

MAPPING COVID-19 EPIDEMIC DATA USING FOSS

P. Zatelli^{1*}, M. Ciolli¹, A. Besana², N. Gabellieri²

¹ Department of Civil, Environmental and Mechanical Engineering, University of Trento, 38123 Trento, Italy -
(paolo.zatelli, marco.ciolli)@unitn.it

² Department of Humanities, University of Trento, 38122 Trento, Italy -
(angelo.besana, nicola.gabellieri)@unitn.it

KEY WORDS: COVID-19, health geography, FOSS GIS, WebGIS, Database.

ABSTRACT:

The recognition of spatial and temporal patterns in the distribution of a pandemic plays a pivotal role in guiding policy approaches to its management, containment and elimination. For this purpose, a database has been built for the COVID-19 pandemic in the Trentino Province, in the eastern Italian alps, near the border between Italy and Austria. The database management system and the WebGIS mapping these data is based on Free and Open Source Software. The Data Base Management System (DBMS) runs on MySQL, available under the GNU General Public License, storing and processing geographic data. A custom procedure has been created to update the dataset, with the capability to import data from suitably formatted spreadsheets by an authenticated administrator. To ensure flexibility and responsiveness on desktop and mobile devices, the WebGIS has been created with a client-side approach, using the Leaflet and Bootstrap JavaScript language libraries, available with Open Source Licenses. These libraries, with additional custom scripts, create the user interface and render geographic data into maps. The exchange of data between the DBMS server and the client is performed using geojson tables. To protect the privacy of the patients, WebGIS users cannot access the source data even though maps and graphs can be downloaded as pictures. Geo-statistical analysis aimed at the detection of spatial and temporal patterns is underway.

1. INTRODUCTION

An important factor in determining policy approaches to pandemic management, containment, and eradication is the detection of spatial and temporal patterns in the disease's distribution. Four steps are required to provide information about spatial and temporal patterns of a phenomenon: the collection of data, their organization and management (Hu et al., 2020), their representation as tables, charts and maps, and finally their analysis with geo-statistical tools (Trias-Llimós et al., 2020) (Pranzo et al., 2023). Spatial and temporal patterns are extremely useful in many research fields, for example they can be used to determine anthropic interactions with fauna to reduce Human Wildlife Conflicts (Corradini et al., 2021).

The collection of pandemic data poses a challenge: on the one hand, the highest possible spatial and temporal resolutions are required to make the detection of patterns more effective (Carballada and Balsa-Barreiro, 2021), allowing the application of containment tools as local as possible, on the other hand it presents major privacy problems, pushing towards data aggregation. For these reasons public COVID-19 datasets and maps are usually available at low spatial and temporal resolutions (Franch-Pardo et al., 2020), because averaging over time and space automatically provides a layer of anonymization by data aggregation (Patel, 2020).

WebGISs are an effective way to make the spatial components of the information about the pandemic available to non specialists (Lipeng et al., 2020) (Mooney and Juhász, 2020).

The Italian Civil Protection Department, part of the Italian Presidency of the Council of Ministers, has published a "Coronavirus dashboards" page (Coronavirus dashboards - Italian Civil Protection Department, 2023), where interactive dashboards display the overall number of reported cases, people who are

currently positive, people who have been cured, and people who have passed away (Figure 1). It is possible to explore COVID-19 data for a date in the past from January the 1st, 2020. It is also possible to download the data provided by the Ministry of Health in open format, with CC-BY license and metadata. While the temporal resolution is high because daily data are available, the spatial resolution is very low, since the information is provided at the regional level: this makes the dataset unsuitable for the detection of spatial patterns and for the analysis of potential correlations with other variables.

A similar dashboard has been created by Trentino Digitale SpA (Trentino coronavirus dashboard, 2023), in-house organ of the local government, for desktop and mobile clients. While the spatial resolution is at the municipality level with daily temporal resolution, only three variables (total cases, current cases and increment) are provided and it is possible to map only the current cases value and only for the current date. Furthermore, the system has ceased to be updated on April the 29th, 2020. For a more general overview of geodashboards dedicated to the dissemination of COVID-19 in Italy, see (Gerbino, 2020).

In this research, a database has been built for the COVID-19 pandemic in the Trentino region, in the eastern Italian alps, near the border between Italy and Austria. The database contains COVID-19 information at the municipality level, with weekly cases. A WebGIS is available to explore the dataset using maps and graphs.

