1 An innovative webGIS system for dissemination and visualization of official statistics and geospatial analysis

Emanuela Recchini, Alessandro Capezzuoli ([1])

1.1. Introduction

Statistical information is essential for decision makers, politicians, researchers, journalists and the general public. They need relevant, reliable and even more timely data for their work.

Official statistics are best suited to meet this need. In fact, data that are produced under the statistical programme of a given country or an international organization are characterized by the highest quality possible. Not only are quality criteria met for the data themselves but also strict conditions concerning processes are in place, following the United Nations Fundamental Principles of Official Statistics and the European Statistics Code of Practice.

ESS Vision 2020 emphasizes the importance to complement existing data derived from traditional sources with newer emerging ones, including geospatial and where possible big data. It is also stressed the relevance and the opportunity of sharing tools within the ESS and the usefulness of proper dissemination channels that meet the needs of as many people as possible.

In this respect, it is evident how crucial the role played by National Statistical Offices (NSOs) may be in developing suitable tools for the integration of different types of information and the dissemination of statistical data.

In the present chapter, after an overview of main formats, standards and methods used for data dissemination, data visualization issues and webGIS systems are considered, with particular reference to Istat experience. A specific focus is put on StatVIEW, a tool which has been created for purposes linked to this topic and can be conveniently shared among NSOs; concrete examples are shown to highlight its possible uses. Some concluding remarks emphasize the main features of StatVIEW.

[1] Italian National Institute of Statistics (Istat).
The views expressed in this paper are solely those of the authors and do not involve the responsibility of Istat. Addresses for correspondence: emanuela.recchini@istat.it and alessandro.capezzuoli@istat.it.

1.2. Dissemination and visualization of official statistics

1.2.1. Main formats, standard and methods for data dissemination

At present, one challenge for NSOs is the digital transformation going on world-wide. The ever increasing availability of data – facilitated by advances in information technology and the Internet – requires the implementation of methods and tools for the proper assessment, interpretation and dissemination of the same data.

Dissemination of high quality data gives users the opportunity to compare and integrate statistical information from different sources to best analyze the phenomena under investigation.

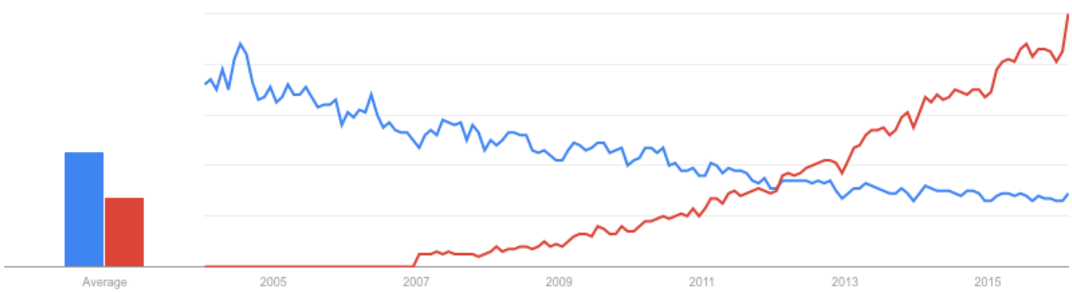
Before the creation of web-sites, the main channel of data dissemination was publication on paper. Now, thanks to Internet, the speed at which information travels and the evolution of tools available to users (PCs, tablets, smartphones, apps) have redefined the way of data dissemination and utilization. The web regularly provides up-to-date data: this is a crucial point that distinguishes this type of dissemination from paper publication.

Solicited by ESS, NSOs have started exploring new technology-driven areas when dealing with very large amount of data and are developing suitable tools for the integration of different types of data and its dissemination, i.e. platforms for data storage, analysis and visualization.

There are several formats, standards and methods used for data dissemination on the web. In this paragraph, two open formats regularly in use are considered: XML (eXtensible Markup Language) and JSON (JavaScript Object Notation). They are both defined as technical standard by the World Wide Web Consortium (W3C), an international community intended to realize the full potential of the Web.

