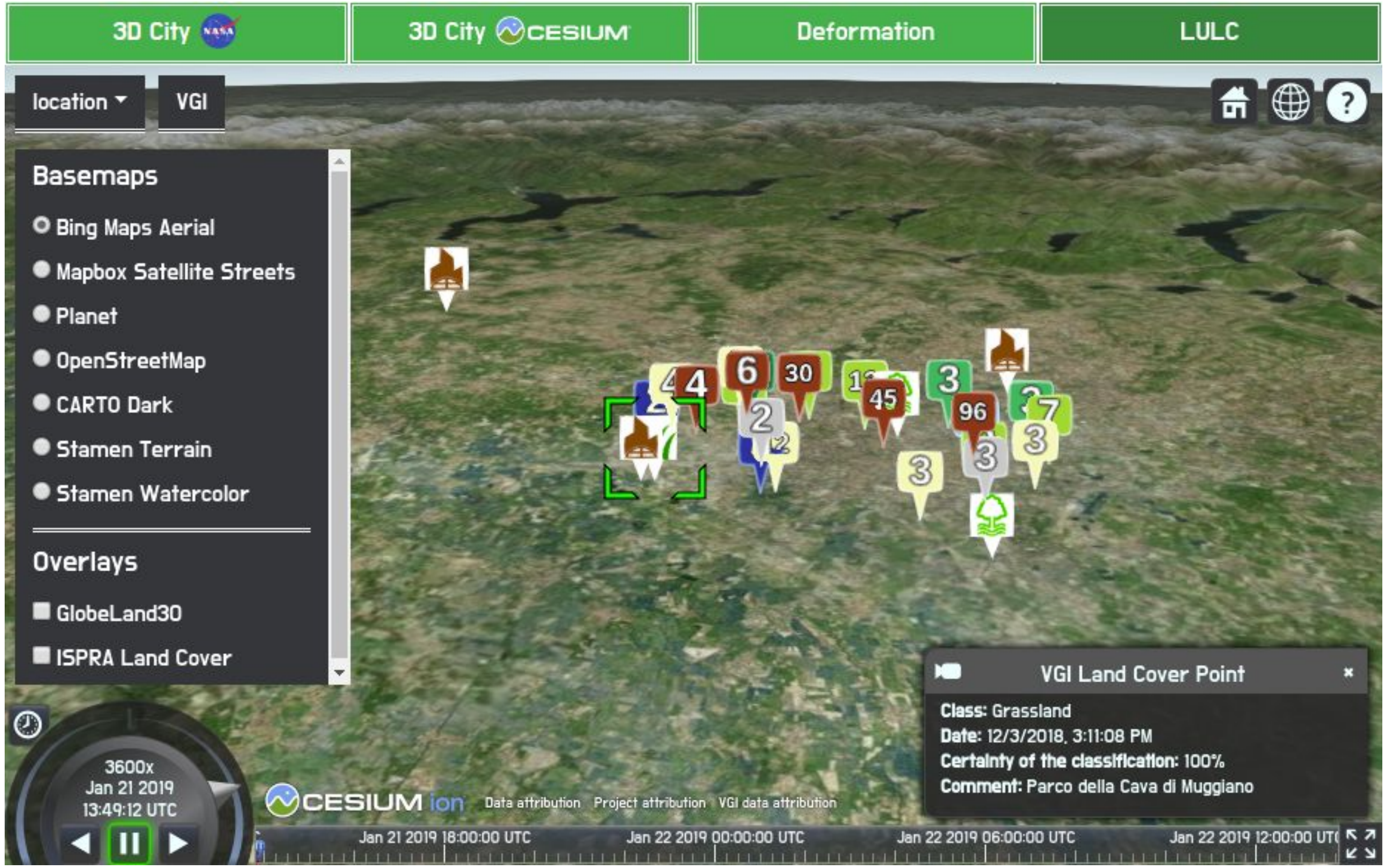

While submitting a land cover point, its class according to GL30 classification is set.