2. MATERIALS AND METHODS

The database development and the geo-portal for data exploration is the result of two projects: GEO-MARCA, funded by the Ministry of University and Research through the FISR-COVID19 Italian national fund, and GEO-SMART, funded

* Corresponding author

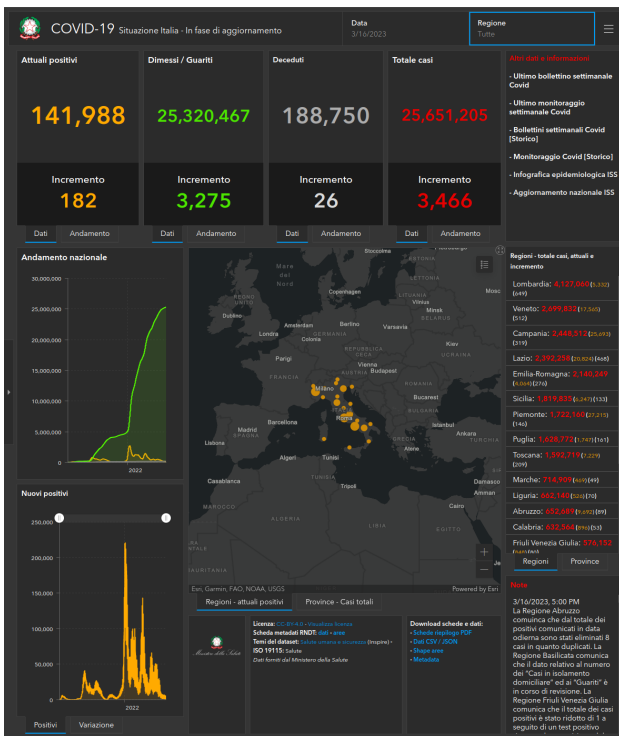


Figure 1. Coronavirus dashboard for desktop clients from the Italian Civil Protection Department.

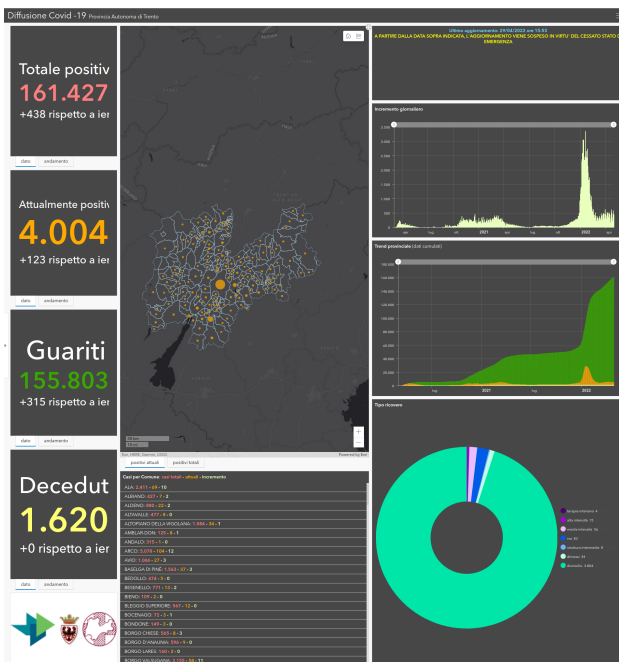


Figure 2. Coronavirus dashboard for desktop clients from Trentino Digitale SpA.

through the 2020 COVID-19 Internal Call of the University of Trento. The main aims of the projects are the availability of a spatial representation of the epidemic in Trentino for the general public and the analysis, statistical modeling and cross-referencing with social and environmental data of the COVID-19 information, for the benefit of scholars and public decision makers.

2.1 Materials

The Province of Trento, with a population of about 542,000 inhabitants, represents the primary corridor for transporting people and products between Italy, Austria and Germany. The area has also an intense tourist development, in particular for winter sports, with the presence of ski slopes, ski lifts and hotels. These two features have played an important role in the diffusion of COVID-19 in the region because the movement of people, both through the main communication routes and the movement of tourists in the lateral valleys, has been the main driver in the virus spread. Therefore, the availability of a reliable database collecting COVID-19 cases is fundamental to map the pandemic evolution (Mollalo et al., 2020).

