



Publishing Eurac Research data on the GEOSS Platform

Roberto Roncella, Bartolomeo Ventura, Andrea Vianello, Enrico Boldrini, Mattia Santoro, Paolo Mazzetti & Stefano Nativi

To cite this article: Roberto Roncella, Bartolomeo Ventura, Andrea Vianello, Enrico Boldrini, Mattia Santoro, Paolo Mazzetti & Stefano Nativi (2023) Publishing Eurac Research data on the GEOSS Platform, Big Earth Data, 7:2, 428-450, DOI: [10.1080/20964471.2023.2187659](https://doi.org/10.1080/20964471.2023.2187659)

To link to this article: <https://doi.org/10.1080/20964471.2023.2187659>



© 2023 The Author(s). Published by Taylor & Francis Group and Science Press on behalf of the International Society for Digital Earth, supported by the International Research Center of Big Data for Sustainable Development Goals, and CASEarth Strategic Priority Research Programme.



[View supplementary material](#)



Published online: 20 Mar 2023.



[Submit your article to this journal](#)



Article views: 1161



[View related articles](#)



[View Crossmark data](#)



Citing articles: 1 [View citing articles](#)

Publishing Eurac Research data on the GEOSS Platform

Roberto Roncella ^a, Bartolomeo Ventura ^b, Andrea Vianello ^c, Enrico Boldrini ^a,
Mattia Santoro ^a, Paolo Mazzetti ^a and Stefano Nativi ^a

^aNational Research Council of Italy – Institute of Atmospheric Pollution Research, Division of Florence – Via Madonna del Piano 10, Florence, Italy; ^bInstitute for Earth Observation, Eurac Research – Viale Druso 1, Bozen, Italy; ^cCenter for Sensing Solution, Eurac Research – Viale Druso 1, Bozen, Italy

ABSTRACT

This paper is the third of a series that introduces some of the main dataset resources presently shared through the GEOSS Platform. The GEOSS Platform is a brokering infrastructure that brokers more than 190 autonomous information systems and data catalogs; it was created to provide the technological tool to implement the Global Earth Observation System of Systems (GEOSS). This manuscript focuses on the analysis of Eurac Research datasets and illustrates the data publishing process to enroll the Eurac Research Data Provider to the GEOSS Platform through the administrative and technical registrations. The study provides an analysis of the GEOSS user searches for Eurac Research data in order to understand the main use of datasets of an important Data Provider.

ARTICLE HISTORY

Received 2 November 2022

Accepted 1 March 2023

KEYWORDS

GEOSS; data interoperability; data sharing; Eurac Research; Earth observation

1. Introduction

This is the third publication of a series of manuscripts introducing some significant datasets, which are currently published and accessible through the GEOSS Platform. GEOSS (the Global Earth Observation System of Systems) is a social and software ecosystem sharing independent and open Earth observation (EO) data, information, and processing services. The GEOSS Platform (formerly known as GCI: GEOSS Common Infrastructure) is the interoperability cornerstone enabling the GEOSS ecosystem, which is contributed by many enterprise systems that are managed by the GEO members (Boldrini et al., 2021; Craglia et al., 2017; Nativi et al., 2015). After more than 15 years, the GEOSS concept is evolving looking at the Digital Twin pattern, which is facilitated by a flexible and scalable digital ecosystem (Guo et al., 2020; Nativi & Craglia, 2021; Nativi & Mazzetti, 2021; Nativi et al., 2021; Santoro et al., 2020). The evolution of the new GEOSS Platform will also enable model sharing such as demonstrated in the proof-of-concept during the GEO Plenary held in Canberra in 2019 (Ollier, 2019).

GEOSS was initiated and is operated by the Group on Earth Observation (GEO).¹ GEO is an intergovernmental partnership working to improve the availability, access, and use of

CONTACT Roberto Roncella  roberto.roncella@cnr.it  National Research Council of Italy – Institute of Atmospheric Pollution Research, Division of Florence – Via Madonna del Piano 10, Florence, Italy; Bartolomeo Ventura  bartolomeo.ventura@eurac.edu  Institute for Earth Observation Eurac Research – Viale Druso 1, Bozen, Italy
 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/20964471.2023.2187659>.

