

The Open Data GeoPortal of the Lamma Consortium

Manuela Corongiu¹, Riccardo Mari^{1,2}, Raffaella Ferrari¹, Lorenzo Bottai¹, Valentina Grasso^{1,2}, Federica Zabini^{1,2}, Luca Fibbi^{1,2}, Daniele Grifoni^{1,2}, Claudio Tei^{1,2}, Francesco Pasi^{1,2}, Bernardo Gozzini^{1,2}, Simone Giannecchini⁽³⁾

¹ Consorzio LaMMA, via Madonna del Piano 10, 50019 Sesto Fiorentino (FI) Italy

² CNR Ibimet, via Caproni 8, 50145 Firenze, Italy

³ Geosolutions s.a.s. Via di Montramito 3/A, 55054 Massarosa (Lucca) – Italy

Corresponding author:

Manuela Corongiu¹

Email address: corongiu@lamma.rete.toscana.it

ABSTRACT

The new LaMMA Open Data platform (<http://dati.lamma.toscana.it>) allows data download related to information delivered / managed by the Consortium, encouraging the reuse both at technical and legal level. At the moment the Database is a main collection of more than 200 weather forecast layers, together with other basic geospatial layers (NDVI, LandCoverChangeFlows, Landslides, Natural Risk Zones), that are in continuous updating. LaMMA open data platform integrates in a harmonised interface, most of the spatial dataset already available through the LaMMA geoportal (<http://geoportale.lamma.rete.toscana.it/MapStore/public/>), making them available for a direct download with a specific licence of use.

Keywords: Geoportal, Open Data, Open Source, OGC, CKAN, Inspire, Spatial Data Infrastructure, Weather, GIS, Web Services

INTRODUCTION

The Public Authorities and the Research Bodies, among their several roles, have to deliver, manage and store data as a result of their institutional tasks. Government data is being put online to increase accountability, contribute valuable information about the world, and to enable government, the country, and the world to function more efficiently (Berners-Lee, 2009). Some of these data are constrained by the privacy protection or by intellectual property, while others may be freely disseminated. Open data means contributing to potential development of innovative services, where applications must better organize their information. For that reason, open data can have an important economic impact on society. For instance, an infinite number of possible reuses can be originated by weather data, such as operational systems that assess environmental impacts (e.g. Fire as well as industrial risks), or ad-hoc applications for territorial planning as well as for citizen leisure activities.

THE SPATIAL DATA INFRASTRUCTURE

The LaMMA SDI (Spatial Data Infrastructure) has been designed according to the classical multi-tier (layer) system with a communication paradigm based on open source service oriented architecture, where each component (service), interacts with the others through a set of messages written in a standard format. With the term service we do not only include each one of the three layers but also all the components inside them, allowing the integration of several multifaceted computational units inside a unique system. The SDI is much more than data and goes far beyond surveying and mapping, it provides an environment within which organisations and/or nations interact with technologies to foster activities for using, managing and producing geographic data (Rajabifard et al., 2001). More specifically, the LaMMA SDI comprehends the OpenData Portal, the Geoportal platforms, the source data and a middleware level that allows the interchange among them, integrating them in a harmonised interface both spatial and no spatial datasets. Geoportal make accessible LaMMA datasets, together with other basic geographical information, recalling them thanks to OWS (Opengeospatial Web Services) standard services, available in other geoportal such as the Tuscany administration geoportal named Geoscopio (<http://www502.regione.toscana.it/geoscopio/cartoteca.html>). The Open Data platform (<http://dati.lamma.toscana.it>) allows data download related to information delivered / managed by the Consortium, encouraging the reuse both at technical and legal level. The datasets, over 220, mostly related to weather forecast and geographical (NDVI, LandCoverChangeFlows, Landslides, Natural Risk Zones) topics, but also no spatial data (such as administrative documentation), integrating in a harmonised interface, most off the spatial dataset already available through the Lamma geoportal (<http://geoportale.lamma.rete.toscana.it/MapStore/public/>).