At the same time, the status of autonomous region of the Provincia Autonoma di Trento allows greater discretion in the organization of health data, their scientific use and their dissemination. In this context the local government and the University of Trento, in particular its the Geo-cartographic Center (GeCo), have signed an Agreement for sharing COVID-19 data and their analysis (Gabellieri et al., 2021). The resulting dataset collects the official number of the infected, infected in healthcare residences (RSA, Residenze sanitarie assistenziali), clinically recovered, deceased people, and their age group. The dataset contains daily data at the municipal level, starting from the beginning of the COVID-19 epidemic in March 2020 until the whole 2022. Data anonymization has been carried out by aggregating data on a weekly basis and by hiding data with small numbers, with the threshold set to 5 cases. The sole use of official data created by public agencies tasked with managing public health, specifically the local Health Authority (Agenzia Provinciale per i Servizi Sanitari, APSS), ensures the validity of their production process and strict observance of patient data confidentiality.

Two maps have been used for the map geometry and for the background: the municipalities boundaries have been provided by the Provincia Autonoma di Trento under the CC0 1.0 Universal Public Domain Dedication license, while the OpenStreetMap (OSM) maps used for the background are available with the Open Database License.

2.2 Methods

All the data collection, management, analysis and visualization have been carried out using Free and Open Source Software (FOSS). FOSS for dealing with geographic information has proven useful for processing spatial data in all areas, including education (Ciolli et al., 2017) (Quinn, 2021), environmental suitability for fauna (Tattoni et al., 2019) and tourism (Cannata et al., 2022), image analysis (Gobbi et al., 2019) (Zatelli et al., 2022), and evaluation of landscape metrics (Zatelli et al., 2019). In particular, it has been demonstrated that FOSS DBMS can adapt to a variety of types of data and database structures, while easily integrated with a WebGIS (Simeoni et al., 2014) for data visualization.

The system has three main components: the Data Base Management System (DBMS), the web server and the client. To keep the system as simple as possible for speed and future maintenance and update, the server side only provides the data to create maps and graphs to the clients as geojson tables. The web interface, graphs and maps are all created on the fly on the client side (Figure 3).

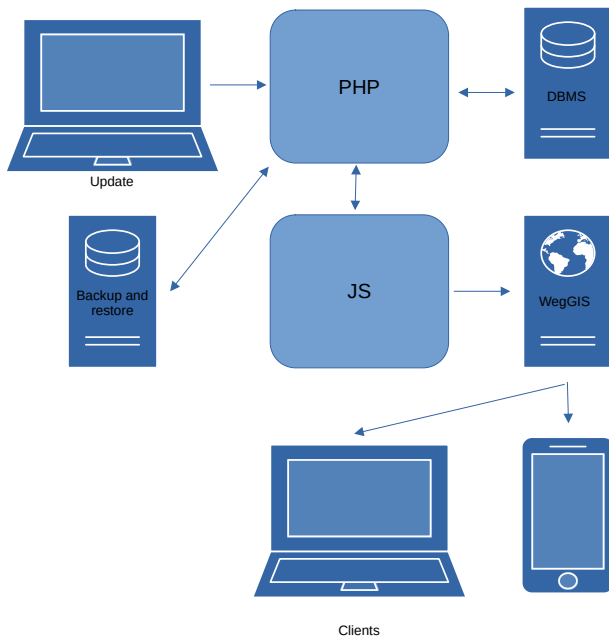


Figure 3. System architecture.

3. DATABASE

The back end of the system runs a Data Base Management System (DBMS), which organizes the data, including the spatial components, and a web server, which provides access to the users.

The DBMS runs on MySQL 8.0.32 (MySQL DBMS, 2023), a relational database management system (RDBMS) available as Free Software under the GNU General Public License. MySQL provides the capability of storing and processing geographic data, following the OpenGIS data model.

The database contains 9 tables. The table structure is the same for all the tables: one record for each of the 166 municipalities of the Provincia Autonoma di Trento and one field for each week, starting from Monday the 4th, February 2020. The tables are: weekly contagions, weekly deceased, weekly recovered, weekly contagions in healthcare residences, weekly active cases, weekly cumulated cases, weekly recovered cumulated, weekly cumulated deaths and weekly active cases in healthcare residences.

The link between tables is given by the use of an unique code for each municipality, used as key field in all the tables. This code, assigned to each municipality by the Italian National Institute of Statistics (Italian: Istituto nazionale di statistica, ISTAT), is a 6-digit code: the first 3 digits indicate the province, the last three digits contain a progressive number for the municipalities. Since all the municipalities in the database belongs to the province fo Trento, the first 3 digits are the same (022) for all the records.