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

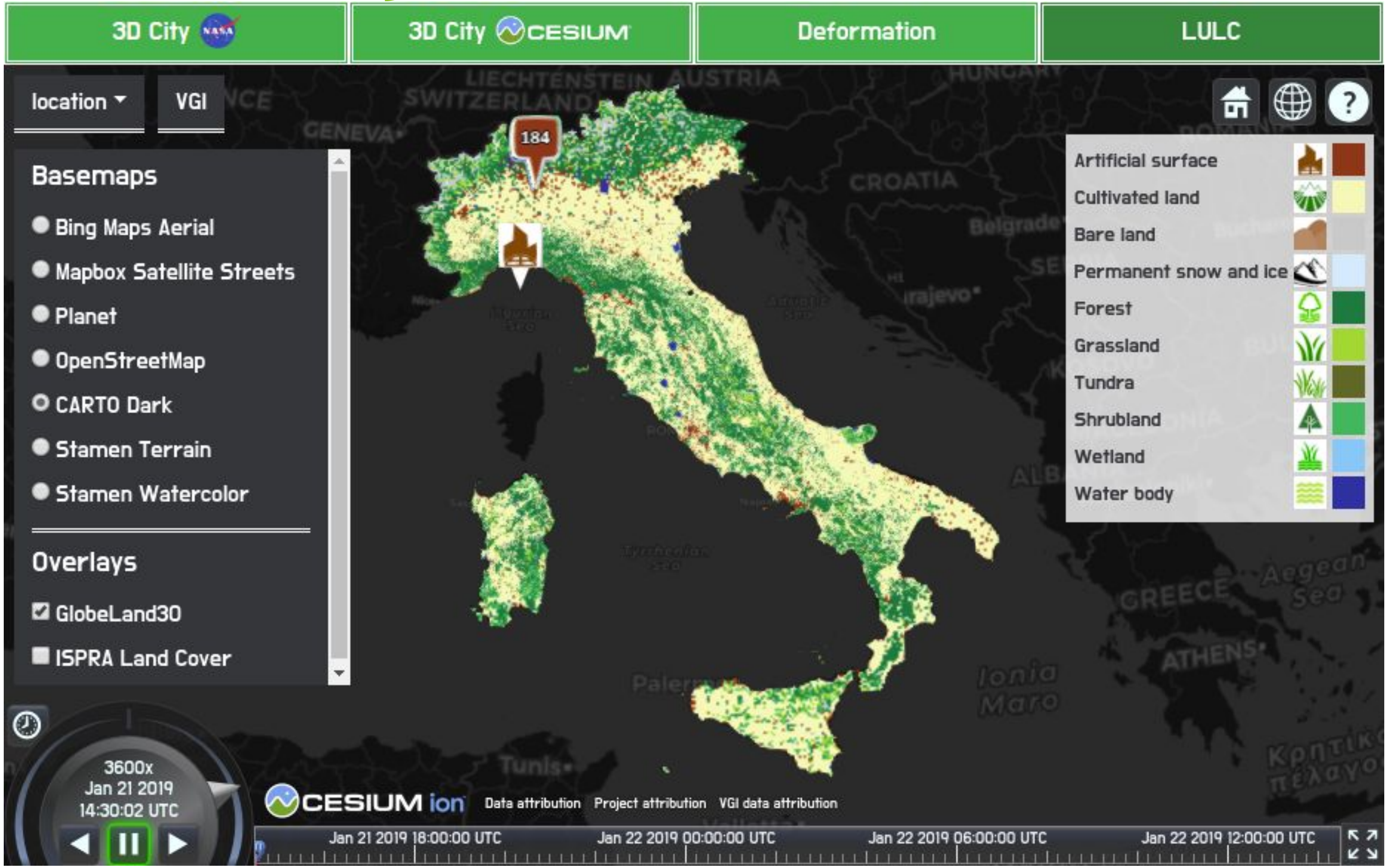

The collected land cover data can be visualized as clusters, aggregated based on the declared land cover class.