© 2023 The Author(s). Published by Taylor & Francis Group and Science Press on behalf of the International Society for Digital Earth, supported by the International Research Center of Big Data for Sustainable Development Goals, and CASEarth Strategic Priority Research Programme. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

open Earth observations, including satellite imagery, remote sensing and in situ data, to impact policy and decision-making in a wide range of sectors (GEO, 2022b).

According to a recent study focusing on the GEOSS data content (Boldrini et al., 2021), remote sensing is a predominant item among other major thematic areas such as human activities, hydrology, climate, pollution, geology, meteorology, oceanography and sustainable development goals. The top data providers contributing with the most records are the European Commission, the Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI), the United States Geological Survey (USGS), China GEOSS, and the Incorporated Research Institutions for Seismology (IRIS). Nevertheless, an important added value of GEOSS is also to be a data hub designed for lesser known data providers sharing a less amount of data, but more specific and not elsewhere discoverable.

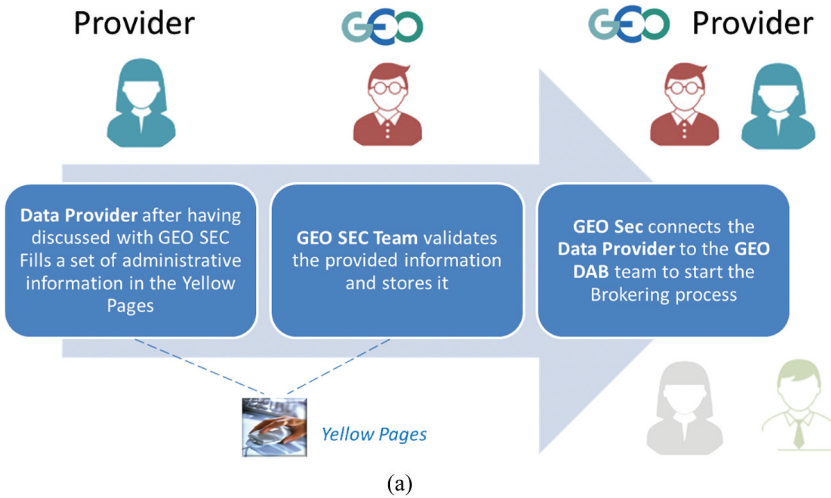
Established in 2005, today, GEO is a partnership of more than 100 national governments and more than 100 Participating Organizations. GEO, GEOSS, and the GEOSS Platform were fully introduced in the first two manuscripts of the series (Roncella, Zhang, et al., 2022; Roncella, Boldrini et al., 2022). The China satellite data represented in the first published paper is one of the leading satellite data providers. It also consolidates China's participation in GEOSS, making it an important strategic resource for the GEOSS Platform. The second manuscript describes the NextGEOSS Catalog based on the Horizon 2020 European framework program, with the aim to build a European data hub for Earth Observation, capable of collecting satellite data and in situ data from several different scientific communities, helping users to ensure food security, monitoring air pollution and contributing to sustainable urban development. In this third paper, we will focus on the introduction of the Eurac Research Datasets. The main features of this interesting use case are the great variety of dataset collections and the fact that they source from a private research center.

1.1. *Becoming a provider of GEOSS ecosystem and publishing datasets*

The organizations that want to join the GEOSS ecosystems and contribute their datasets through the GEOSS Platform must follow a rather simple procedure, consisting of a couple of processes: a registration and an interoperability one (GEOSS Infrastructure Development Task Team, 2017). The procedure was already described in details in the first manuscript of the series (Roncella, Zhang, et al., 2022). For the Readers' convenience, the simplified utilized schemas representing the procedure are depicted again also in this manuscript – see [Figure 1](#). The interoperability process (and hence the GEOSS ecosystem) applies a service brokering approach allowing enterprise data systems free to evolve and keep their autonomy (Craglia et al., 2011; Nativi et al., 2013; Santoro et al., 2016).