XML defines a set of rules for encoding documents in a format that is both human-readable and machine-readable. JSON provides a format that uses human-readable text to transmit data objects consisting of attribute–value pairs. It is speedy and easy to understand and use. In recent years, thanks also to web 2.0 and the considerably increasing popularity of the JavaScript language, the use of JSON format has largely replaced XML. The graph below shows the interest of users in textual search of ‘Json API’ (red line) versus ‘XML’ (blue line) on Google (Figure 1).

Figure 1: XML vs JSON (users' interest) – Years 2005-2015



Source: Google Trend.

Being a JavaScript Object Notation, JSON is the natural environment for the construction of web applications using programming languages (like Javascript, Ajax, PHP, JSP, PYTHON) combined with *ad hoc* frameworks as appropriate (e.g., Node.js, OpenLayers, Leaflet, Raphael JS).

Different purposes may be served by the available frameworks: examples are graph construction, database management, asynchronous calls to external resources, cartographic applications, etc.

OpenLayers and Leaftlet are the main frameworks used for cartographic applications. The latter, in particular, provides a set of powerful and complete tools and plugins for data integration and the use of web resources.

JSON and XML differ in the type of architecture used: the former uses a resource-oriented architecture, while the latter uses a service-oriented one.

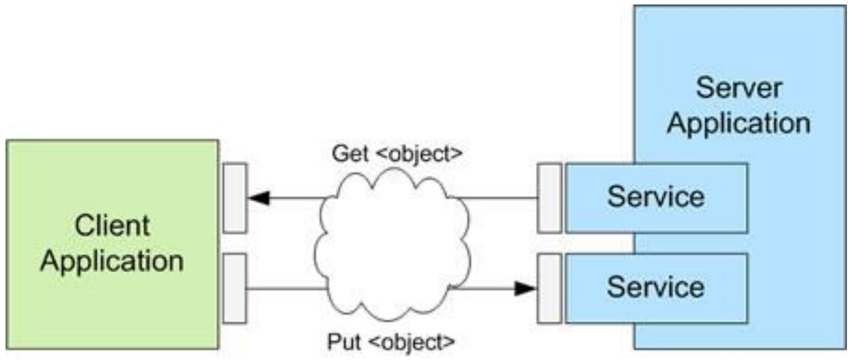
It is worth noting that the protocol used to uniquely identify (through a URL, Uniform Resource Locator) and then accessing web resources is the HTTP (HyperText Transfer Protocol). REST (Representational State Transfer) defines a set of guidelines to use the HTTP protocol in order to perform four operations summarized in the acronym CRUD (Create, Read, Update, Delete), by means of API (Application Programming Interface). See figures 2 and 3.

Figure 2: HTTP methods vs CRUD operations

HTTP METHODS	CRUD OPERATIONS	DESCRIPTION
POST	CREATE	Create or add a new resource
GET	READ	Read, retrieve, search, or view existing resources
PUT	UPDATE	Update or edit existing resources
DELETE	DELETE	Delete/deactivate/remove existing resources

An API is a web resource that is based on a client server architecture. It provides a service when called. In this case, the service is represented by the data obtained. Figure 3 suggests how an API interface with REST architecture allows users to work in a simple and well-ordered way.

Figure 3: REST API Architecture



1.2.2. Data visualization and webGIS systems

In the last few years, there has been a rapid increase of languages, frameworks and tools for the web, allowing users wide opportunities for cartographic visualization and geospatial analysis of statistical data. In order to produce simple density maps or dashboard containing tables and graphs, many commercial and open source tools have been developed (for example, Tableau, Microstrategy, Spago BI, etc.). However, these tools do not often take advantage of the potential of geo-referenced data, simply providing users with little more than infographics.

WebGIS systems are more advanced tools, allowing users to satisfy their different needs. This is because they enable not only data dissemination and cartographic visualization, but also dynamic territorial analyses. WebGIS systems are very useful to produce *ad hoc* applications for real-time and time series analyses, monitoring of specific phenomena such as earthquakes, urban mobility, environmental pollution, as well as data integration using REST (or SOAP) web services and standards like JSON-stat.