The rows contain the municipality name, its ISTAT code and a value for each week. Weeks, and the corresponding columns, are identified by the year and the week number.

The geometry is stored in a set of tables: a geometry columns table provides an index of the two available geometry tables, which contain the polygons representing the municipalities and their centroids coordinates, and the table containing the datum definition. Polygons and centroids coordinates are stored in the Well-Known Binary (WKB) format, while the datum definition is stored in the Well-Known Text representation (WKT).

The system uses the WGS84 latitude/longitude (unprojected) datum, with EPSG code 4326.

Two additional tables are used to set the scale factor of the graphical representation using centroids, one for the absolute values and the other for the percentages. A further service table contains the list of the login parameters (email address and hashed password) for the authenticated users, which are allowed to perform maintenance operations.

A custom procedure has been created in PHP to update the dataset, with the capability to import data from suitably formatted spreadsheets. A roll back option is provided in case of failure of the import procedure. Data base management and update functionalities are available only to authenticated WebGIS administrators and accessible through a dedicated web page (Figure 4).

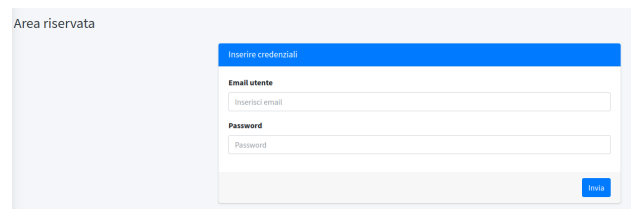


Figure 4. Login prompt for privileged users.

4. WEBGIS

The main goals in the design and development of the WebGIS have been the ease of use and clarity of data presentation, both on large screens for desktop and laptop clients and on mobile devices. This approach maximizes the user performance while exploring the data, by splitting the processing tasks and load between server and clients.

The client side uses the Open Source Leaflet (Agafonkin, 2023) JavaScript libraries, version 1.8.0, available under a BSD 2-Clause License, with custom scripts, for map rendering. The Bootstrap library (Bootstrap · The most popular HTML, CSS, and JS library in the world, 2023) version 4.4.1 is used for the frontend page structuring. This approach ensures flexibility and responsiveness on desktop and mobile devices. Finally, the web pages are served using the Apache web server (Apache, HTTP server project, 2023), version 2.4.41.

The exchange of the data between the server and the client is performed using geojson tables, created on the fly according to the user's requests. In a similar way, the data temporal variation graph is created by the js library, which automatically reads the date selected by the user, extracts the relevant data from the database and creates the graph. As long as they fit within the database structure, the system automatically uses all the accessible data. To protect the privacy of the patients, WebGIS users cannot access the source data, even though maps and graphs can be downloaded as pictures.

Cartographic data include background maps from the OpenStreetMap (OSM) project and a map of the municipalities boundaries for the Province of Trento, which serves as a spatial basis for the dataset.

After an initial page describing the service, the interface allows the exploration of the data in the 9 tables described in Section 3, therefore creating a map and a graph for weekly contagions, weekly deceased, weekly recovered, weekly contagions in healthcare residences, weekly active cases, weekly cumulated cases, weekly recovered cumulated, weekly cumulated deaths and weekly active cases in healthcare residences (Figures 5 -

10).

Absolute values are represented with circles of varying sizes locate in each municipalities centroid (Figure 5). A range of colors associated to the administrative polygon vector layer is used to represents values given as ratio between affected people and total population (Figure 6).

It is possible to choose a week using a slider in the interface: for maps and graphs representing current cases (weekly contagions, weekly deceased, weekly recovered, weekly contagions in healthcare residences, weekly active cases, and weekly active cases in healthcare residences) data for the chosen week is used, while for cumulative maps and graphs (weekly cumulated cases, weekly recovered cumulated and weekly cumulated deaths) the sums of cases up to the chosen week (included) are used.

For absolute values, the representation using circles allows the interaction with the timeline to observe the occurrence of localized clusters and the permanence of enduring hotspots.

