

URBAN GEO BIG DATA

Maria A. Brovelli ^{1,3*}, Piero Boccardo ², Gloria Bordogna ³, Antonio Pepe ³, Mattia Crespi ⁴, Michele Munafò ⁵, Francesco Pirotti ⁶

¹ Politecnico di Milano, Department DICA, Politecnico di Milano, Italy, maria.brovelli@polimi.it

² Politecnico di Torino, Department DIST, Italy, piero.boccardo@polito.it

³ CNR IREA - Istituto per il Rilevamento Elettromagnetico dell'Ambiente, bordogna.g@irea.cnr.it; pepe.a@irea.cnr.it

⁴ Sapienza University of Rome, Rome, Italy, mattia.crespi@uniroma1.it

⁵ ISPRA - National Institute for Environmental Protection and Research, Rome, Italy, michele.munafò@isprambiente.it

⁶ CIRGEO - Interdepartmental Research Center of Geomatics, Department TESAF, University of Padova, Italy, francesco.pirotti@unipd.it

Commission IV, WG IV/4

KEY WORDS: Big Geodata, Urban, Infrastructure, Mobility, Soil Consumption, Displacement

ABSTRACT:

The paper deals with the general presentation of the Urban GEO BIG DATA, a collaborative acentric and distributed Free and Open Source (FOS) platform consisting of several components: local data nodes for data and related service Web deploy; a visualization node for data fruition; a catalog node for data discovery; a CityGML modeler; data-rich viewers based on virtual globes; an INSPIRE metadata management system enriched with quality indicators for each dataset.

Three use cases in five Italian cities (Turin, Milan, Padua, Rome, and Naples) are examined: 1) urban mobility; 2) land cover and soil consumption at different resolutions; 3) displacement time series. Besides the case studies, the architecture of the system and its components will be presented.

1. INTRODUCTION

Nowadays about 54% of world population lives in urban areas and, according to the 2014 UN-ESA report, this percentage is expected to increase up to 66% by 2050. We are clearly facing a rapid and global trend that will affect daily life in the next few decades.

Meanwhile, according to the well-known sentence “80% of data is geographic”, much of the data in the world are geo-referenced, usually in the global spatial reference frame connected to the Global Navigation Satellite Systems (GNSS), often called WGS84.

Geospatial data can be collected and analyzed using a variety of geomatics sensors and methodologies GNSS and terrestrial surveying, photogrammetry and remote sensing, laser scanning, mobile mapping, geo-located sensors, geo-tagged web contents, and Volunteered Geographic Information - VGI).

This is the why the efficient geospatial big data handling and integration is of key importance, in order to benefit of them as much as possible to managing the social and cultural change connected to the worldwide urban growth in a much more sustainable way, compared to what was done in the past. It is therefore timely and important to establish methodologies to evaluate the quality, extract and integrate the relevant information which can be supplied by this huge and continuously growing amount of geo data.

The overall goals and the detailed objectives of the Urban GEO BIG DATA project are just placed in this framework. Hereafter the research activities carried out with respect to the three pillar topics (urban mobility, land cover and soil consumption, displacement time series) of the project are

summarized, together with their main achievements up to now.

2. THE URBAN GEO BIG DATA PROJECT

The Urban GEOmatics for Bulk Information Generation, Data Assessment and Technology Awareness (Urban GEO BIG DATA) is a national (Italian) project (<http://www.urbangeobigdata.it/>) developed under the framework of projects of national interest of the Italian Ministry of Education, Universities, and Research (MIUR). The Principal Investigator of the project is Politecnico di Milano. It started in February 2017 and lasts three years.

The project aims to develop innovative Geographic Information System (GIS) methodologies and tools to exploit the integration of traditional geomatics data, Earth Observations (EO) and statistics data with new user-generated contents for promoting more effective management of urban resources and infrastructures.

The data-driven approach represents the first key concept in the methodology that was adopted. Management policies cannot be effective if they are not based on the detailed knowledge of the urban processes: a data-driven decisional approach is required to transform the huge amount of daily-generated data into useful information for decision-makers. Focusing on the spatial information, the data sources that enable us to monitor the status of the urban areas are very extensive: from topographic to passive geo-crowdsourced data (e.g. Twitter; see Molinari et al., 2018).

Another important point is that the data mining will be particularly focused on the valorization of the Open Data, which have been released in recent years from both national mapping and spatial agencies. Many of this data were collected and analyzed to understand their possible contribute

* Corresponding author

to enhancing the knowledge of the city. These steps involve an important challenge: the management of (geo) big data. In fact, many of the above-listed data meet most of the key traits that literature attributes to the big data, among these: volume (huge amount of data), velocity (real-time generated), variety (structured, unstructured) and veracity (messy/noisy data with uncertainty/error). Three main topics have been explored: mobility, soil consumption and displacement..

The selection of these topics was due to their temporal, spatial and semantic differences. Soil consumption and displacements must be studied in the long-term period and they are distributed on an “areal” dimension while mobility processes are temporally more limited and mainly represented in terms of linear graphs. Moreover, soil consumption and displacement are principally evaluated based on satellite image time series (mainly with multispectral images the former and with SAR data the latter); the mobility information is instead mainly derived from in situ sensor data, requiring completely different procedures and methods.