1.2.3. Istat experience

The usability of data is related to the format (XML, JSON, CSV, etc.) and the model adopted for their representation. At present, the Italian National Institute of Statistics (Istat) uses different channels for data dissemination and visualization. The main ones are:

- The single exit point for exchanging and delivering data to Eurostat and other Institutions through SDMX format;
- I.Stat, the data warehouse of statistics currently produced by Istat, where data are organized in a coherent and homogeneous way and are constantly upgraded. I.Stat permits data downloading in several formats (csv, excel, SDMX, etc);
- A set of API REST to disseminate data in the JSON-stat format.

SDMX (Statistical Data and Metadata eXchange) is an international standard providing: a logical model to describe statistical data, together with guidelines on how to structure the content; a standard for automated communication machine to machine; a technology supporting standardized IT tools that can be used by all parties involved in the data exchange and processing[2]. In recent years, other models have appeared, in addition to SDMX (for example JSON-stat, DDI, etc.).

Istat data warehouse is conceived on the basis of the multi-dimensional array of data OLAP Cube (On-Line Analytical Processing). An OLAP cube is a method of storing data in a multidimensional form, generally for reporting purposes. The faces of the cube contain the dimensions – i.e. the dimensional spreadsheets containing the descriptive attributes of the measures (region, province, gender, date of birth, etc.) –, while each cell contains the value related to the intersection of dimensions. The arrangement of data into cubes overcomes a limitation of relational databases, which are not well suited for near instantaneous analysis and display of large amounts of data.

The cube model is at the basis of the JSON-stat format, which is a JSON-based data cube packaging format. The JSON-stat format is a simple lightweight JSON format for data dissemination. It is based on a cube model that arises from the evidence that the most common form of data dissemination is the tabular form. In this cube model, datasets are organized in dimensions, which in turn are organized in categories[3].

1.3. StatVIEW: a webGIS system for data visualization, dissemination, monitoring and geospatial analysis

1.3.1. What is StatVIEW?

StatVIEW (http://www.statview.eu) is an innovative tool which can conveniently support whatever analysis focused on one or more subject areas covered by official statistics (social, demographic, economic, environmental, etc.). StatVIEW uses JSON-stat API for the dynamic construction of density maps based on data from different sources. It is a user-friendly webGIS system, developed

[2] SDMX is an international initiative that aims at standardizing and modernizing the mechanisms and processes for the exchange of statistical data and metadata among international organizations and their member countries. For further details, see: https://sdmx.org/?page_id=3425 and http://ec.europa.eu/eurostat/data/sdmx-data-metadata-exchange.

[3] For further details on JSON-stat, see: https://json-stat.org/.

by making use of open source technologies; it permits to gather, link, standardize, disseminate and visualize statistical data in different formats.

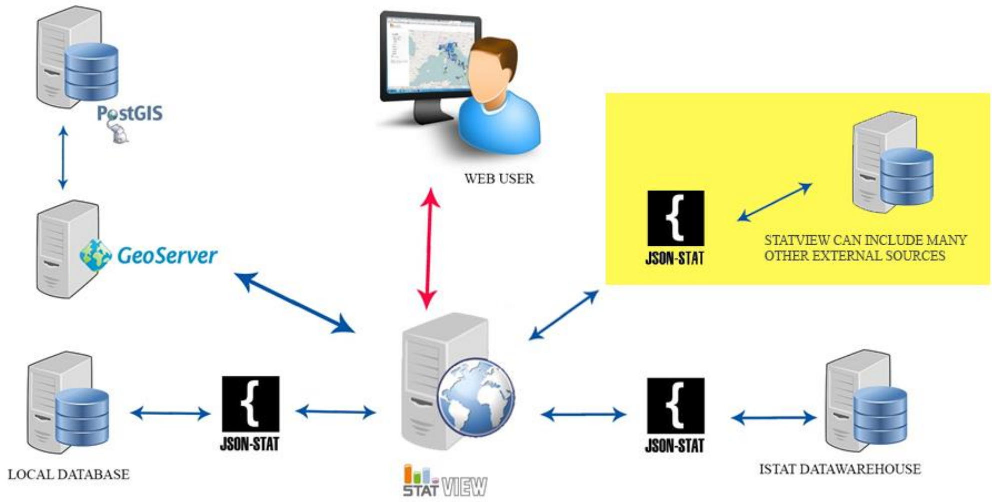
Similarly to data on population, health, education, labour, environment, national accounts and social security and welfare produced and provided by Istat currently visualized through StatVIEW, data produced by NSOs, research institutes and government agencies, available according to different geographical units, can be disseminated through the web using web services complying with the JSON-stat standard or other data exchange standards (i.e., SDMX, OData); also data other than official statistics can easily enter the webGIS system. The release of geospatial data in the form of Web Map Service (WMS) and Web Features Service (WFS) is managed by Geoserver, which provides data on territorial boundaries and georeferenced point data. StatVIEW may be seen as a JSON-stat hub that can be used as a web service to further disseminate data in a machine-readable format.