Both the registration and interoperability processes are managed by the GEO Secretariat (GEOSec), in keeping with the GEOSS data sharing and data management principles (GEO, 2015). These principles address the needs of discoverability, accessibility, usability, preservation, and the curation of data and their related shared resources. In 2022, the GEO Data Working group introduced some new guidelines supporting the implementation of these principles for the management of all Earth Observation data products and services (GEO, 2022a).

Workflow: Administrative Registration



Workflow: Brokering Process

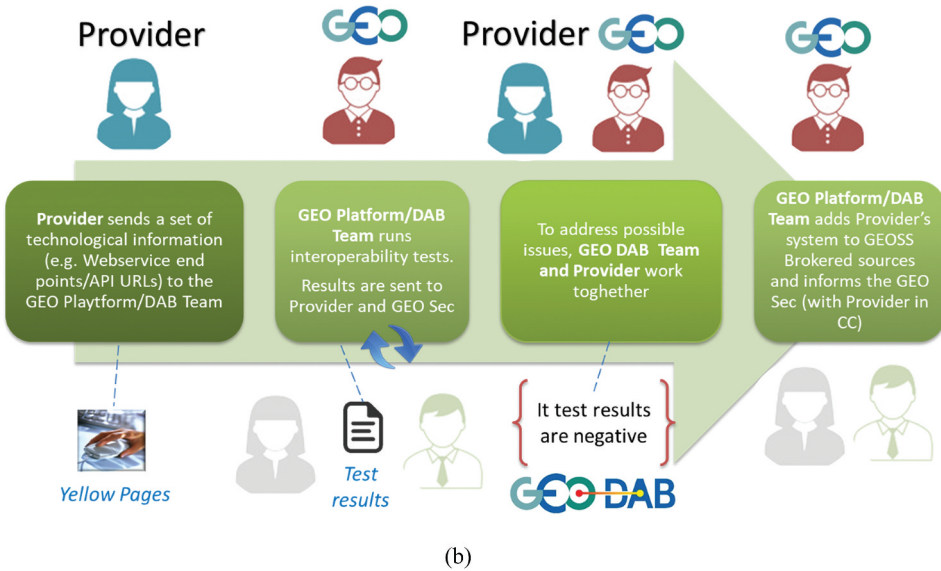


Figure 1. The two steps constituting the process to become a data/service provider of the GEOSS platform: (a) depicts the administrative registration and (b) represents the technological registration (GIDTT, 2017).

The next section will introduce the Eurac Research Datasets. Section three will cover the implementation of the registration and interoperability processes. The conclusion section will discuss the significant contribution made by the Eurac Research Data Provider presenting an analysis of relevant statistics information on data usage.

2. Eurac Research datasets

Eurac Research is composed of 18 research institutes and competence centers that study, monitor, and forecast the environment and all the dynamics of the world around us, at different scales, such as snow coverage, soil moisture, forest change, drought risk, etc. It collects, manages, and processes many datasets and information coming from different platforms such as satellites (remote sensing), UAVs (proximal sensing), and ground stations (in situ sensing).

The datasets collected are the results of the work of our researchers in many national and European projects in the last 30 years. Each project is organized independently according to its regulations and partners.

Because of the heterogeneous topics and sources, data are usually of different data formats and require to be organized in diverse data repositories. For this, it is not easy for researchers to discover datasets and access them simultaneously. Consequently, it is crucial to organize data and research outputs so they can be more easily accessed, exchanged, and reused, providing clear and complete documentation. The first step toward a new research output is always to discover the necessary input datasets.

Currently we publish more than 400 datasets publicly available, harvested by the GEOSS catalogue. We manage much more datasets in our Institute that we are going to publish if they will be considered useful and of common interest.

2.1. Main data repositories

Datasets collected by Eurac Research can be grouped into three main categories: time series from ground stations measuring physical parameters such as Normalized Difference Vegetation Index (NDVI), soil moisture, and air temperature; time series raster data from remote sensing or from data processing describing, for example, snowmelt, run-off, vegetation phenology; vector layers derived from data processing or from external partners to represent and describe environmental and artificial phenomena or structures, such as building density per municipality, climatic factors, environmental indexes (NDVI, drought), social indexes, traffic monitoring, etc.