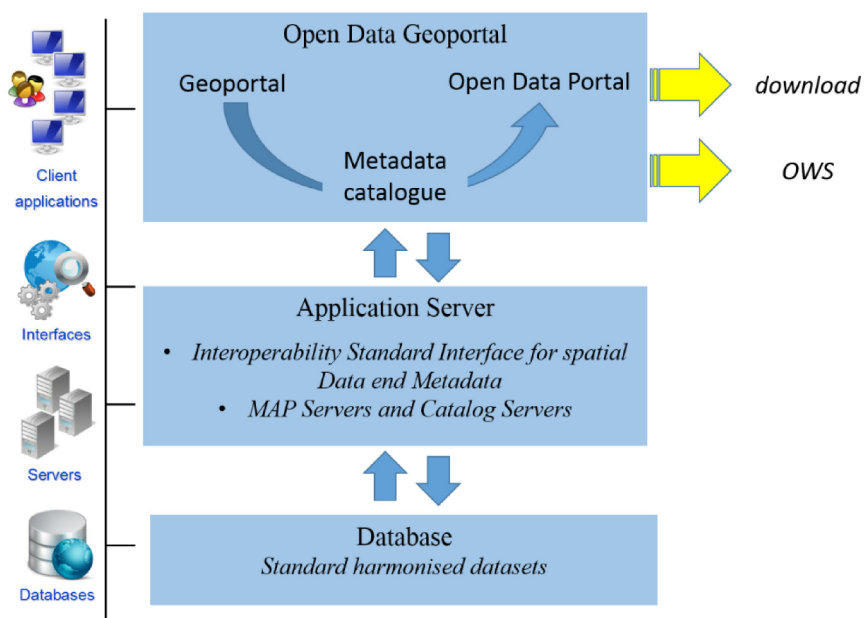


Figure 1: The Lamma SDI (Spatial Data Infrastructure)

The layers published on the Geoportal and made available on open data platform are a selection of most important variables in the meteorological models. They are accessible to the public and can be viewed in the Viewer integrated with the catalogue or downloaded free of charge as georeferenced images (GeoTIFF).

DATA AVAILABILITY: DYNAMICITY AND UPDATE

The particularity of meteorological information is their organization in models, archives and formats according to the type of information, source of acquisition and level of elaboration. These formats are not all functional or directly manageable in their entirety, as data to be made available and immediately accessible. The datasets therefore require a preliminary phase of evaluation and analysis of the contents to identify the most appropriate elements for publication via filters and elaborations that maintain the significance of the variables to be highlighted. Indeed, as many people are aware, weather data are made available on geographical charts only after elaborations. Related information, sometimes complex, run in meteorological models from raw data. The definition of the algorithms and variables in play constitute the core of the contents, as these are otherwise not directly observable by the main users of weather data, even if specialized, in the form of environmental and spatialized data.

The key point of the Geoportal is the possibility of coherently overlaying forecasts for geophysical parameters coming from the meteorological models elaborated internally together with additional information created and managed by the LaMMA Consortium, like in-site observations about weather collected in near real-time from the Italian and international observation networks. This information, although having a spatial component, had neither, up to now, been exploited in a geospatial context nor visualized in a GIS environment, but it was rather distributed to the end users in text form, having in mind specific elaborations or simply used for the production of charts.

Meantime the main issue to solve has been to update data according to frequency of the meteorological model runs. A time window of 3 days is currently maintained for the meteorological models, i.e. all the data and related metadata are available for the 3 days prior to the date of access to the Geoportal/Open Data platform. Each dataset published refers to the source model frequency.