1.3.2. StatVIEW architecture

StatVIEW, based on a microservices architecture pattern, is composed of a map server and a geospatial database (Geoserver, Postgres, PostGIS), different data sources (data warehouse, web services, database, etc.), interrogation engine and user interface (Figure 4). By using a map server you can import shapefiles in geospatial database containing geographic data and disseminate them through the HTTP protocol, in compliance with the standards defined by the OGC Consortium (WMS, WFS, WCS), in a format that suits the web – GeoJSON, TopoJSON, etc. – and existing frameworks (OpenLayers, Leaflet, D3.js).

Though StatVIEW is a highly scalable system that can support different data sources (databases, files, web services), different models (SDMX, DDI, JSON-stat) and different formats (XML, JSON), it was chosen to standardize the source output in the JSON-stat format. This choice, in addition to standardizing data from different sources, permits to integrate the frameworks to work with (JSON-stat framework toolkit, OpenLayers, Leaflet and Raphael JS) and permits to unify the programming language (Javascript). Unlike many tools that enable users only density maps visualization, StatVIEW allows them to perform operations normally included in GIS Desktop and WebGIS systems (Topology Overlay; Network analysis; Buffering).

Figure 4: StatVIEW architecture



When multidimensional cubes are updated in the Istat data warehouse, updated data are automatically visualized in the user interface. Similarly to what happens in the case of Istat data warehouse, it is possible to visualize or monitor data from NSOs, research institutes and government agencies.

1.3.3. Graphical representations and geospatial analysis through StatVIEW

1.3.3.1. DENSITY MAPS, GRAPHS AND TABLES

Among the different possible types of graphical representations, StatVIEW permits to visualize statistical data in the form of density map, graph or table (table visualization is achieved through JSON-stat Table Browser). The box below shows the case of Istat data related to residence permits of non-EU citizens. By clicking on the menu available on StatVIEW, it is possible to:

- select the dimensions associated to the dataset;
- explore the entire content of the dataset;
- choose the type of data representation: cartographic (Figure 5); graphic (Figure 6); tabular (Figure 7);
- customize the research (for instance, by choosing dimension, territorial level, type of visualization, or by splitting density areas; several other possibilities are also available). Gradients and subdivisions can be easily customized, adding extra layers in the database management. The customization depends on the layers available in the chosen map. In the case of residence permits of non-EU citizens, the available layers are: the layer containing the georeferenced organizations providing assistance to migrants, the layer that, via geospatial query, creates a bubble chart representing the number of the georeferenced organizations active in the territory (Figure 8). By selecting the layers and, for example, different base map or colour shade or partition (quartile, quintile), it is possible to get a different visualization, conducing to a different interpretation of the phenomenon;
- visualize data according to different geographical units.

A legenda allows users to know the threshold values associated with colour shades.