# Land Cover VGI Visualization

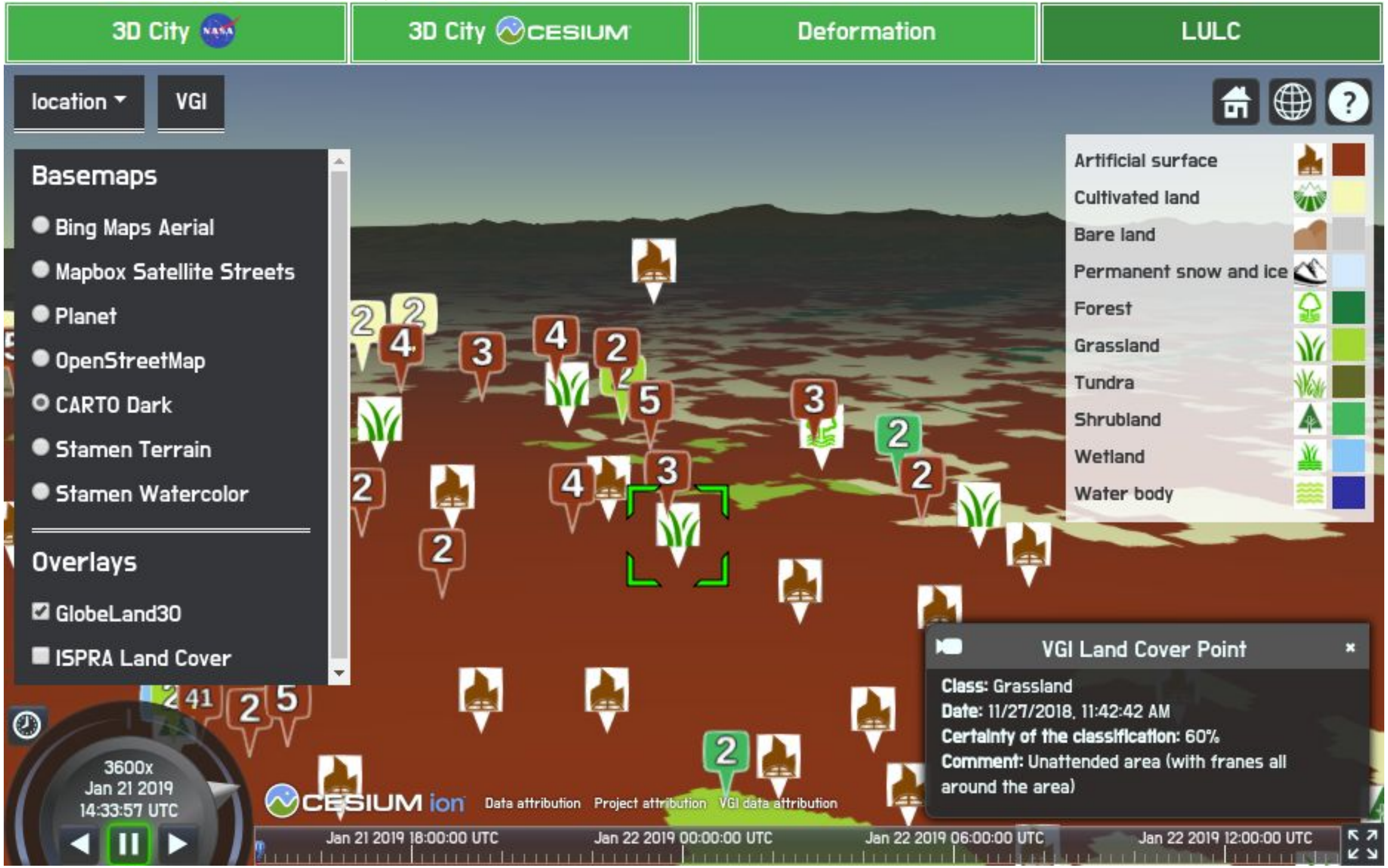



- Data collected using Land Cover Collector application can be visualized as "billboard" clusters on CesiumJS-based virtual globe. Icons and cluster colors are adapted from GL30. Each billboard can be queried. (Photos of north, east, south, and west are to be added.)