The three cases have been designed and implemented on the base of a Free and Open Source (FOS) distributed and acentric interoperable geospatial data infrastructure sharing on the Web the project’s multi-source heterogeneous geo-data. Virtual globes have been chosen as clients, instead of the more “classic” 2D viewers (Veenendaal et al, 2017), for the web visualization of the studied phenomena in five cities: Milan, Rome, Turin, Padua, and Naples.

The paper presents the outcome of the project after more than 2 years of activities. The remaining of the paper is organized as follow: at the beginning, the three case studies are discussed. Then the structure of the Urban GEO BIG DATA system is presented, starting from the adopted “urbanscape” data model and going to the description of the developed infrastructure on both server and client side.

3. CASE STUDY 1 - DATA: MOBILITY

Mobility is strictly connected to social and economic growth of urban areas. Urban transport allows more and more people to move, increasing the quality of life, but in the meantime generates serious problems that affect human activities and the environment. To cope with these problems the concept of “Sustainable Transport” has emerged in last years: the European Commission since 2006 has included the concept in its main objectives for the European development strategy (Council of the European Union, 2006). The Sustainable Transport concept aims to “ensure that our transport systems meet society’s economic, social and environmental needs whilst minimizing their undesirable impacts on the economy, society and the environment”.

To deal with transport and mobility issues, it is important to take into account the multidisciplinary nature of transport. Studies on mobility and transport traces its historical roots in geographic and spatial science, but nowadays the increasing capacity of producing and acquiring a large variety of spatial data, including flows and movements of passengers and freight, pose specific questions. Technologies as GIS and Intelligent Transportation System (ITS) are becoming key factors to manage this data flow. In the Urban GEO BIG DATA project, the first main research question concerned the integration of different kinds of data in a unique GIS spatial data model. This was approached and eventually solved by

modelling all the input data, by defining a comprehensive geodatabase, by tackling issues as: different data format and standards, data inconsistencies, different spatial and temporal granularities, different methods and enforcing rules that relates measures, events and physical infrastructures.

This interoperable data model has been used to test the delivery of all the principal mobility services (traffic conditions, routing, limited traffic area accessibility, public transportation management, etc.), comparing simulated data with real ones acquired by public agencies by means of a fixed sensor network. Considering actual situation in most European cities (where these kind of sensor network are not available at all), Floating Car Data (FCD) have been used. FCD are based on the collection of georeferenced data, by mean of GNSS receiver, inertial platforms, accelerometers and odometers, regarding speed, direction of travel and time information from On Board Unit (OBU) in vehicles that are being driven.

Using this data, different mobility services are going to be implemented: traffic condition (with an almost real time mapping generation as depicted in Figure 1), road graph impedances association (using calculated mean travel speed and times), routing preferences (to be considered as an input of a traffic balancing model) and eventually spatial origin/destination matrix as an innovative service for routing and planning operations (Figure 2). Hereafter, the methodology developed within the Urban GEO BIG DATA project to support the mobility services implementation is shortly summarized, with respect to FCD related to the Turin Municipality.

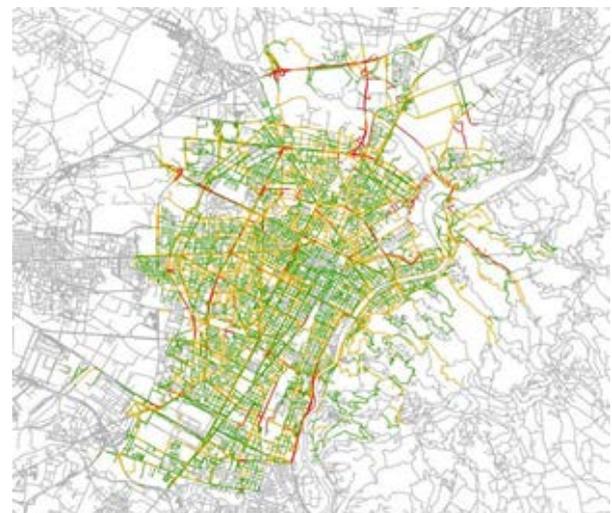


Figure 1. Vehicle density during morning rush hour estimated from FCD (Turin)

The largest part of movements in an urban environment is constrained to the road network. This is the why FCD, even if based on GNSS data only, represent a great wealth of movement information, especially if compared with more traditional traffic monitoring methods such as loop detectors or automatic plate number recognition (Ravanelli and Crespi, 2018), (Yang and Gidófalvi, 2018). These data contain the needed key information for estimating traffic impedance maps, also in real-time. GNSS sensors indeed track the position of vehicles at high rate (e.g. 1 second or more), so that their continuous movement is recorded as a trajectory containing a sequence of sampled points. Unfortunately, GNSS data are obviously corrupted by the noise generated by

the pretty well known problems affecting GNSS signal in urban environment (obstructions, multipath, etc.), so that proper processing methodologies have to be implemented to correctly extract the valuable information from the data.