Residence permits data visualization

Figure 5: Density map

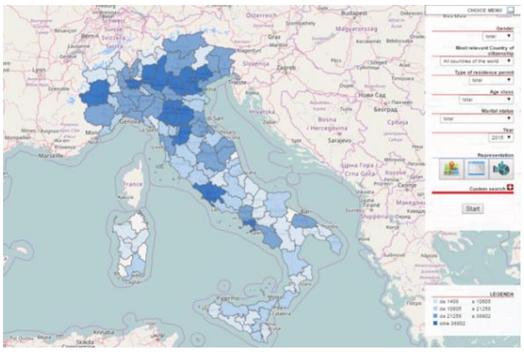


Figure 6: Graph

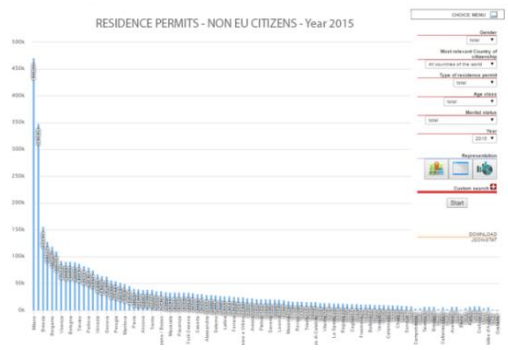


Figure 7: Table

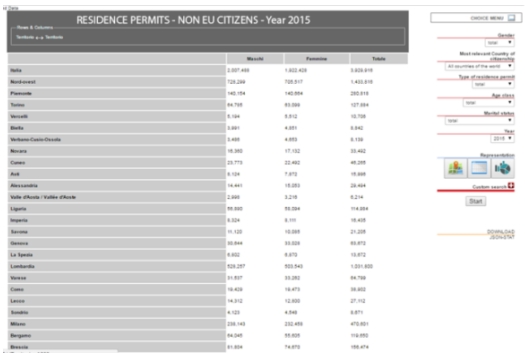
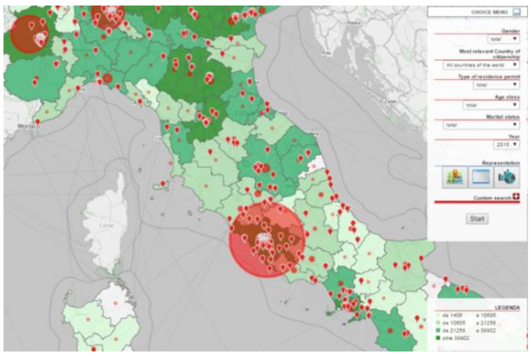
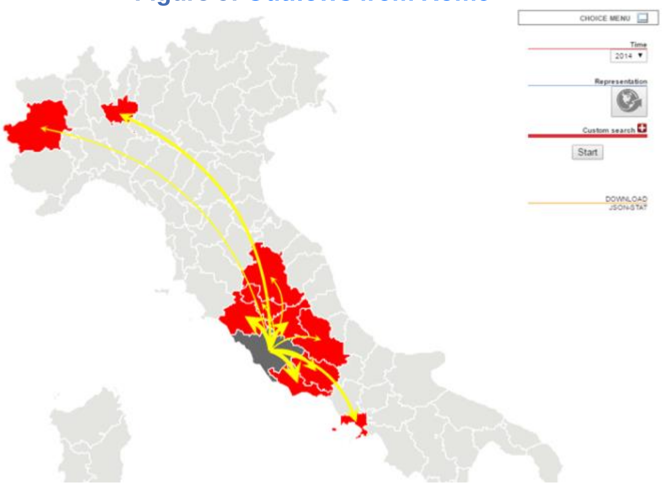


Figure 8: Topology overlay



1.3.3.2. INFLOWS AND OUTFLOWS

Figure 9: **Outflows from Rome**



Transfers of residence

Through StatVIEW it is possible to represent inflows and outflows to and from a given territorial border.

Figure 9 shows outflows of foreign citizens cancelled from the population register of Rome and registered in the population register of another Italian municipality.

It is possible to set the maximum distance or the distance range for the desired flows.

For example, you may be interested in visualizing the flows within 100 km from a given province or within a distance ranging 100-300 km. The dimension of the arrows gives an idea of volumes: the thicker the arrow, the greater the flow.

1.3.3.3. ROUTING

Figure 10: **Routing**

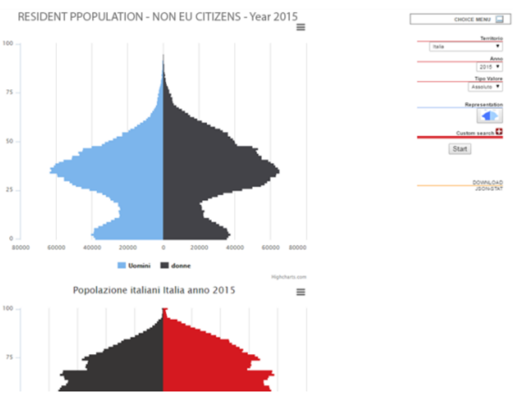
OSRM Routing machine

StatVIEW uses the OSRM (Open Source Routing Machine) routing algorithm.