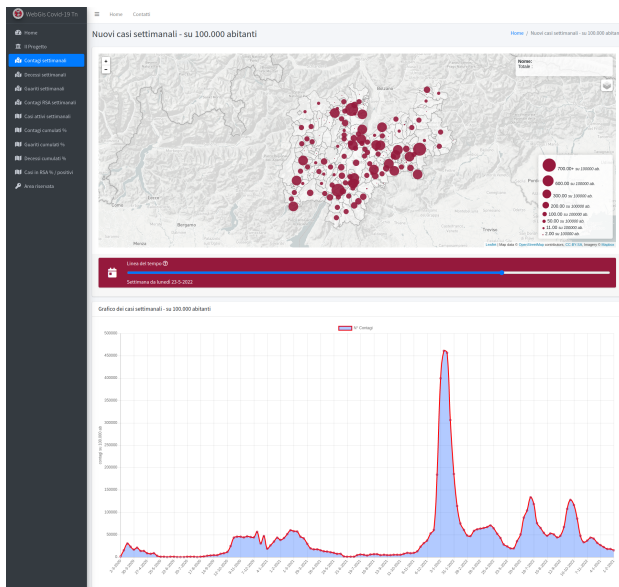


Figure 5. Weekly cases (week from May the 23rd, 2022).

When the user selects a municipality the system displays its value for the chosen table with a popup window (Figure 7).

The server side approach has the advantage of automatically adapting to the type of client. Figures 5 - 10 are from a laptop computer, Figure 11 has been taken from a smartphone.

A virtual machine that houses both software and data powers the system on the server side.

5. CONCLUSIONS

The WebGIS provides a simple way to interact with the spatial representation of the spreading of the COVID-19 pandemic and the associated variables in the Trentino Province. It is possible in this way to spot local clusters and to observe their variation in time. Using the slide control which selects the week it is possible to explore the dataset time wise, making it possible to appreciate the different phases of the pandemic, from the initial spreading period to the receding stages. At the same time, the

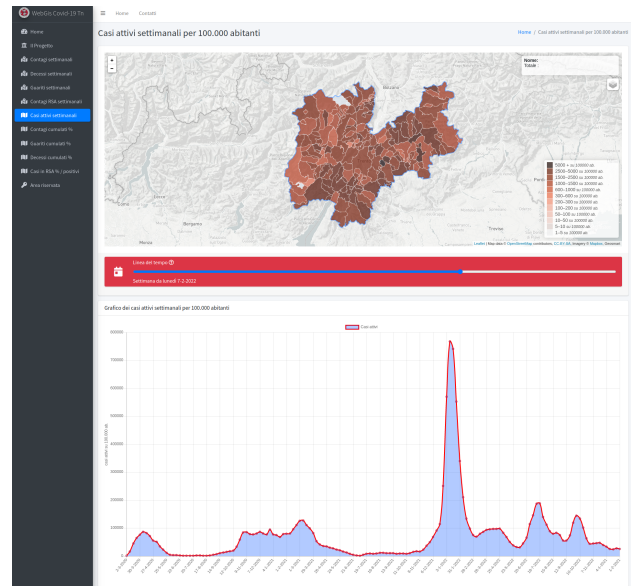


Figure 6. Weekly active cases (week from February the 7th, 2022).

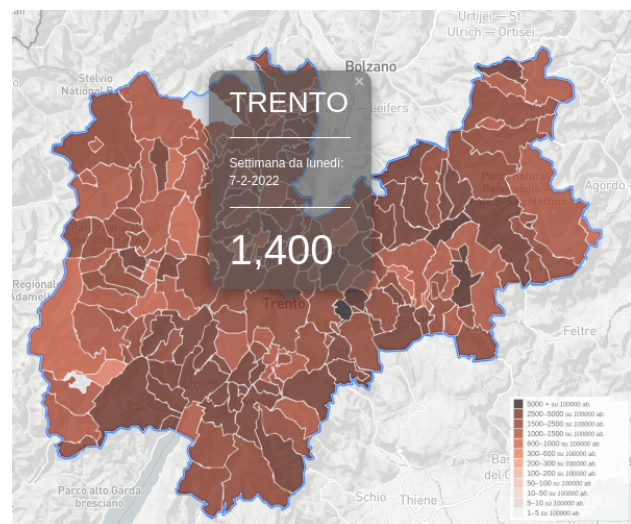


Figure 7. Popup for the municipality of Trento for weekly active cases (week from February the 7th, 2022).

graph representation of the timeline allow the user to detect the distribution of the peaks of the infections and the rate of abatement of the pandemic.

Both the general public and public health managers can benefit from information provided by spatial data on the epidemiological dynamics, such as the identification of locations that are more affected throughout various time periods, recovered concentrations, and mortality rates. The advance over static maps is self-evident, as data exploration across space and time allow the visual appreciation of clusters and trends.