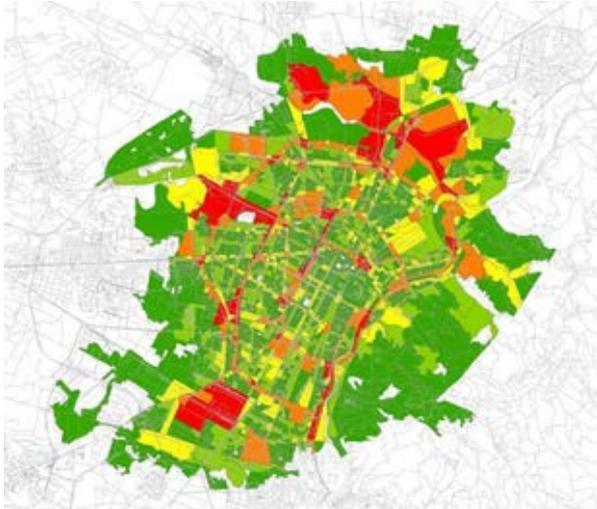


Figure 2. Census areas symbolized by total length of travel path crossing them. Legend: green color = lower values, red color = higher values (Turin)

In details, accurate traffic impedance maps can be estimated only if the raw GNSS trajectories can be correctly mapped to the road network. For this reason, a preliminary analysis was performed in (Ravanelli and Crespi, 2018), where a first strategy to process the FCD of the Turin Public Transportation system was implemented, in view of an automatic and possible real-time impedance map generation. The issues related to the management and visualization of such a huge amount of data were also addressed. Furthermore, some outlier removal methodologies were investigated (Pirotti et al, 2018) and a preliminary temporal analysis was carried out, showing higher velocities in weekend days and not peak hours, as could be expected. Finally, a methodology to assign the velocities to the line network topology, by projecting the raw FCD to the route lines, was developed and successfully tested.

4. CASE STUDY 2 - DATA: LAND COVER AND SOIL CONSUMPTION

Soil can be considered one of the most important non-renewable resource since its major role in many ecosystem functions (e.g. biomass production, filtration and transformation of many substances, habitat provision, cultural and historic function) and the slow process of formation (European Commission, 2011).

Land monitoring and land cover change data provide a description of the surface of the Earth by its biophysical characteristics, including the vegetation, bare soil, open bodies of water and artificial surfaces that can be observed by any earth observation platform and help to assess soil status and, particularly, soil loss due to its consumption (soil consumption or land consumption). Soil consumption refers to an increase of the artificial covering of the ground, due to settlement dynamics. It is defined as a change from a non-artificial to an artificial land cover (UN, 2015; ISPRA, 2018).

Some of the limitations of the existing land cover dataset are related both to the spatial resolution and to the classification adopted (Munafò et al., 2013). In most classification systems, a clear distinction between land use and land cover is lacking, giving rise to hybrid classifications (Sallustio et al., 2016). For example, in Corine Land Cover (CLC) products, land use and land cover are often confused within mixed classes and this data are not properly suitable to detect land consumption. Moreover, land cover classes, with a minimum mapping unit of 25 hectares, do not provide a reliable accuracy for a comparison of small territories (Decoville and Schneider, 2016) such as Italian municipalities.

Soil consumption maps produced within the Urban GEO project (in Figure 3 see, as an example, the map of Padua) are based on the semi-automatic identification of artificial land cover by exploiting the spatial and spectral resolutions of optical EO data as Sentinel-2 images integrated with radar Sentinel-1 data. In addition, ancillary data, such as regional topographic databases, VGI and in situ data, as well as manual photo-interpretation of high-resolution images were employed to improve the identification of minor artificial buildings, sealed areas, extraction sites, infrastructures, etc.

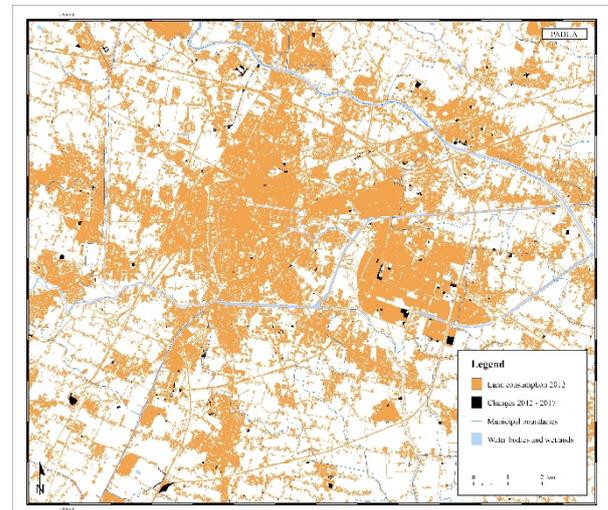


Figure 3. Soil consumption map of Padua (orange: soil consumption in 2012; black: soil consumption 2012 - 2017)

5. GENERATION OF GROUND DISPLACEMENT TIME-SERIES

The availability of huge archives of radar images acquired by the satellites operated over almost the last three decades has opened the possibility to detect and map, at a regular basis, the changes of the Earth's surface. In this framework, nowadays, the Differential Synthetic Aperture Radar (DInSAR) technique (Massonet and Feigl, 1998) represents a well-employed technique for computing the changes of the ground surface, with a centimeter-to-millimeter accuracy.

DInSAR basically operates by computing the phase difference between pairs of Synthetic Aperture Radar (SAR) images acquired over the same area at different times and from slightly different orbital positions. More recently, advanced multi-temporal DInSAR techniques have been developed to obtain time-series of the ground changes over time. This is accomplished by processing sequences of SAR images covering long time spans, and by properly combining the information derived from every single interferogram (i.e., the

phase difference between two SAR images).