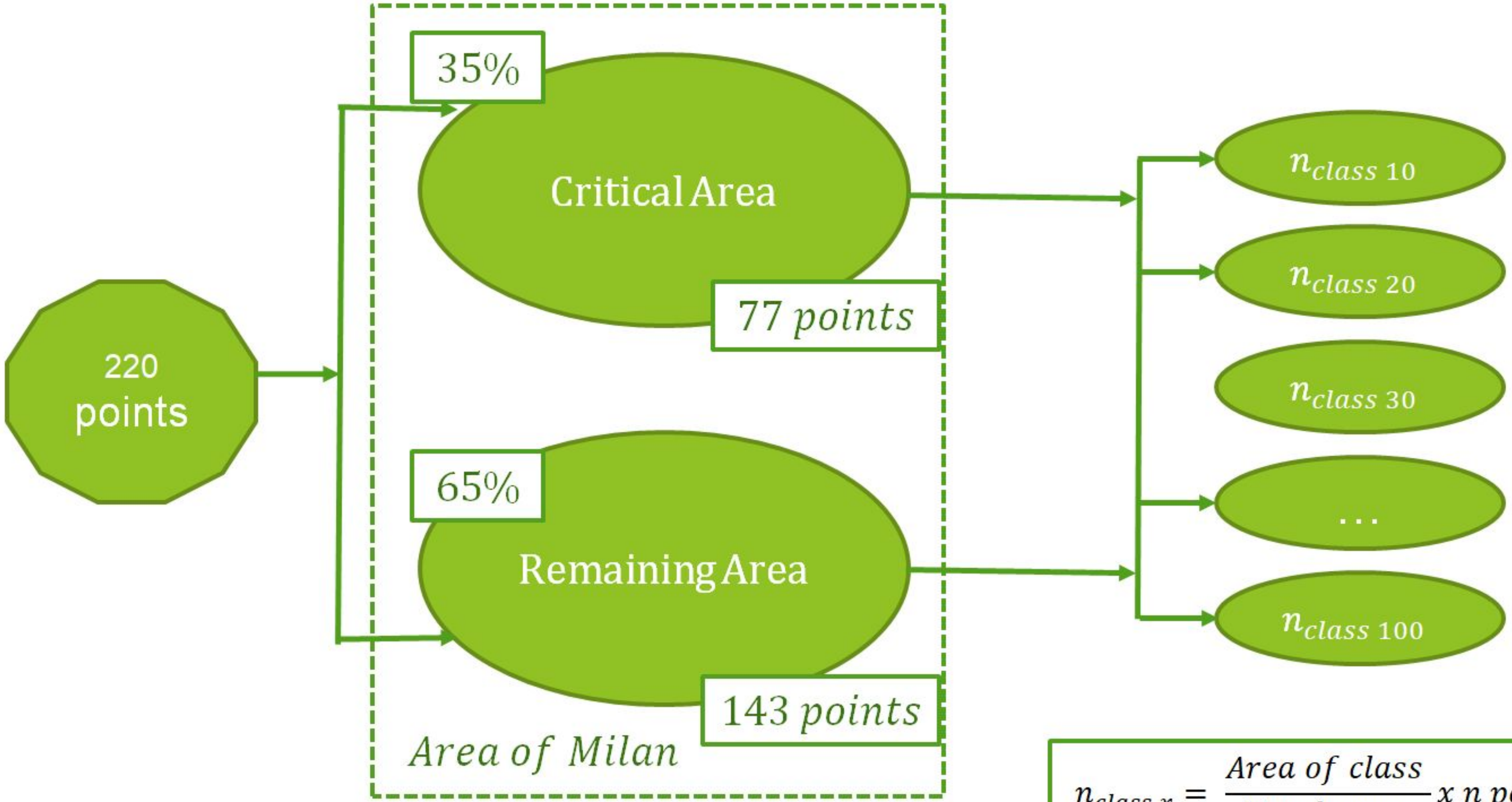
# GL30 and Land Cover VGI



# GL30 and Land Cover VGI

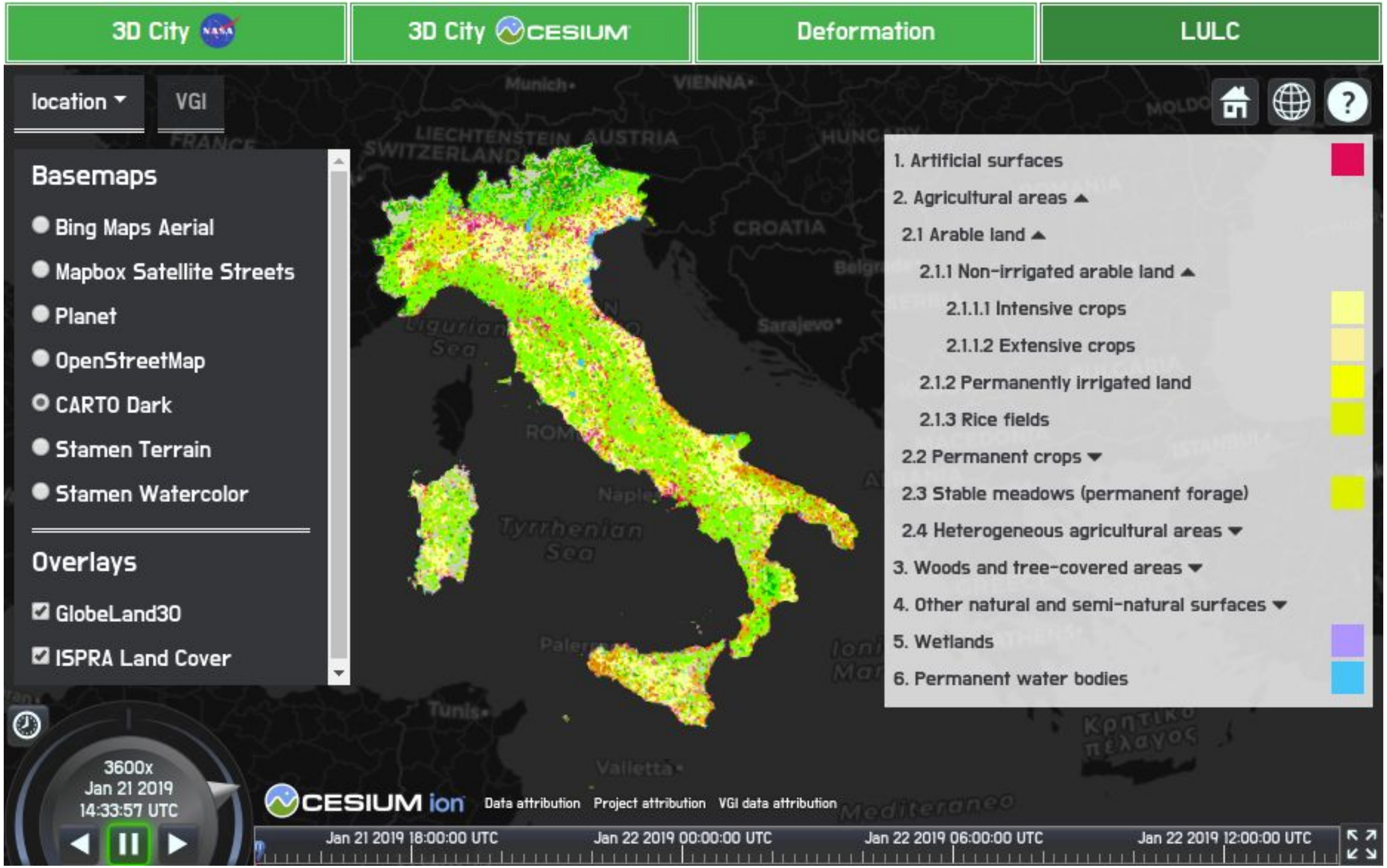


# VGI LULC validation data




$$n_{class\ x} = \frac{Area\ of\ class}{Total\ area}\ x\ n\ points\ in\ area$$

# ISPRA Land Cover



# Big Raster Data Visualization (in progress)

- Following data will be visualized using WMTS and ImageMosaic through GeoServer and timeline and animation widgets of CesiumJS to detect soil consumption and LULC change in time visually:
  - Binary soil consumption data from ISPRA, 10m, [2012, 2015, 2016]
  - Land cover data from ISPRA, 10m, [1990, 2000, 2006, 2012]
  - Global Human Settlement Map, 40m, [1975, 1990, 2000, 2014]

- Additional data to be visualized to enrich the Web GIS:
  - Europe Settlement Map, 10m, [2016]
  - Global Human Footprint, 12m
  - ...
  - (Urban Atlas for five cities, vector, [2016, 2012])

# Query and Processing Big Multidimensional Raster Data (to-do)

- Existing technologies will be reviewed:
  - Digital Earth Australia
  - Swiss Data Cube
  - EarthServer
  - E-sensing platform
  - Google Earth Engine
  - rasdaman
  - …
- The best one for our purpose will be chosen that is free and open source and supports the functionalities we aim to use.