By using this algorithm, it is possible to obtain the shortest route to a geo-localised point in the map server used by StatVIEW (e.g., hospital, police station, etc.) from any point in the map, as well as the respective travel time (Figure 10).

1.3.3.4. CUSTOM GRAPHS

Figure 11: **Pyramid graph**



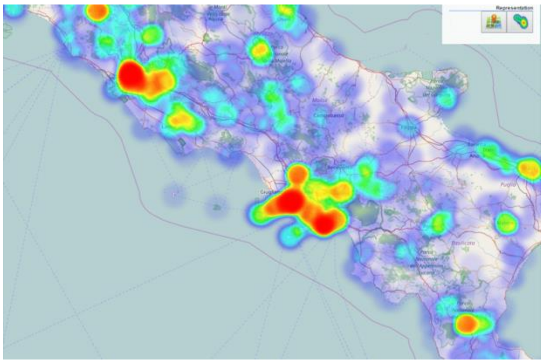
Pyramid graph

The standardization of sources' output allows users the easily creation of *ad hoc* plugins for the generalized visualization of multi-source graphs or dashboards.

Figure 11 presents a pyramid graph showing the age of foreign citizens resident in Italy, compared to that of resident Italian citizens.

1.3.3.5. HEATMAPS AND CLUSTERS

Figure 12: **Heatmap of private museum**



Private museums

StatVIEW permits to create heatmaps and clusters of geolocated points to represent their density and distribution on the territory.

Figure 12 represents the heatmap of private museums resulting from the census survey carried out annually by Istat, the Ministry for Cultural Heritage and Activities and Tourism, the Regions and Autonomous Provinces.

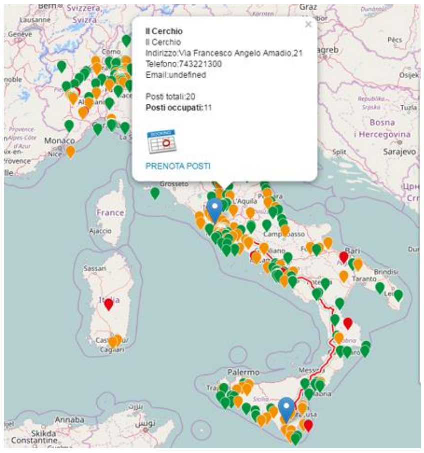
1.3.3.6. AD HOC APPLICATIONS

Management of immigration centres

Thanks to the support of geospatial database, StatVIEW permits the development of *ad hoc* applications.

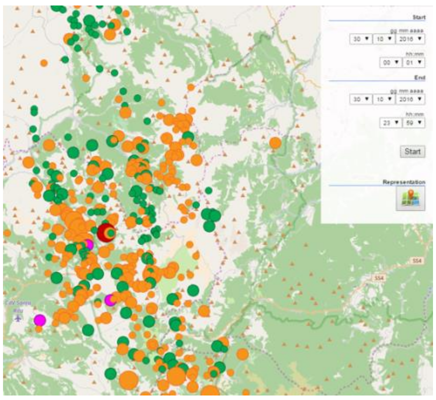
The map in Figure 13 indicates the position of immigration centres. The colour indicates the number of available sleeping accommodations. By clicking on the marker of an immigration centre, an interested user can book a certain number of accommodations, in order to equally distribute immigrants. In a reserved area, it is possible to manage the immigration centre and the booking.