Several multi-temporal DInSAR methods have been proposed in the literature. They can be globally categorized in the two main classes of the Permanent Scatterers (PS) (Ferretti et al., 2001) and of the Small Baseline (SB) (Berardino et al., 2002, Mora et al., 2003) methods. The former work at the scale of resolution of single pixels and are based on the discrimination of the deformation signals related to isolated, high-coherent, point-wise structures on the ground, which preserve the deformation information even when using large-baseline interferograms. On the contrary, the second group consists of methods based on the selection of a proper set of SB interferograms. DInSAR is nowadays a mature technology, and operational products, also routinely obtained using automatic procedures and exploiting cloud platforms, are available. At the same time, further efforts are still needed for improving the potential of existing multi-temporal DInSAR methods for: i) the estimation and removal of the atmospheric artefacts from the generated displacement time-series (Ferretti et al., 2001, Li et al., 2012), ii) the mitigation and filtering of noise, also using space-time approaches (Pepe 2019), iii) the combination of DInSAR products obtained with different SAR instruments (Samsonov et al., 2015). One of the most popular DInSAR approaches is the Small Baseline Subset (SBAS) algorithm (Berardino et al. 2002, Lanari et al. 2004), which allows to obtain spatially-dense mean displacement maps and time-series of deformation both in urban and rural areas.

Within the project, the interest is on studying the historical deformation phenomena that affected some urban areas in Italy. To the purpose, long-term displacement time-series from 1992 to 2011 have been obtained by applying the SBAS technique to sequences of archived SAR images collected by the ERS and ENVISAT satellites over the cities of the projects (Pepe et al., 2018).

6. MODELING THE URBANSCAPE: CITYGML

Modelling in three dimensions has been a natural follow-up to using 2D or 2.5D paradigm in geographical information systems. The importance of explicit 3D information becomes evident in the complexity of urban spaces. One of the objectives of the project was to analyse best solutions for 3D modelling of urban data. CityGML has been a widely used model for quite some time, and a natural choice for implementing such model in the project. It is worth noting that CityGML is an XML-based format (Yao et al., 2018) and recent work is focusing on moving towards CityJSON (Ledoux, 2018). CityJSON was at the initial stages when this investigation started, therefore CityGML was adopted as format for creating the 3D urban data of the five cities that are involved in the project. The foremost challenge is the typical extract transform and load (ETL) step, which means that existing 3D information, which is not in explicit 3D formats, is to be converted to CityGML.

Several tools are available that support this ETL procedure, but in this project a specific Python-based script was implemented for converting automatically cartographic data in ESRI Shapefile format to CityGML. An online web portal developed in R language and several packages (e.g. Shiny for the web application, Geos for spatial processing) for uploading data and following steps for spatial joins that are required for adding attributes that are available in the CityGML schema (Fissore and Pirotti, 2018).

The web portal is called ZCityGML and the backend Python programme for ETL procedure for processing the ESRI Shapefile to CityGML is called shp2city and is available online in Bitbucket with GNU GPLv3 license. Figure 4 depicts the steps for the procedure from top to bottom; the intermediate step, which uses the Python-based shp2city module, is automated from the input inserted by the user in the web portal.

The output is a CityGML file with LOD1 building objects. The CityGML file can be downloaded by the user directly, or easily integrated in a more structured framework that imports in more robust database systems such as Postgresql supporting further extensibility that is inherent to the CityGML model (Biljecki et al., 2018). As mentioned, LOD1 is extracted from cartography, but higher levels can be reached if other sources are used, such as LiDAR data or derived products (digital terrain and surface models). An exploratory investigation was done in this sense and presented in Fissore and Pirotti, (2019); results support existing literature that shows how building footprints from digital cartography and height information from 3D data sources like LiDAR can be integrated successfully to determine LOD2 details theoretically. The main difficulty is geometric interpretation of the sampled 3D information, which requires simplification into primitives which is not always a straightforward process

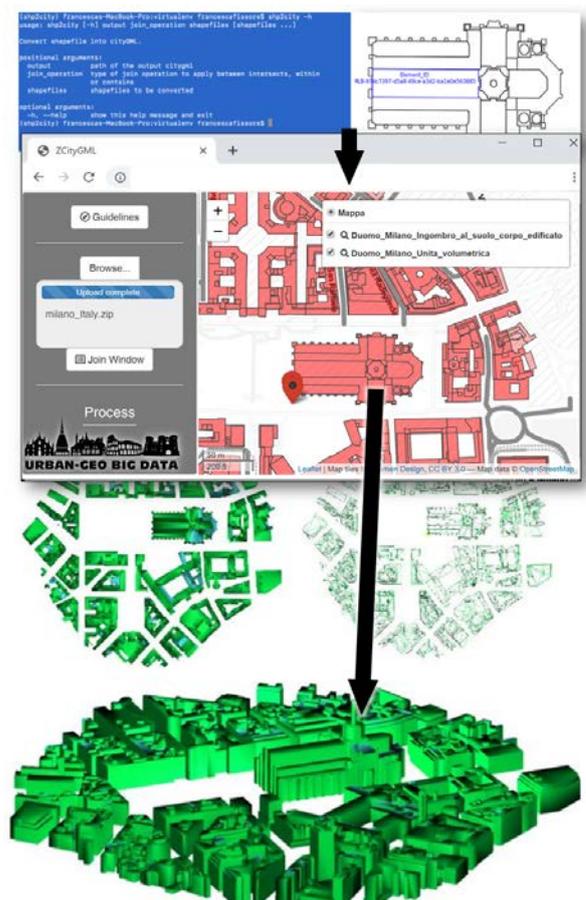


Figure 4. ETL procedure for converting ESRI Shapefiles to a CityGML Model; the area is the central square of the city of Milan (Fissore and Pirotti, 2018).

