Technical Report

ISPRS Open-Data Collector for Supporting Remote Sensing Analyses over Agriculture and Natural Ecosystems by Sharing Sampled Ground Truth (Shy)

Principle Investigator: Francesco Pirotti, University of Padova, Italy

Co-investigators: Mitsunori Yoshimura, Center for Spatial Information Science, The University of Tokyo, Japan; Jaime Hernandez P., Universidad de Chile, Chile; Brigitte Leblon, University of New Brunswick, Canada; Maria Antonia Brovelli, Politecnico di Milano, Italy; Meguni Yamashita, Tokyo University of Agriculture and Technology, Japan.

Abstract

Sampling data over areas covered with vegetation can be particularly expensive, time-consuming and complex (thick canopy/crop, wet areas, limited accessibility), but it is a crucial information for calibrating, correcting, training, validating and testing models and processes.

The implemented ISPRS-Shy spatial web-platform provides a user-friendly and guided interface for uploading spatial datasets with rich metadata regarding timestamp, description, authors, license, links and references to related literature (DOIs). Both metadata and spatial information is searchable to allow users to find data that might support their research. Contributors can choose to leave their data fully open to public, also to non-registered users, or limited to specific projects or groups of registered users. It is expected that the effort of the data owners is rewarded through dissemination of scientific products and public engagement through ISPRS-Shy.

Currently the platform is available at the link https://isprs-shy.cirgeo.unipd.it/ and has 20 registered users and data of three projects from national and international funding with a total of 47 datasets.

Specific details are available online on the ISPRS Archives publication dedicated to this initiative at the link https://doi.org/10.5194/isprs -archives-XLIII-B5-2022-45-2022

Report

The 2021 Scientific Initiatives in ISPRS funded this project called ISRS-SHY from “SHare mY ground truth”. It was intended as a collector of geographic data to support image analysis by sharing the necessary ground truth data needed for rigorous analysis. Regression and classification tasks that use remote sensing imagery necessarily require some control on the ground. The rationale behind this project is that often data on the ground is collected during projects, but is not valued by sharing across projects and teams globally. Internet has improved the way that data are shared, but there are still limitations related to discoverability of the data and its integrity. In other words, data are usually kept in local storage or, if in an accessible server, they are not documented and therefore they will not be picked up during search. In this initiative we created a portal using the Geonode environment to provide a hub for sharing data between research groups and openly to the community. The portal was then tested within the framework of three projects, with several participants each. The data that was uploaded and shared covered all types of geographic data formats and sizes. Further sharing was done in the context of teaching activities in higher education.

The results show the importance of creating easy means to find data and share it across stakeholders. Qualitative results are discussed, and future steps will focus on quantitative assessment of the portal's usage, e.g. number of registered users in time, number of visits, and other key performance indicators. The results of this project are to be considered also in light of the effort in the scientific community to make research data available, i.e. FAIR - Findability, Accessibility, Interoperability, and Reuse of digital assets.

SHY aimed at providing a user-friendly and guided interface for uploading spatial datasets and respective metadata, e.g. date, description, authors, license, links to references to literature. Both metadata and spatial information will be searchable to allow users to find data that might support their research.

SHY has been installed in a remote machine with 128 MB of RAM and 16 available processors and a Linux Ubuntu 20.4 operating system. Implementation consisted on extending an open source project, Geonode. Geonode is an "Open Source Geospatial Content Management System - a web-based application and platform for developing geospatial information systems (GIS) and for deploying spatial data infrastructures (SDI)". It is based on Python and other open source projects, such as Geoserver and PostgreSQL for data management. The structure of the main system components, Geonode, Geoserver and PostgreSQL, is organized in different Docker containers for easier management and backup. For now, anyone can sign-up via normal registration or using social accounts from Linkedin or Facebook.

Contributors can choose to upload full datasets or only the description and metadata of geodata shared via an external link or with a contact for requesting the full dataset. This option is ideal for very large datasets that might be problematic to upload. One example for the latter option that is already included in SHY is a UAV survey with more than 1000 images and the resulting point clouds with ~1 billion points; the actual data are stored respectively in a Google Drive folder and in an online viewer (*Potree*) that are searchable and linked in the portal. Users can either download the data directly or send an email to authors asking for the full dataset, in case only metadata and author reference is provided.