Figure 13: **Ad hoc application**



1.3.3.7. MONITORING

Figure 14: **Earthquakes real-time monitoring**

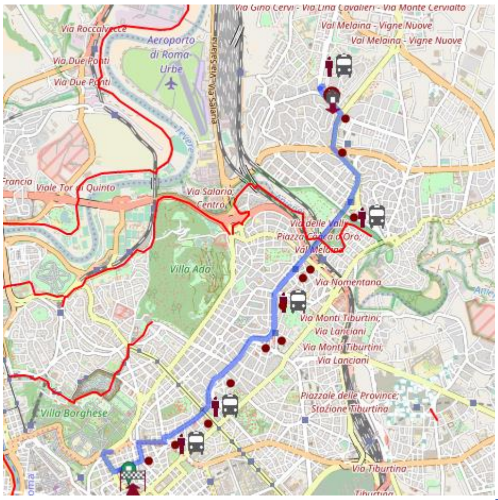


Real-time monitoring

Micro services architecture at the basis of StatVIEW permits the creation of different monitoring system (real-time and time-series) to study and control several phenomena.

Figure 14 is an example of how StatVIEW enable users to real-time visualize and monitor earthquakes and their main features (location – with the precise indication of geographic coordinates –, magnitude, depth, day and time of occurrence). In the case of earthquakes, the data source is INGV, the Italian National Institute of Geophysics and Volcanology, devoted to 24-hour countrywide seismic surveillance, real-time volcanic monitoring, early warning and forecast activities.

Figure 15: Urban mobility real-time monitoring



Urban mobility

StatVIEW permits the exploitation of urban mobility data to real-time control and monitor road traffic, timetables and public transports routes.

In Figure 15 the case of collective mobility in the Metropolitan Area of Rome is represented. The website of the Municipality of Rome provides a set of APIs through which it is possible to monitor the positions of public transports (buses, trolleybuses, trams, metro lines, regional railways, etc.), know routes and route planners to follow to get from one point to another of the city using public transport.

The overlay, by combining for example routes and bus positions to the tracks of the bike paths and car sharing areas, facilitates urban mobility.

1.4. Concluding remarks

In the Internet Era, the ever increasing availability of data, essential in supporting research and decision making, necessarily involves the use of proper tools to best analyze, assess and interpret their meaning in an interactive fashion. Solicited by ESS, NSOs, playing an important role in developing, producing and disseminating statistics that respect standardized quality criteria, have started developing suitable tools for the integration of different types of data and its dissemination, i.e. platforms for data storage, analysis and visualization. A particularly interesting way to integrate different sets of data is to bring in a spatial dimension; in this way, the complexity of phenomena under investigation is still there but their analysis can be facilitated and early warning is made more at hand.

StatVIEW, the proposed webGIS system, is an innovative tool providing a wide range of solutions to easily analyze the many phenomena for which statistical data exist as well as to quickly monitor the evolution of different phenomena at the same time. It is currently in an experimental phase, consequently a web analytic tool to keep track of visitors, although already in place in a basic form, does not permit at the moment a deeper analysis on traffic and insight on visitors' behavior in order to create strategies according to it. What is certain is that the platform, already presented in numerous national and international conferences, is receiving tangible and more than positive feedback from academics, researchers and statistics stakeholders, mostly interested in the use and replicability of it.

One of its significant features is represented by the opportunity of having constantly updated statistical information, given the fact that the webGIS system does not imply the loading and transferring of data, while, instead, it provides a dynamic link.

Overall, this open source tool can be conveniently shared among NSOs and extended to any institution that performs data dissemination through API REST.

Furthermore, a webGIS system such as StatVIEW is along the lines of ESS Vision 2020, in the sense that it helps official statistics to engage users proactively and to meet their demands in a cost-effective and responsive manner.

References

De Smith M.J., Goodchild M.F., Longley P.A. (2007), *Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools,* Troubador Publishing Ltd.

Henderson C. (2014), *Mastering Geoserver*, Packt Publishing Ltd.

Kalkhan M.A. (2011), *Spatial Statistics: GeoSpatial Information Modeling and Thematic Mapping*, CRC Press.

Obe R.O., Hsu L.S. (2015), *PostGIS in Action, 2nd Edition,* Manning Publications

Richardson L., Ruby S. (2007), *RESTful Web Services. Web services for the real world*, O'Reilly Media.

Schabenberger O., Gotway C.A. (2005), *Statistical Methods for Spatial Data Analysis*, Chapman & Hall/CRC Press.

Webber J., Parastatidis S., Robinson I. (2010), *REST in Practice: Hypermedia and Systems Architecture*, O'Reilly Media