7. THE INFRASTRUCTURE OF URBAN GEO BIG DATA

Spatial Data Infrastructures (SDI) are complex systems in which a network of interconnected but autonomous components is used for the exchange and mobility of geospatial information. Their successful exploitation requires technologies, policies, common standards and harmonised rules. The INSPIRE Directive (Directive, 2007) has fostered the creation of a European SDI that delivers integrated spatial information services to the users. Following the Directive, components of European SDIs include metadata, spatial data themes, spatial data services; network services and technologies; agreements on data and service sharing, access and use; coordination and monitoring mechanisms, processes and procedures. These services should allow users to discover and possibly access spatial or geographical information from a wide range of sources, from the local to the global level, in an interoperable way for a variety of uses. To implement such an SDI for multi-source geospatial data sharing on the Web effectively and efficiently, in compliance with interoperable OGC standards, an acentric distributed SDI architecture has been conceived with multiple end point nodes.

The choice of a distributed architecture offers the advantages of the more flexible and low-cost solution with respect to the centralized one. In fact, as data increase it is possible to perform horizontal scaling by adding new hardware resources and new end point nodes when needed, while a centralized solution would require scaling up the hardware which generally is more costly.

Furthermore, in a distributed SDI multiple concurrent requests to data of different themes can be addressed to the appropriate data node by applying a distributed information retrieval approach (Bordogna et al, 2003), thus allowing an optimization of the network traffic and a more efficient management of the requests, which in a centralized solution might cause inefficiencies and bottle-necks when trying to concurrently access data managed on the same server.

Finally, an acentric solution with several independent data sharing nodes for geospatial data maintenance and sharing is also more respectful of both independence and ownership of the data providers. Specifically, within the Urban GEO BIG DATA project several nodes are deemed for the geospatial data sharing (data nodes), which have been set up by partitioning thematically the physical organization of data.

This choice allows each project's unit to maintain locally its own data for which it is responsible and for which a local expertise is available for their creation, organization and updating. To this end, each data sharing node hosts a geo Web server that manages thematically homogenous geospatial data related to whole spatial domain of interests, i.e., the five cities. In such a way, for example, all information about to the characterization of the long-term behavior of the deformation mechanisms in all the urbanized areas is managed and shared by a single end point of the SDI.

Data in each end point node are organized according to a suitable common data model deemed appropriate for representing a given theme in compliance with OGC standards. For example, all information related to soil consumption and mobility are organized according to their specific agreed common data models and are shared on the web by distinct end point nodes of the SDI. In this context,

standard data models trying to represent the city as a whole (such as the certificated OGC standard CityGML) are applied.

In order to offer users the possibility to inquire what geospatial data products are made available by the project a common end point node of the SDI provides the discovery facility by enabling the formulation of content and spatial queries on the metadata of the single geospatial data products and services.

Finally, an end point node enables the consultation, visualization and analysis of all the distributed geospatial information, by accessing remotely the requested data in a transparent way to the final user who is not requested to be aware of the location of the accessed data.

To implement the SDI, the use of FOS software (Brovelli et al, 2017) is another key aspect of the project to ensure the interoperability, replication and reutilization of the applications. Last but not least, since the targeted users are public administrations, by the adoption of FOS the project aims to demonstrate both a best practice and a feasible low-cost solution for managing and sharing on the Web geospatial information in compliance with governmental standards required to public administrations.

Interoperable standards are used for the development of geoservices by exploiting OGC standards such as WMS (Web Map Service), WMTS (Web Map Tiles Service), WFS (Web Feature Service), WCS (Web Coverage Service) and WCPS (Web Coverage Processing Service), CSW (Catalogue Service for the Web), the OGC best practice for implementing WMS with time support (WMS-T), and, considering the new researches currently addressed on these themes.

In order to ease the sharing of geospatial data, the GET-IT platform developed within the RITMARE Italian flagship project has been used by some units (Lanucara et al., 2017).

GET-IT is designed as an interface between the user and GeoNode and Geoserver FOS tools to allow the ingestion, storage, display and sharing on the Internet of different types of geographical data, whether digital images, processed maps, data acquired from sensors (observations) or georeferenced textual documents. This is the first FOS tool that allows integrated management for these different categories, providing the users with a single working interface allowing personalized protection of the shared data.

8. METADATA AND CATALOGUE SERVICE

One key aspect of a SDI is the facility for enabling users to get knowledge of the available data and their characteristics, and the contacts and policy for accessing and possibly downloading data. This facility known as discovery is performed through the use of a Catalogue service by users having information needs, expressed by queries. A catalogue service exploits a Web database management system to manage metadata, i.e., information about geospatial data, and to answer users' queries. Metadata describe the characteristics of the geospatial format, authorship and ownership, semantics, geographic and temporal context, and ancillary information on their quality, which is necessary to allow consumers and stakeholders to interpret the data meaning and to understand if the correspondent data fits their information needs.