These three diverse types of data are collected and managed using the following different, independent, and open-source repositories: OpenEO-API² (big raster time series), Geonode³ platform (vector and raster layers), SOS 52° North⁴ and Databases (sensor time series and other tabular datasets). To gather metadata from these repositories, we use a Geonetwork⁵ (GN) catalog application that harvests automatically for new or updated entries, providing a single access point for discoverability for researchers and external users. It is extremely useful to discover about 400 datasets by using filters for time, topic, geographical area, etc. It provides the Catalogue Service for the Web (CSW) standard used by our custom graphical user interface, called the Environmental Data Platform (EDP) catalog, and by external web portals such as the GEOSS data portal and

other ones in the local community. All metadata describe datasets using standard fields such as title, abstract, contacts, authors, license, etc., to provide complete information about the resources.

We chose the Geonetwork application for our catalogue because it is one of the most robust applications available used by the open-source community. Currently, all our metadata are publicly available, but some datasets may have restrictions for download.

The main idea of openEO (<https://openeo.eurac.edu/>) is to create an API for interoperability between Earth Observation data centers, supporting actions like data discovery, access, processing, and retrieval as well as authentication, access control, and billing. This is realized on the client API level. Client APIs are linked through the core API to the backend drivers. The whole interaction is designed via REST calls between client to core and core to corresponding back-ends. In Eurac, we use openEO to share large datasets from remote sensing or generated by data processing that are organized in our datacube repositories. There are about 60 collections publicly available by the API for processing on the server or through our web interface. Additional collections are available after authentication for our authorized users.

Maps-portal (<https://maps.eurac.edu/>), based on Geonode software, is a data management web application to manage and share spatial datasets and maps. It is based on Geoserver, PostgreSQL, Mapstore, PyCSW, and other components to provide all the functionalities available. It allows, for example, to upload resources (shapefiles, GeoTIFF, time-series vector layers), to set permissions for your resources and maps (view, download, etc.), to create working groups to share resources only with your project's partners/colleagues and to access data using WMS by desktop GIS client. The idea is to provide a self-service application to let users be autonomous in all the necessary steps from the upload to the metadata compilation, maps creation, etc. On this portal, our users uploaded more than 250 datasets and 45 maps. This platform allows our researchers to share a common environment to set up a simple WebGIS to share datasets belonging to different projects, reducing the cost of the WebGIS setup and maintenance. Authenticated users can access more datasets, according to the membership groups. Users can create working groups for specific projects, or groups of datasets, to easily restrict permission on the uploaded resources.

Fifty-two degree North Sensor Observation Service (SOS) is a reference implementation of the OGC Sensor Observation Service specification (version 2.0). It provides an environment and web interface to manage data from sensors and share them through a REST API and store data in a PostgreSQL database. In Eurac Research, we use them to collect and share datasets from ground stations for projects that require a standard API to share the data. We developed, in the MONALISA project, an automatic system (Ventura et al., 2019) to upload data collected in the FTP server to the 52° North SOS by web services to simplify the database filling. In this application, we collected data from about 100 ground stations and more than 75 different phenomena such as soil temperature, soil water content, wind speed, solar radiation, etc.

All other datasets we manage are organized in databases using PostgreSQL and Influx-DB. These two solutions are very robust, largely used, and supported by the open-source software community. We use Influx-DB for time series datasets coming from sensor or IoT devices to have a powerful web API to store datasets and to set up web applications. All other datasets are stored in the PostgreSQL database which is a more versatile database.

InfluxDB is a time series platform that has grown a lot in the last years and can provide powerful API & toolsets for real-time applications, a high-performance time series engine, and a massive community of cloud and open-source developers. We use it for automatic data ingestion from remote sensors and IoT devices. It does not provide a real standard API, but it is largely used for this type of application and offers a ready-to-use platform for data management and visualization.

PostgreSQL is an open-source relational database management system (RDBMS) that we use to organize data and share them with our researchers as input datasets for their work. We store here many diverse types of datasets such as spatial data (especially vector data), time-series data coming from ground station sensors but also from remote sensing analysis (NDVI, drought indices, etc.), and other descriptive tabular data. This RDBMS system is largely used and has good documentation and support from the community. It does not provide a standard web service, but it is possible to connect to the DB in many by many common clients like R, MatLab, Python, QGIS, and many frameworks to develop web applications.