A critical point in the creation of a database and a WebGIS for epidemic data is to find the optimal balance between the will to providing data with the highest possible resolution, both space-wise and time-wise, and the need of safeguarding patients' privacy, also in relation to the legal framework. For this reason the dataset uses the municipality as spatial unit and the week as time resolution.

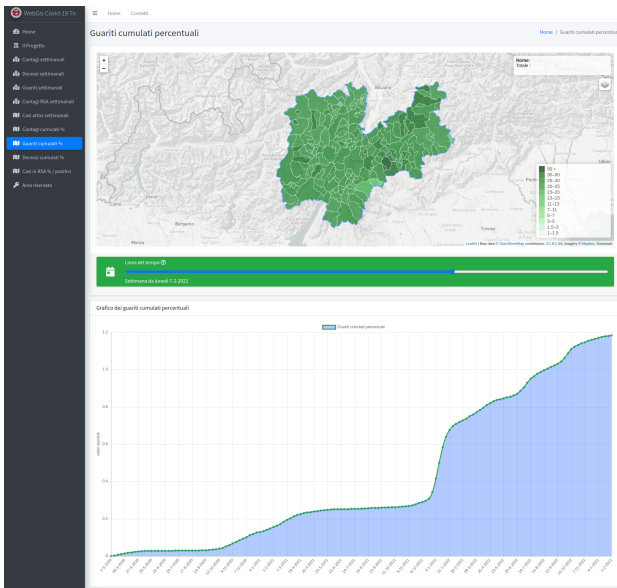


Figure 8. Weekly cumulative recovered (week from February the 7th, 2022).

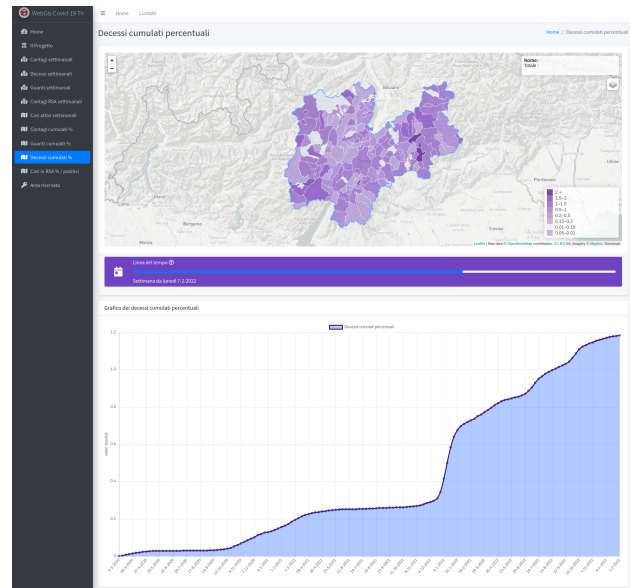


Figure 10. Weekly cumulative deaths (week from February the 7th, 2022).

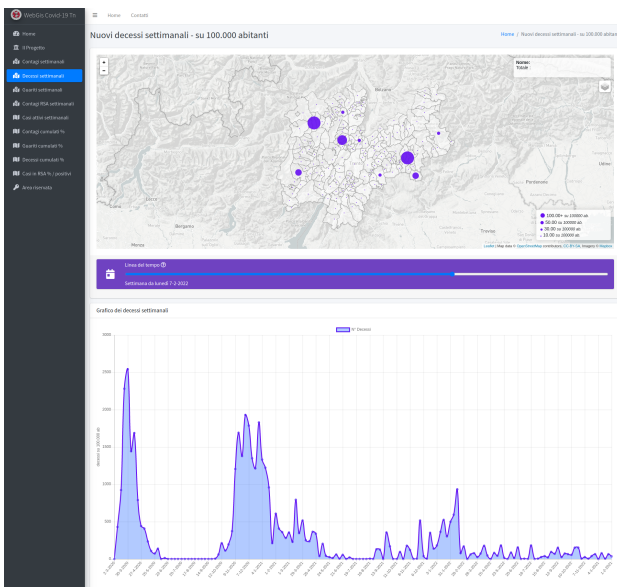


Figure 9. Weekly deaths (week from February the 7th, 2022).