Generally catalogue services follow standards as far as the format and structure of the metadata for which the INSPIRE directive and National authorities have defined interoperable

schemas (Italian authority has defined the RNDT metadata schema that is an extension of the INSPIRE schema for encoding metadata of catalogue service in charge of public authorities). Catalogue service can be implemented through some products (either proprietary or not) that declare their compliance to such metadata schemas. Users' conditions, expressed through discovery service clients, are matched against archived metadata, describing associated geo-data, in order to give an answer to the above conditions. The matching result is a list (sometimes empty) of records of metadata satisfying the match.

The bottleneck in setting up a catalogue service is the need to perform the burdensome activity of manually creating metadata by data providers or authors. This task must be carried out each time a new data set is published in a SDI, and it must be periodically performed in the case of time series of geospatial data.

To address this activity, the European Open Source Metadata Editor (EUOSME) has been developed to help create metadata compliant with the INSPIRE Directive (2007/2/EC) and the INSPIRE Metadata Regulation (1205/2008). Nevertheless this editor is bounded to a single metadata schema. As part of the GET-IT platform, the metadata editor EDI is available as FOS to guide users in the creation of metadata compliant with several standard schema: once created GET-IT automatically deploys both the CSW (Catalogue Service for the Web) through Geonode tool and data and relative services through Geoserver Web GIS server.

EDI is a template-driven metadata authoring tool that can be easily customized to any XML-based metadata format and to a specific workgroup, institute, or project. Within the Urban GEO BIG DATA project, EDI has been used by some units to create the metadata compliant with RNDT schema. One important facility offered by EDI is the availability of domain thesauri and information about the projects, such as the units and staff profiles which can be selected to fill the metadata specific fields. Nevertheless, in the case of the displacement time series, when metadata have to be created mainly with the same contents for each city and at each update, a semi-automatic workflow for Web deployment of both time series and the creation and updating of metadata have also been performed.

Finally, metadata are periodically harvested by the discovery node (<http://ugbd-catalog.get-it.it/geonetwork>) that hosts a Catalogue service implemented by the FOS Geonetwork, that has been customized to index and manage RNDT metadata.

9. DATA SERVICES

In recent years, a consistent debate has been raised in relation to geospatial data and related services. In 2017, the National Geospatial Advisory Committee (NGAC) received guidance from the Federal Geographic Data Committee (FGDC) provided forward-looking advice and guidance to maximize the value and utility of Data-as-a-Service (DaaS) as a delivery mechanism for Federal data repositories (e.g., GeoPlatform, Data.gov, etc.).

In the Urban GEO BIG DATA project, the concept of DaaS has been implemented, in order to guarantee the main benefits that should be provided by DaaS: data availability from one common location and a single point of contact to manage, maintain, and update any given data source, reduced management and delivery costs, datasets easier to manage,

secure, and maintain, support for scientific research and innovation, and eventually support to National geospatial data repository in an interoperable data format.

Taking into consideration these goals and the requirement to manage large volumes of data, including data generated by sensors, the infrastructure has to be designed and configured in order to effectively and automatically manage the insert of several package of data in parallel (or the bulk insert of large chunks of data): as an example, in the case of Floating Car Data (FCD), several concomitant data package transmitted through the communication network must be handled and transformed into entries in the geodatabase.

The completion of the inserting process should then automatically trigger highly optimized procedures, capable to process and generate updated outputs without saturating computing resources. Finally, spatial data services are responsible for sharing data and their representation in standard formats, searchable and discoverable by means of catalogue services and exploitable by desktop or web-based clients. In Figure 5 an example related to vehicle fluxes representation in the mobility domain, is presented.

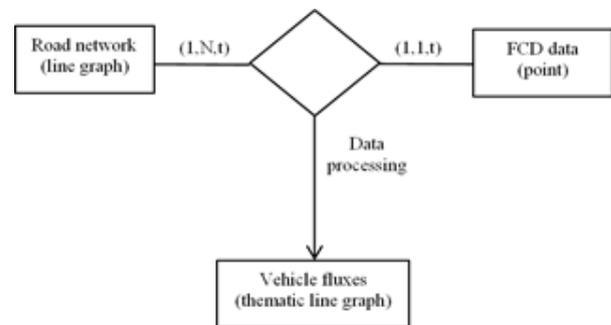


Figure 5. Example of a geodatabase embedded processing service

The implemented solution could also allow different possible advantages (as reported also by NGAC):

1. legality and licensing, since all geospatial datasets included and/or produced in the framework of this project are available under an Open Data license;
2. data description, access, origin, etc. structured in standard and accessible metadata;
3. efficient data search and discovery, as reported in the previous chapter, based on metadata content;
4. versioning capabilities (DBMS inheritance);
5. interoperable data access (well describe data model, standard data formats and data access services);
6. trustworthiness of data (origin certification) and pre-calculated operational dataset (e.g. in the case of the mobility domain) ready to use into operational services.

10. THE 3D CLIENT

Common web mapping is still based on two-dimensional clients, specifically when big data are involved in the visualization. On the opposite, within the Urban GEO BIG DATA project, the choice was that of developing multidimensional viewer based on existing open source virtual globes. Among the different available systems, the most mature are at the moment NASA World Wind (Brovelli and

Zamboni, 2012; Brovelli et al, 2013; Valentini et al, 2014; Brovelli et al, 2016; Kilsedar et al, 2019) and CesiumJS.