2.2. Metadata workflow

To collect and share metadata from these three main repositories, we use the GeoNetwork (GN) metadata catalog, an open-source platform that allows us to manage all the metadata and provide an API to external web portals and clients. GN can harvest metadata from the Maps, consuming directly its CSW standard. On the contrary, for openEO and SOS metadata, it was necessary to develop some scripts to convert them into a compliant schema accepted by GeoNetwork. Further metadata, describing datasets not managed by the previous three repositories, are created manually directly in GN web interface. Most of these datasets consist of tables of databases concerning environmental datasets such as climatic data or data from sensors or other research outputs.

The main idea is to collect all the available data sets in a unique internal EDP-catalog (<https://edp-portal.eurac.edu/discovery/>) able to assist the user to discover information, process, and analyze data systematically and efficiently. This approach has a series of advantages such as a rapid search for data availability, avoiding duplicates, increasing visibility, and DOI information, performing the cross-correlated search, to use standard-based metadata. To access the EDP it is not necessary to be registered but if you would like to benefit from all functionalities of the OpenEO editor, the Jupiter lab, or other web tools, authentication is required. To make datasets findable, we decided to use both one of the latest specifications developed, SpatioTemporal Asset Catalog (STAC),⁶ and the catalog application GeoNetwork.

STAC evolved from different organizations coming together to increase the interoperability of searching for satellite imagery. Geonetwork is a catalog application to manage spatially referenced resources. It provides powerful metadata editing and search functions as well as an interactive web map viewer. It is currently used in numerous Spatial Data Infrastructure initiatives across the world. Both STAC and Geonetwork are open-source, and users can easily interact with the community to suggest improvements and/or contribute with new ideas.

STAC specification provides a method to describe a range of geospatial metadata information, so it can more easily be indexed and discovered. The primary module for this

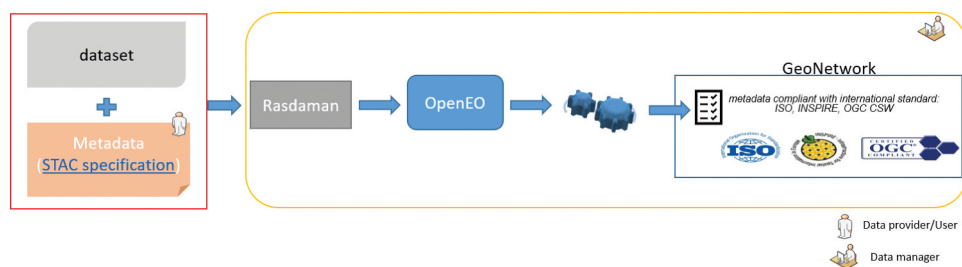


Figure 2. Workflow to transform STAC metadata into ISO standards, supported by GeoNetwork.

service is a spatiotemporal asset, which is any file that represents information about a planetary body acquired at a specific location and time. A JavaScript Object Notation (JSON) document is also created for the asset data, which is simple, straightforward, and can be customized using specific extensions related to the field of interest of the described data. Geonetwork offers the possibility to describe data using standards defined by the ISO and OGC community. Each metadata has a series of fields to be filled (some are mandatory) by the owner or producer of the data. Once the metadata is complete, it is possible to perform validation with the standard of reference to verify if it is fully compliant or if there are some errors concerning schematron rules. To link together these two different catalogs, an automatic script converts STAC metadata into GeoNetwork ones. This script is described in [Figure 2](#).

Once the JSON file describing the collection is imported into a data service, for example, Rasdaman⁷, this last one can be queried using the OpenEO API where the metadata is accessible but still in JSON format. As Geonetwork is not able to harvest metadata in JSON format, we use an XSL Transformation. The use of an XSLT⁸ is fundamental. XSL stands for Extensible Stylesheet Language, and it is a stylesheet language for XML documents. The main advantage of using an XSLT consists in having the possibility of adapting all the information written in the STAC metadata in the corresponding standard to be used or directly customized to the requirements of a specific project or goal. An example is illustrated in [Figure 3](#) where a section of an STAC metadata and the corresponding XSL commands are shown.