Datasets coming from meteorological models are:

- GFS (Global Forecast System) global model, with spatial resolution 50 km, 180 hour weather forecasts, updating frequency 4 times per day.
- WRF (Weather Research and Forecasting model) limited area model, with spatial resolution 12 km, GFS formatted data, with domain extended to the entire Mediterranean and 120 hour weather forecasts (med_gfs_12km*), updating frequency twice daily
- WRF limited area model, with spatial resolution 12 km and ECMWF (European Centre for Medium-Range Weather Forecasts) formatted data, with domain extended to the central-western Mediterranean, with 120 hour weather forecasts (arw_ecm_12km*), updating frequency twice daily
- WRF meteorological model, inserted in the 12 km model on ECMWF (arw_ecm_12km),

with spatial resolution 3 km, domain extended to Italy, with 48 hour weather forecasts (arw_ecm_3km*), updating frequency twice daily

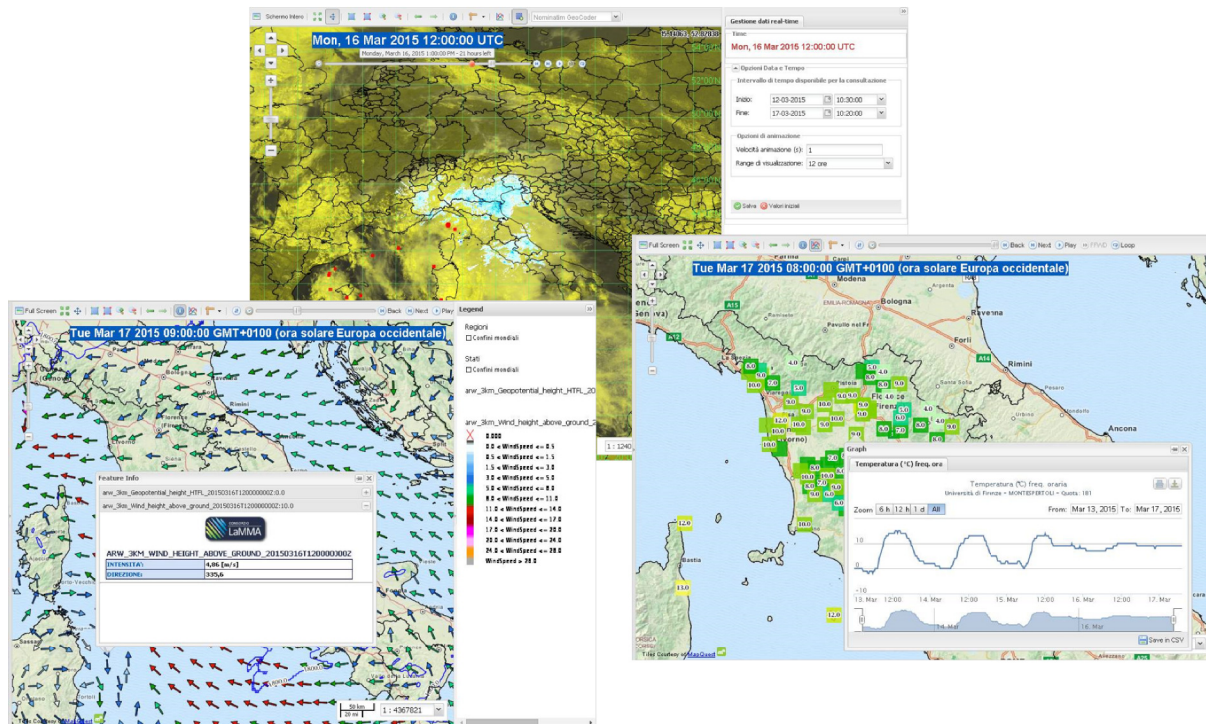


Figure 2: meteorological datasets available on LaMMA Geoportal

In addition to meteorological models, raster layers are also produced in near real-time exploiting raw data from the Meteosat MSG2 (Meteosat Second Generation) and MSG3 (Meteosat Third Generation) geostationary meteorological satellites managed by EUMETSAT (European Organisation for the Exploitation of Meteorological Satellites) and the RADAR images coming from the Italian Civil Protection. Finally, some geographic datasets, harmonised following the related schemas of the Inspire data specifications are made available as examples of the transformation service for a Spatial infrastructure. That datasets refers to landslides and land cover themes derived from regional archives. These islands of data of different standards and quality (Smits, 2003) has been organised for creating an SDI to provide potential users access to (spatial) data, sharing of data, save resources, time and effort, by avoiding the duplication of effort required to acquire and maintain the data (Rajabifard et al., 2001).

