The Open Data GeoPortal of the Lamma Consortium

Manuela Corongiu[(1)], Riccardo Mari[(1,2)], Raffaella Ferrari[(1)], Lorenzo Bottai[(1)], 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)] and 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*

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. 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 (eg. Fire as well as industrial risks), or ad-hoc applications for territorial planning as well as for citizen leisure activities. 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.

The datasets, over 220, belong to the weather forecast and geospatial topics above all, but they are in continuous updating, both spatial and no spatial (such as administrative documentation). Lamma open data platform integrates in a harmonised interface, most off the spatial dataset already available through the Lamma geoportal (http://geoportale.lamma.rete.toscana.it/MapStore/public/), now available for download as open data. 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, sometimes complex, of the raw data in order to run meteorological models in which 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 this tool 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.

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

The layers published on the Geoportal and made available on open data platform are a selection of most important variables in the meteorological models mentioned above. 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). 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.

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. But, in general, because of the dynamicity of meteorological datasets, the focal point of all the work was to set up a pre-processing and publishing infrastructure that 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. It is also important to note that, 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).

So, a synergic and integrated infrastructure for spatial data has been carried out through open source softwares. The LaMMA Geoportal 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 reutilize) 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 are made available according to the CC-BY license - Attribution Creative Commons.

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.