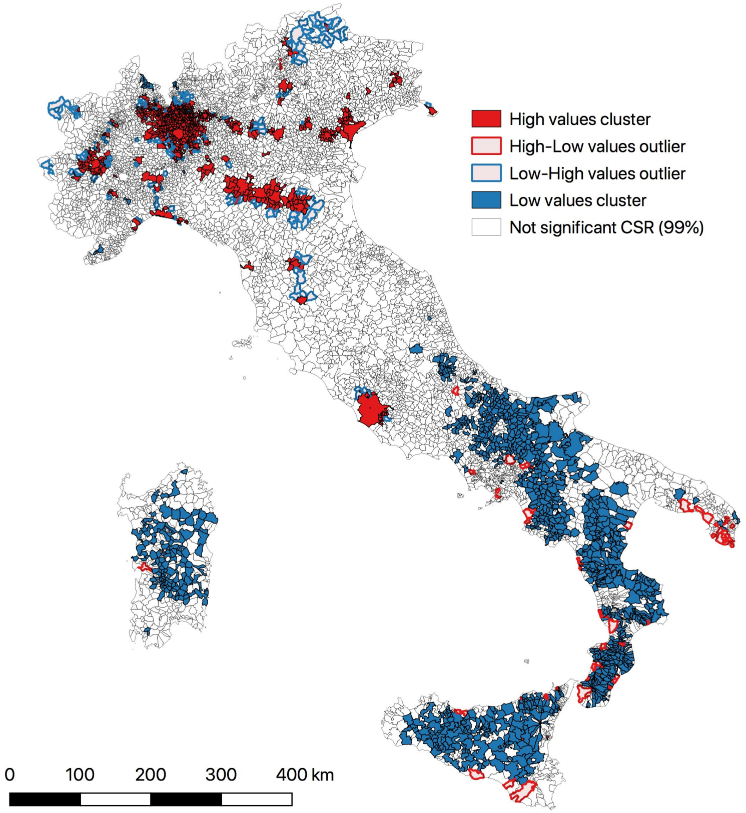
# Query and Processing Big Multidimensional Raster Data (to-do)

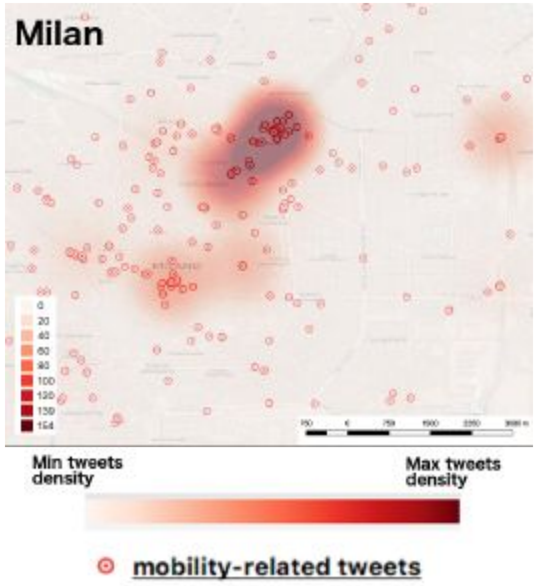
- Query and processing of big multidimensional raster data within the Web GIS with the selected technology will be developed.
- For the listed soil consumption and LULC data, on the Web interface the following operations are planned to be allowed:
  - click on a pixel, select a date, return the value 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;
  - …

# Exploratory Spatial Data Analysis (ESDA)

- Development of ESDA methods (spatial statistics + visualisation techniques + GIS software) to identify spatial patterns in a generic spatial dataset
- Focus on multivariate spatial association analysis
- Preliminary applications on PRIN core data (e.g. soil consumption and crowdsourcing data)

Bivariate Local Moran's I for the total soil consumed and the average income per capita in the Year 2012.



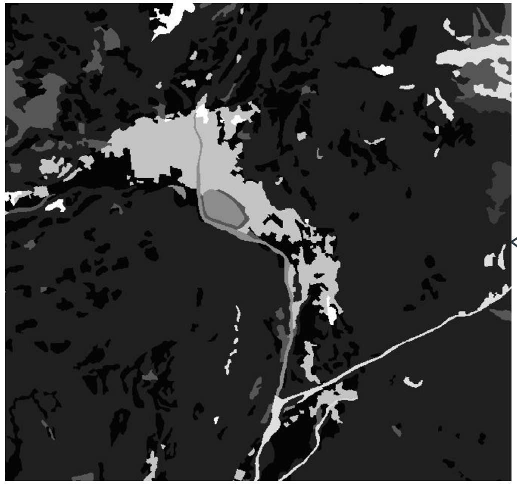



# Enhanced Accuracy Assessment of Global LULC products (ongoing)

- High-resolution global land cover accuracy assessment against local land cover products

GlobLand30
(30m)

∩



DUSAF
(1:10000)

global / local **spatial association analysis on disagreements** (multi-class)

traditional accuracy assessment (global indices)

**Enanched Accuracy Assessment**

- Purposes
  - Better comprehension of classification error patterns
  - Point out pros & cons of global LULC products applications at a local scale (e.g. urban studies)

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, Daniele Oxoli, Gorica Bratic**

Department of Civil and Environmental Engineering | GEOlab, Politecnico di Milano