In general, because of the dynamicity of meteorological datasets, the focal point of all the work has been to set up a pre-processing and publishing infrastructure. SDI would have been able to automatically process, catalogue and publish in near real-time the huge volume of data acquired by LaMMA, in order to create layers and mash-ups with highly valuable information content and always up-to-date. Moreover, in order to reduce the hardware and software resources necessary to run the infrastructure, it was decided to limit the temporal window of the data available online, by relying on automatic procedures that would run at night, i.e. when accesses are scarce, to remove the obsolete data (e.g. weather models outputs older than 3 days).

METHODS AND MATERIALS

A synergic and integrated infrastructure for spatial data has been carried out through open source software. The open-source software provides great potential to make available components for SDI implementations that are affordable by resources poor organisations (Reid at al., 2001). In fact, the LaMMA Geoportale integrates, in a single simple but powerful interface, the functionalities of research, display and download of the available data. This objective is to provide a ready-to-use tool for all users who do not intend to connect directly to the services offered or to download (and therefore reuse) the data: in this case, we relied on the software Open Source MapStore.

The open data platform is directly connected to the Geonetwork metadata catalogue that in turn automatically provide a real-time ingestion of datasets in geoportal. For that, each metadata must include resources for download when already available on geoportal as well as open data platform, such as WMS and WMTS for time and elevation weather parameters. The Lamma open data infrastructure has been implemented by the use of CKAN software, which is the world's leading platform for portals of open-source data, developed by the Open Knowledge Foundation, a no profit organization that promotes free knowledge. All the datasets have been made available according to the CC-BY license - Attribution Creative Commons.



Figure 3: the interconnection between geoportal and open data platform through the metadata catalogue

That choice will allow an easier federation with Open Tuscany (<http://dati.toscana.it/>), the open data portal of Tuscany Regional Government that until now has hosted, as supplementary task, some Lamma Consortium datasets.

The open data infrastructure has been implemented thanks to the Life+IMAGINE European contribution and with the support of the Geosolutions company.

CONCLUSIONS AND PERSPECTIVES

A first implementation of the linked infrastructure between geoportal and open data framework has been set-up by the use of exclusively open source software. The amount of datasets will be in continuous growth and update, but the issues about this dynamicity have already been taken into account during the designing phase. For the future, some critical aspects have to be analysed in depth, related to the continuous updating of data and metadata particularly taking care on:

- How to align the LaMMA metadata catalogue to the Italian RNDT (National Territorial Data Inventory) metadata catalogue, through the metadata harvesting process;
- How to implement the federal system with the Tuscany Regional Platform named OPEN TOSCANA (<http://dati.toscana.it/>).

That approach make available data and both for expert and non-expert users. The first one can download directly their datasets in specific software to reuse data and metadata to do more and different derived analysis; the second can view and access to information by the use of the web client application, avoiding to elaborate other kind of information by their own desktop client. Moreover, these approaches have been addressed to link and make interoperable both data and metadata on the web: toward the RNDT national metadata catalogue as well as toward the Tuscany Regional federal opendata framework.

REFERENCES

- Berners-Lee T., Putting Government Data Online, <http://www.w3.org/DesignIssues/GovData.html>, 2009.
- Rajabifard A, Williamson IP (2001): Spatial data infrastructures: concepts, SDI hierarchy and future directions. Proceedings Geomatics 2001, Australia.
- Reid J, Martin F (2001): The open source movement and its potential in implementing Spatial Data Infrastructures. Proceedings International Symposium on Spatial Data Infrastructure, Melbourne, Australia. Available http://www.geo.unizh.ch/publications/sstein/pubothers/reid_martin_foss_for_sdi_2001.pdf.
- MSmits P (2003): INSPIRE: Infrastructure for Spatial Information in Europe. Presentation held on Sept. 17 2003, Open European Symposium on GI Interoperability, Segovia, Spain.