The significant flexibility offered by FOSS for spatial analysis, such as QGIS, MySQL and the Leaflet and Bootstrap libraries, as well as the simplicity of their combined use, has been important for the installation and the setup of the system. Furthermore, the deployment of FOSS ensures the maintainability of the system, thanks to the constant availability of updated versions of the software. For these systems, the ability to access the source code to write custom scripts for some of the activities is essential.

The WebGIS is available at <http://covid19mappa-trentino.geco.unitn.it/geosmart/index.php>

6. ACKNOWLEDGEMENTS

This study was supported by the University of Trento, “Bando Strategico di Finanziamento COVID-19” 2020 Funding, GEO-

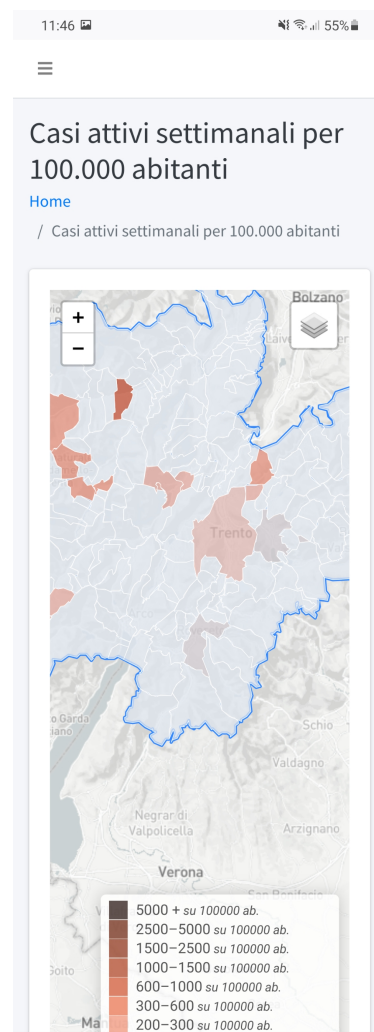


Figure 11. View on a mobile phone.

SMART Grant. The funding entity had no involvement in the study design and realization.