maria.brovelli, candaneylul.kilsedar, daniele.oxoli, gorica.bratic@polimi.it

URBAN-GEO BIG DATA

POLITECNICO MILANO 1863

istituto per il rilevamento elettromagnetico dell'ambiente

ISPRA
Istituto Superiore per la Protezione e la Ricerca Ambientale

# Shp2city: conversion between cartographic data to cityGML

F. Fissore, F. Pirotti

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

MINISTERO DELL'ISTRUZIONE DELL'UNIVERSITA' E DELLA RICERCA

# Goal:

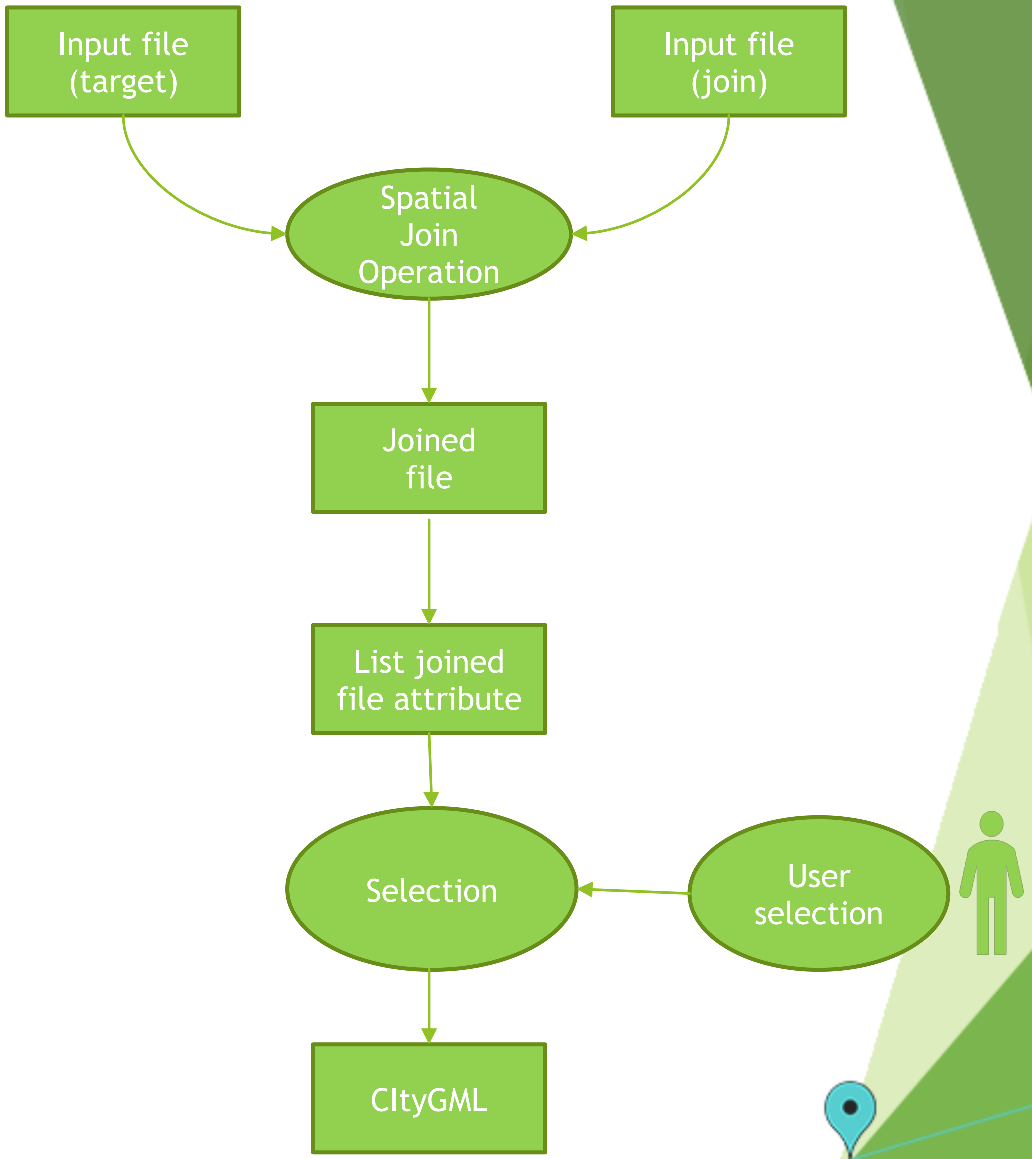
- Find an automatic way to convert data in ESRI shapefile format to 3D cityGML model.
- Data Test: Italian Cartography

↓

## Shp2City

# Shp2City

- Python Language: Geopandas & lxml
- Command line program
- Modular structure: two foudamental operations
  1. Spatial Join -> Optimum Shapefile
  2. Creation 3D cityGML model



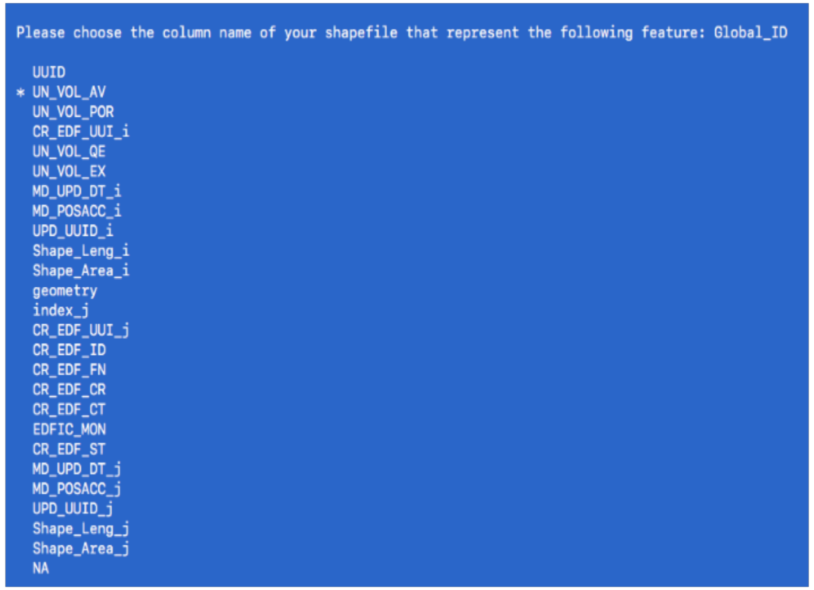

```
(shp2city) francescas-MacBook-Pro:virtualenv francescafissore$ shp2city -h
usage: shp2city [-h] output join_operation shapefiles [shapefiles ...]

Convert shapefile into cityGML.

positional arguments:
  output          path of the output citygml
  join_operation  type of join operation to apply between intersects, within
                  or contains
  shapefiles      shapefiles to be converted

optional arguments:
  -h, --help      show this help message and exit
(shp2city) francescas-MacBook-Pro:virtualenv francescafissore$
```