NASA World Wind has been available as open source since 2003 in C#, since 2006 in Java, since 2012 in Android, since 2013 in iOS and since 2014 in JavaScript (Pirotti et al, 2017). The Javascript implementation allows the development of interfaces which, being accessible directly via a web browser, are easier and most immediately usable by common users. Moreover, the API is being enriched with functionalities from the Java version and this makes it mature and performant.

CesiumJS, started as a project in 2011, is an open source Javascripts library that offers cross-platform visualization and exploration of 3D geographic data on Web. Among the tools that can visualize data in CityGML or formats that can be converted to CityGML on Web, CesiumJS surpasses the rest as it supports various spatial reference systems, standard data formats, and OGC standards; streams terrain; has means to define interaction; uses WebGL which provides hardware-accelerated graphics and as a result improves the performance significantly while handling large complex 3D city models.

The decision of the team was to test the functionalities of the two products, using case by case the most efficient. The client is therefore an hybrid one, where services are offered with one of the two virtual globes. Figures 10 and 11 show two visualisation, the former of OSM data with NASA World wind, the latter of ground displacement based on CesiumJS.



Figure 10. The Urban GEO BIG DATA client: OpenStreetMap data visualisation with NASA World Wind (Naples)



Figure 11. The Urban GEO BIG DATA client: ground displacements with Cesium JS (Naples)

11. CONCLUSIONS

The future will see a growing amount of spatial data available for processing. The presented work reports the efforts necessary for collecting, integrating and sharing spatial

information through open tools, providing access to researchers and other stakeholders that can create added value to the data. The bottom line is to support the key process of converting spatial data to spatial information, which is of foremost importance in present and future geomatics approaches towards big data.

ACKNOWLEDGEMENTS

This research is supported by the Italian Ministry of Education through the project of national interest (PRIN) Urban-Geo Big Data (project code: 20159CNLW8 - PE10).

REFERENCES

- Berardino, P., Fornaro, G., Lanari, R., Sansosti, E., 2002. A new algorithm for surface deformation monitoring based on small baseline differential SAR interferograms. *IEEE Trans. Geosci. Remote Sens.*, 40, pp. 2375-2383.
- Biljecki, F., Kumar, K., Nagel, C., 2018. CityGML Application Domain Extension (ADE): overview of developments. *Open Geospatial Data, Softw. Stand.*, 3, 13. doi:10.1186/s40965-018-0055-6
- Bordogna G., Pasi G., Yager R., 2003. Soft Approaches to distributed information Retrieval. *Int. Journal of Approximate Reasoning*, 34(2-3), pp. 105-120.
- Brovelli, M. A., Zamboni, G., 2012. Virtual Globes for 4D Environmental Analysis. *Applied Geomatics*, 4(3), pp. 163–172.
- Brovelli, M. A., Hogan, P., Minghini, M., Zamboni, G., 2013. The Power of Virtual Globes for Valorising Cultural Heritage and Enabling Sustainable Tourism: NASA World Wind Applications. *Int. Arch. Photogramm. Remote Sensing and Spatial Information Sciences*, XL-4/W2, pp. 115–120.
- Brovelli, M. A., Kilsedar, C. E., Zamboni, G., 2016. Visualization of VGI Data Through the New NASA Web World Wind Virtual Globe. *Int. Arch. Photogramm., Remote Sensing and Spatial Information Sciences*, XLI-B4, pp. 205–209.
- Brovelli, M. A., Minghini, M., Moreno-Sanchez, R., Oliveira, R., 2017. Free and Open Source Software for Geospatial Applications (FOSS4G) to Support Future Earth. *International Journal of Digital Earth*, 10(4), pp. 386–404.
- Council of the European Union, 2006. Review of the EU Sustainable Development Strategy (EU SDS) - Renewed Strategy. <http://register.consilium.europa.eu/doc/srv?l=EN&f=ST%2010117%202006%20INIT> (29 May 2019)
- Decoville A., Schneider M., 2016. Can the 2050 zero land take objective of the EU be reliably monitored? A comparative study. *Journal of Land Use Science*, 11(3), 331–349.
- Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 (INSPIRE), online: <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32007L0002>, 2019.
- European Commission, 2011. Communication from the Commission to the European Parliament, the Council, the