REFERENCES

- Agafonkin, V., 2023. Leaflet, an open-source javascript library for mobile-friendly interactive maps. leaflet. <https://leafletjs.com/> (22 March 2023).
- Apache, HTTP server project, 2023. Apache. <https://httpd.apache.org/> (22 March 2023).
- Bootstrap · The most popular HTML, CSS, and JS library in the world, 2023. Bootstrap. <https://getbootstrap.com/> (22 March 2023).
- Cannata, M., Strigaro, D., Spataro, A., Marotta, F., Achille, C., 2022. TOURISM, NATURAL PROTECTED AREAS AND OPEN SOURCE GEOSPATIAL TECHNOLOGIES. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 48(4/W1-2022), 81 – 88. doi.org/10.5194/isprs-archives-XLVIII-4-W1-2022-81-2022.
- Carballada, A. M., Balsa-Barreiro, J., 2021. Geospatial Analysis and Mapping Strategies for Fine-Grained and Detailed COVID-19 Data with GIS. *ISPRS International Journal of Geo-Information*, 10(9), 602. doi.org/10.3390/ijgi10090602.
- Ciulli, M., Federici, B., Ferrando, I., Marzocchi, R., Sguerso, D., Tattoni, C., Vitti, A., Zatelli, P., 2017. FOSS tools and applications for education in geospatial sciences. *ISPRS International Journal of Geo-Information*, 6(7). doi.org/10.3390/ijgi6070225.
- Coronavirus dashboards - Italian Civil Protection Department, 2023. Coronavirus dashboards. <https://mappe.protezionecivile.gov.it/en/emergencies-dashboards/coronavirus> (23 March 2023).
- Corradini, A., Peters, W., Pedrotti, L., Hebblewhite, M., Bragalanti, N., Tattoni, C., Ciulli, M., Cagnacci, F., 2021. Animal movements occurring during COVID-19 lockdown were predicted by connectivity models. *Global Ecology and Conservation*, 32. doi.org/10.1016/j.gecco.2021.e01895.
- Franch-Pardo, I., Napoletano, B. M., Rosete-Verges, F., Billa, L., 2020. Spatial analysis and GIS in the study of COVID-19. A review. *Science of The Total Environment*, 739, 140033. doi.org/10.1016/j.scitotenv.2020.140033.
- Gabellieri, N., Geri, F., Pranzo, A. M. R., Zatelli, P., 2021. Bridging research and dissemination in the CoViD-19 era: a WebGIS dashboard for the Autonomous Province of Trento (Italy). *Proceedings of the ICA*, 4, 34. doi.org/10.5194/ica-proc-4-34-2021.
- Gerbino, C., 2020. Emergenza COVID-19: dati, standard ed interoperabilità o solo dashboard? *GEOmedia*, 24(1). <https://mediageo.it/ojs/index.php/GEOmedia/article/view/1719>.
- Gobbi, S., Ciulli, M., La Porta, N., Rocchini, D., Tattoni, C., Zatelli, P., 2019. New Tools for the Classification and Filtering of Historical Maps. *ISPRS International Journal of Geo-Information*, 8(10), 455. doi.org/10.3390/ijgi8100455.
- Hu, T., Guan, W. W., Zhu, X., Shao, Y., Liu, L., Du, J., Liu, H., Zhou, H., Wang, J., She, B., Zhang, L., Li, Z., Wang, P., Tang, Y., Hou, R., Li, Y., Sha, D., Yang, Y., Lewis, B., Kakkar, D., Bao, S., 2020. Building an Open Resources Repository for COVID-19 Research. *Data and Information Management*, 4(3), 130 – 147. doi.org/10.2478/dim-2020-0012.
- Lipeng, J., Xuedong, Z., Jianqin, Z., Zhijie, X., Shaocun, D., 2020. Design and Development of a Visualization System for COVID-19 Simulation Based on WebGIS. *Proceedings - 2020 International Conference on Public Health and Data Science, ICPHDS 2020*, 278 – 282. doi.org/10.1109/ICPHDS51617.2020.00061.
- Mollalo, A., Vahedi, B., Rivera, K. M., 2020. GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. *Science of The Total Environment*, 728, 138884. doi.org/10.1016/j.scitotenv.2020.138884.
- Mooney, P., Juhász, L., 2020. Mapping COVID-19: How web-based maps contribute to the infodemic. *Dialogues in Human Geography*, 10(2), 265-270. <https://doi.org/10.1177/2043820620934926>.
- MySQL DBMS, 2023. MySQL. <https://www.mysql.com/> (22 March 2023).
- Patel, N. V., 2020. The best and the worst of the coronavirus dashboards. *MIT Technology Review*. <https://www.technologyreview.com/2020/03/06/905436/best-worst-coronavirus-dashboards/> (19 April 2023).
- Pranzo, A. M. R., Dai Prà, E., Besana, A., 2023. Epidemiological geography at work: An exploratory review about the overall findings of spatial analysis applied to the study of CoViD-19 propagation along the first pandemic year. *GeoJournal*, 88(1), 1103-1125. doi.org/10.1007/s10708-022-10601-y.
- Quinn, S., 2021. Using Free and Open Source software to teach university GIS courses online: lessons learned during a pandemic. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 46(4/W2-2021), 127 – 131. doi.org/10.5194/isprs-archives-XLVI-4-W2-2021-127-2021.
- Simeoni, L., Zatelli, P., Floretta, C., 2014. Field measurements in river embankments: Validation and management with spatial database and webGIS. *Natural Hazards*, 71(3), 1453-1473. doi.org/10.1007/s11069-013-0955-9.
- Tattoni, C., Soardi, E., Prosser, F., Odasso, M., Zatelli, P., Ciulli, M., 2019. Fruit availability for migratory birds: A GIS approach. *PeerJ*, 7. doi.org/10.7717/peerj.6394.
- Trentino coronavirus dashboard, 2023. Trentino coronavirus dashboard. <https://patn.maps.arcgis.com/apps/dashboards/345e288061094c248e44fd4f5c5cda3a> (23 March 2023).
- Trias-Llimós, S., Alustiza, A., Prats, C., Tobias, A., Riffe, T., 2020. The need for detailed COVID-19 data in Spain. *Lancet Public Health*, 5(11), e576. doi.org/10.1016/S2468-2667(20)30234-6.
- Zatelli, P., Gabellieri, N., Besana, A., 2022. Digitalization and Classification of Cesare Battisti's Atlas of 1915. *ISPRS International Journal of Geo-Information*, 11(4). doi.org/10.3390/ijgi11040238.

Zatelli, P., Gobbi, S., Tattoni, C., Cantiani, M., la Porta, N., Rocchini, D., Zorzi, N., Ciolli, M., 2019. Relevance of the cell neighborhood size in landscape metrics evaluation and free or open source software implementations. *ISPRS International Journal of Geo-Information*, 8(12). doi.org/10.3390/ijgi8120586.