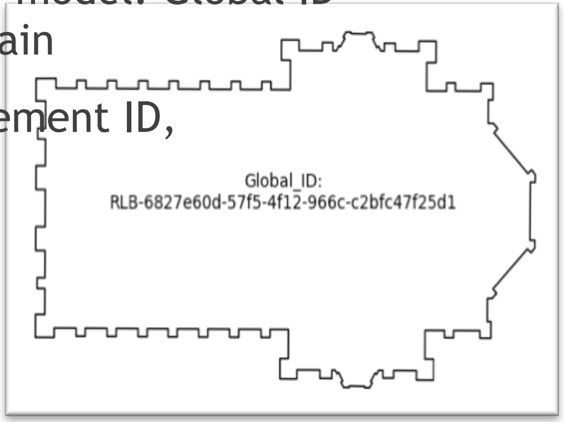
# Requests to satisfy

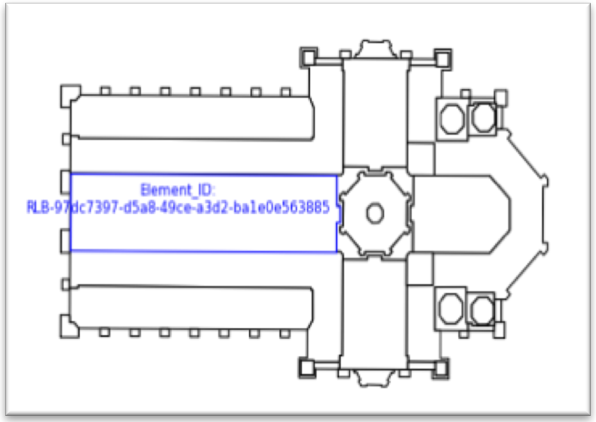
1. Output file name
2. Spatial relationship parameter : Intersects, Within, Contains
3. Paths input files

After the spatial join operation

- User should select fields of the input shapefiles where are stored features usefull to the creation of 3D model.
  - Mandatory features to crete 3D model: Global ID and Geometry height from terrain
  - Fetures to enrich 3D model: Element ID, Temporal Validity, Scale, etc