- European Economic and Social Committee and the Committee of the Regions. Roadmap to a Resource Efficient Europe. COM/2011/0571 final.
- Ferretti, A., Prati, C., Rocca, F., 2001. Permanent Scatterers in SAR Interferometry. *IEEE Trans. Geosci. Remote Sens.*, 39, 8-20.
- Fissore, F., Pirotti, F., 2018. Migration of digital cartography to CityGML; a web-based tool for supporting simple ETL procedures. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, XLII-4, pp. 193–200. doi:10.5194/isprs-archives-XLII-4-193-2018
- Fissore, F., Pirotti, F., 2019. DSM and DTM for extracting 3d building models: advantages and limitations. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, XLII-4.
- ISPRA, 2018. Consumo di suolo, dinamiche territoriali e servizi ecosistemici - Edizione 2018. Rapporti 288/2018.
- Kilsedar, C. E., Fissore, F., Pirotti, F., Brovelli, M. A., 2019. Extraction and Visualization of 3D Building Models in Urban Areas for Flood Simulation. *Int. Arch. Photogramm., Remote Sensing and Spatial Information Sciences*, XLII-2/W11, pp. 669–673.
- Lanari, R., Mora, O., Manunta, M., Mallorqui, J.J., Berardino, P., Sansosti, E., 2004. A Small baseline approach for investigating deformation on full resolution differential SAR interferograms. *IEEE Trans. Geosci. Remote Sens.*, 42, pp. 1377-1386.
- Lanucara S., Oggioni A., Modica G., Carrara P., 2017. Interoperable Sharing and Visualization of Geological Data and Instruments: A Proof of Concept, *Computational Science and Its Applications*, – ICCSA 2017, Gervasi O. et al. (eds) . *Lecture Notes in Computer Science*, 10407 Springer, pp. 584-599.
- Ledoux, H., 2018. val3dity: validation of 3D GIS primitives according to the international standards. *Open Geospatial Data, Softw. Stand.*, 3, 1. doi:10.1186/s40965-018-0043-x
- Li Z. W. Li et al., Correcting atmospheric effects on InSAR with MERIS water vapour data and elevation-dependent interpolation model, *Geophys. J. Int.*, vol. 189, no. 2, pp. 898–910, May 2012.
- Massonet, D., Feigl, K.L., 1998. Radar interferometry and its application to changes in Earth's surface. *Rev. Geophys.*, 36, pp. 441-500.
- Molinari, M.E., Oxoli, D., Kilsedar, C.E., Brovelli, M. A. (2018). User Geolocated Content Analysis for Urban Studies: investigating mobility perception and hubs using Twitter. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-4, 439-442, doi: 10.5194/isprs-archives-XLII-4-439-2018
- Mora, O., Mallorqui, J. J., Broquetas, A. 2003. Linear and nonlinear terrain deformation maps from a reduced set of interferometric SAR images. *IEEE Trans. Geosci. Remote Sens.*, 41, 2243-2253.
- Munafò M., Salvati L., Zitti M., 2013. Estimating soil sealing rate at national level - Italy as a case study. *Ecological Indicators* 26/2013:137–140.
- Pepe, A., Bonano, M., Bordogna, G., Brovelli, M.A., Calo, F., Carrara, P., Congedo, L., Frigerio, L., Imperatore, P., Lanari, R., Lanucara, S., Manzo, M., Munafò, M., 2018. The “Urban Geomatics for Bulk Information Generation, Data Assessment and Technology Awareness” project: detection, representation and analysis of the urban scenario changes. In: *Proceeding of IEEE International Geoscience and Remote Sensing Society Symposium*, July 22-27, 2018. Valencia, Spain
- Pepe A, 2019, Theory and Statistical Description of the Enhanced Multi-Temporal InSAR (E-MTInSAR) Noise-Filtering Algorithm, *Remote Sens.*, 11 (3), pp. 363.
- Pirotti, F., Brovelli, M. A., Prestifilippo, G., Zamboni, G., Kilsedar, C. E., Piragnolo, M., Hogan, P., 2017. An Open Source Virtual Globe Rendering Engine for 3D Applications: NASA World Wind. *Open Geospatial Data, Software and Standards*.
- Pirotti, F., Ravanelli, R., Fissore, F., Masiero, A. 2018. Implementation and assessment of two density-based outlier detection methods over large spatial point clouds. *Open Geospatial Data, Software and Standards*, 3(1), 14. doi: 10.1186/s40965-018-0056-5.
- Ravanelli, R., Crespi, M. 2018. Analysis of the Floating Car Data of Turin Public Transportation system: first results. *The Int. Arch. Photogramm., Remote Sensing and Spatial Information Sciences*, XLII-4, pp. 515 – 521. doi: 10.5194/isprs-archives-XLII-4-515-2018.
- Sallustio L., Munafò M., Riitano N., Lasserre B., Fattorini L., Marchetti M., 2016. Integration of land use and land cover inventories for landscape management and planning in Italy. *Environmental Monitoring and Assessment* 188:1-20.
- Samsonov, S.; d’Oreye, N. Multidimensional time-series analysis of ground deformation from multiple InSAR data sets applied to Virunga Volcanic Province. *Geophys. J. Int.* 2012, 191, 1095–1108.
- UN, 2015. *Transforming our World: The 2030 Agenda for Sustainable Development*, A/RES/70/1, United Nations.
- Valentini, L., Brovelli, M. A., Zamboni, G., 2014. Multi-frame and Multi-dimensional Historical Digital Cities: the Como Example. *International Journal of Digital Earth*, 7(4), pp. 336–350.
- Veenendaal, B., Brovelli, M. A., Li, S. (2017). Review of Web Mapping: Eras, Trends and Directions. *ISPRS International Journal of Geo-Information*, 6(10),317. doi: 10.3390/ijgi6100317.
- Yang, C. and Gidófalvi, G., 2018. Mining and visual exploration of closed contiguous sequential patterns in trajectories. *International Journal of Geographical Information Science* 32(7), pp. 1282–1304.
- Yao, Z., Nagel, C., Kunde, F., Hudra, G., Willkomm, P., Donaubaer, A., Adolphi, T., Kolbe, T.H., 2018. 3DCityDB - a 3D geodatabase solution for the management, analysis, and visualization of semantic 3D city models based on CityGML. *Open Geospatial Data, Softw. Stand.*, 3, 5. doi:10.1186/s40965-018-0046-7.