Three international projects were used as pilots. These projects collected different types of data and were used to test the portal. The first project is VARCITIES a European Union Horizon 2020 Research and Innovation programme. This programme includes data from IoT sensors, position data from smartphones and a 3D model from a laser scanning drone survey. The overall objective of the VARCITIES partners that use ISPRS-SHY is to study the well-being of visitors of the historical garden of a cultural heritage site, Villa Revedin Bolasco. To this aim the sensors will collect microclimatic and air quality data at static points in the area and also via a mobile rover that will carry a set of sensors around the garden. The drone flight collected 3D point clouds from three types of surveys, two with lidar sensors (Riegl VUX-3 and VUX-120) and one with photogrammetry. This produced two point clouds consisting of 1 billion points, and one point cloud with 300 million points.

As mentioned, the large volume of data that point clouds represent can be loaded in the portal by linking a viewer, Potree, to the metadata that is loaded in ISPRS-SHY. With this solution, users can discover the data by using keywords or the spatial location, but the data itself is located in a third party server. This solution optimizes the data mining requirements with the space requirements of large volumes of data. It is of course obvious that uploading to the ISPRS-SHY portal several GB of

point cloud data is not an effective solution, also considering that the Geonode infrastructure provides viewers for raster and vector data, but not for 3D point clouds. Therefore integration of a separated viewer (Potree) using the Geonode infrastructure to link the data viewer with the metadata, improves efficiency. Three-dimensional data are becoming more common in the geospatial community, and Geonode and other data sharing solutions would get much added value if viewers for this type of data can be added. Potree is a very versatile solution that might be embedded in Geonode. An upload system to deal also with 3D data formats, e.g. LAS/LAZ format is feasible to implement to stream LAS format to Potree structure and automatically create the viewer. This might be a future development.

Another project is a national project that aimed at analysing the effects of an extreme event, the VAIA storm. VAIA damaged large parts of production forests in the north-east part of Italy. This led to lying deadwood that will cause further risks related to fire and pathogens (e.g. bark beetle attacks). The VAIA project not only has geographic data and maps loaded in the portal, but also links to non-spatial data such as newspaper articles, and the proposal itself.

Remote sensing data and ground surveys have been stored in the ISPRS-SHY platform to cross-analyse information. Not only products from ad-hoc processing, but also parts of existing satellite-derived products, such as the Forest Cover dataset by (Hansen et al., 2013). This allows users to access data relative to the area and to the event, i.e. in this case the forest cover loss caused by the VAIA storm, and visualize them.

The third project is from the Chilean CONAF “Postulación Fondo de Investigación del Bosque Nativo" titled “Indicadores fenológicos y estructurales de alteración de hábitat en bosques de araucaria”. In this project 12 stands of Araucania-Nothofagus forest were selected in southern Chile, which represented four alteration levels. Drone flights acquired images that produced ortho-imagery and point clouds that were further analysed in the project. A total of 379 plots over the 12 stands collected ground data regarding tree species, and other tree parameters. Six users accessed ISPRS-SHY to load and share the project data. The data from this survey was saved in a table connected to the geolocations of the plots that were surveyed with a GNSS receiver and stored as vector points with a unique id for each plot. ISPRS-SHY allowed to view the plot position and the data relative to each plot.

Each dataset that is uploaded becomes a spatial layer or a document layer, depending on the type of dataset. Non-spatial datasets can be anything related to the project but without direct spatial information, e.g. newspaper articles, research papers. These types of data can be linked to spatial data and maps from the same project, so they can indirectly be assigned to a spatial object. Accessing the data is immediate as a link can be provided to users that want to download or simply view the data with the online tools. Depending on the data sharing policy that is given to each layer, it can be shared also as open access, without requiring the user to register to the platform, or to specific users. The user profiling method allows to assign to users projects to that there is a many-to-many relationship between users and projects.

As concluding remarks, thirty-nine items are shared, consisting in geodata, articles and maps that aggregate geodata layers in maps that can be accessed via browser. Twenty users have tested the platform to the beginning of November 2022, thus after almost one year since the platform prototype was put online. Some initial considerations are worth reporting. One is that registration via social media seems a trivial addition, but definitively increases the accessibility to the portal. Since sharing is an important aspect, faster access helps to attract users. Another consideration is that correct and detailed metadata completion allows better data mining. Often users tend to upload data without

providing accurate information, making it difficult to search for the data as keywords are looked up from the inserted metadata. The title of the uploaded product is important as well, as it is the first text that is associated with the product.

The evolution of the ISPRS-SHY portal is to foster more collaboration between research groups by sharing more data like benchmarks (ISPRS and other societies) and projects. Some ongoing projects already involved in ISPRS-SHY (VARCITIES) will be pilot-cases for integration of further geodata. In particular in the near future the implementation of IoT gateways and devices for this project will be uploaded and data streamed and integrated via PostgreSQL. This area will be monitored over several years for air quality and climate information with data that has to be shared with other partners, and ISPRS-SHY will be the central node for this goal.