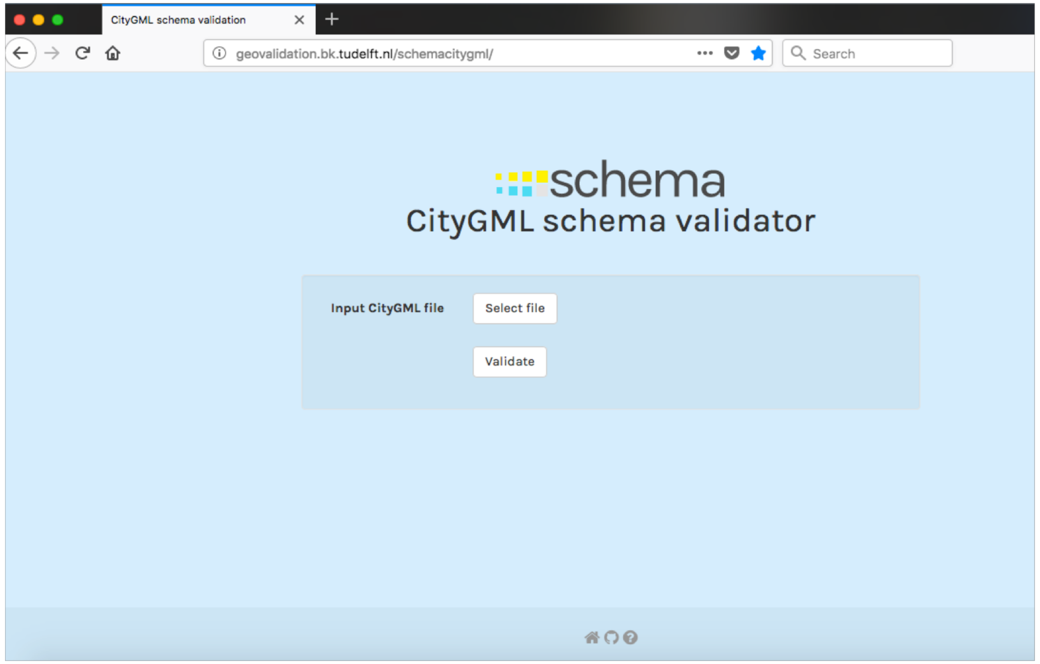



# Validation

- Two external Web tool by TU Delft are used to validate 3D cityGML model

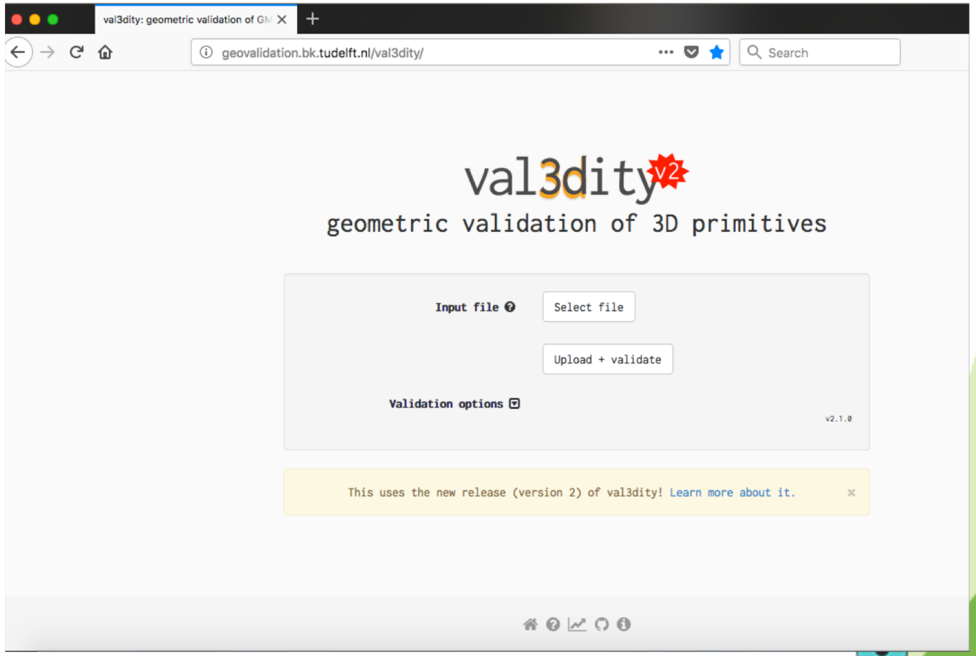
  - «Schema - CityGML Schema Validator»

    Ledoux, H., 2018. CityGML schema validation [WWW Document]. URL http://geovalidation.bk.tudelft.nl/schemacitygml/ (accessed 6.25.18).




  - «val3dity»

    Ledoux, H., 2013. On the Validation of Solids Represented with the International Standards for Geographic Information. *Computer-Aided Civil and Infrastructure Engineering,* 28, 693– 706. https://doi.org/10.1111/mice.12043



# Data-set: Milan

- Dataset used to developed the packages
- Shapefiles used to build citygml: two classes "**Unita' volumetrica**» **and** «**Edificato** « belonging to the theme Edificato
- Coordinate system WGS 84 / UTM zone 32N ( **EPSG:32632** )
- Ground level is assigned to all geometries of shapefiles
- Field I common between two layers that allowed to:
  - validate the results of spatial join operation (remove duplicate geometries)
  - assign global id and element id
- High number of geometries that overload the RAM: Writing in streaming mode to generate the xml file.

# Data-set Turin

- Shapefiles used to build citygml: two classed “**Unita' volumetrica**» **and** «**Edificato** « belong to the theme Edificato
- Coordinate system WGS 84 / UTM zone 32N (**EPSG:32632)**
- Fields in common between two layers
- !!! Ground level has been assigned zero value!!!

↓

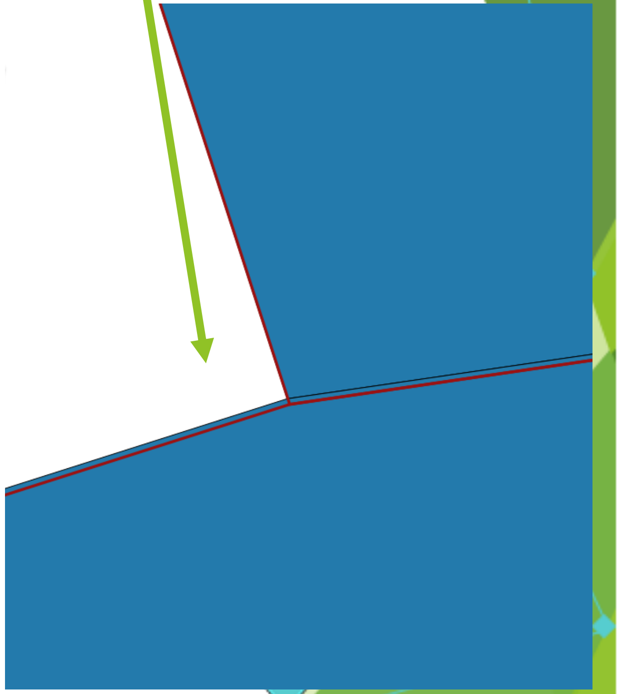
Use of a DTM generated by LiDAR to evaluated the ground level

# Data-set: Padua


- Shapefiles used to build citygml: two classed «**Unita' volumetr** «**Edificato**» belong to the theme Edificato
- Coordinate system: **Monte Mario / Italy zone 1 EPSG:3003 (co**
- Ground level assigned to all geometries of shapefiles
- !!! Not field in common between two layers: it is impossible assign a global id and satisfied a mandatory request !!!

A unique global id for each building was created ad hoc

- !!! Topological errors occur in the two shapefiles: wrong matching of geometries in the spatial join operation and then the presence of portions of a building divided by the main body !!!



- Performed a new spatial join operation: the geometry of the " **Unita' volumetrica**" whose centroid falls within the geometry present in the "**Edificato** " inherits its features.

# Data-set: Naples

- Data used to build citygml are downloaded by OSM: lower information then previous datasets.
- Coordinate system: WGS 84 / UTM zone 33N (**EPSG:32633)**
- !!! Not information regarded ground level and elevation of buildings (mandatory)!!!

↓

Generation of missing data by LiDAR (DTM and DSM)

## Data-set:Rome

# Missing Data

# Thanks for your attention!