In this way, it is possible to have metadata that can be queried both using STAC, and an XML file for a CSW catalog, compliant with the OGC standards as well as for INSPIRE and ISO-19139.

Once the metadata is ready and compliant with international standards, it can be combined and exchanged. The metadata is fully accessible even when, in a certain case, data have a more restricted license. Utilizing an internal GeoNetwork tool, it is possible to register a dedicated DOI, connecting to the DataCite API, to make it uniquely identifiable. We contributed to its improvement by editing the existing code to let the tool work with the ISO 19115-3:2018 schema, required for our datasets.

2.3. Implementation of the FAIR data principles

In the last few years, Eurac Research started a process to foster the use of the FAIR data principles: findable, accessible, interoperable, and reusable (Wilkinson et al., 2016). This

Data discovery Interoperability

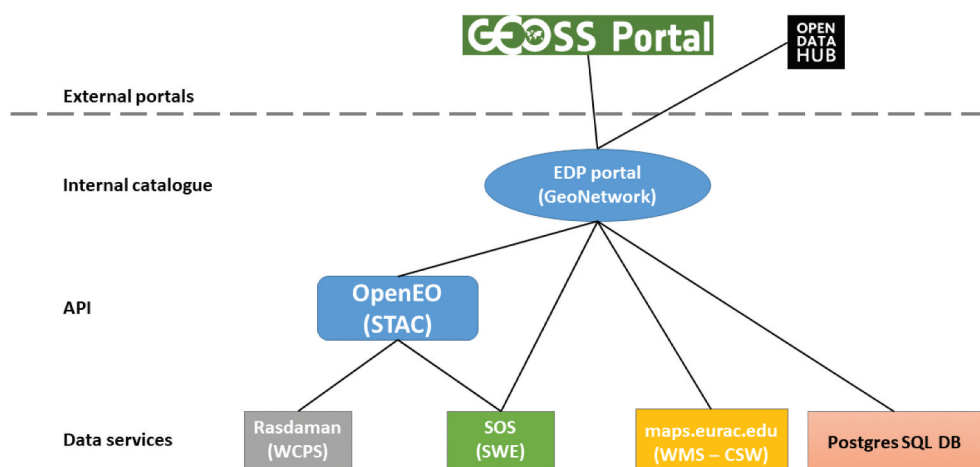


Figure 4. Concept of the data discovery interoperability developed in the Environmental Data Platform.

The EDP also offers us the possibility to write snippet code and documentation and link them to a specific resource type. Users will find this information on the metadata page of the resources.

Figure 4 describes how Eurac Research developed the FAIR principles in the framework of the projects Data Platform and Sensing Technologies for Environmental Sensing LAB (DPS4ESLAB) and openEO.

3. Dataset publishing through the platform

3.1. Administrative (Yellow Page) registration

The administrative registration (see supplemental Annex A) has been completed in May 2017 by the Eurac Research Data Provider. Table 1 shows the required information and the contents filled by the Eurac Research Data Provider.

Protected personal data (contact points names and emails) are not available in agreement with the General Data Protection Regulation (GDPR). The service endpoint is omitted due to policy reasons and to its different status since its start.

3.2. Interoperability (GEO DAB) registration

The interoperability registration is a brokering process that allows to execute interoperability tests on the remote data system in terms of discoverability and accessibility. The tests are conducted jointly with the Data Provider. Feedback and implementation cycles are repeated until the catalog can be fully added to GEOSS.

The component in charge of executing the technical interoperability registration is the GEO DAB, i.e. a middleware software framework based on the open-source technology Discovery and Access Broker (DAB)⁹. The DAB has been developed and applied in the

Table 1. Administrative registration for Eurac Research data provider.

Provider Name*	Eurac Research
Acronym	EURAC
Short Description*	Earth observation from space has opened up new perspectives for environmental research. At the Institute for Earth Observation, we are fusing remote sensing techniques with interdisciplinary approaches for integrated environmental monitoring and assessment of mountain regions. Using satellite data, we monitor the state and temporal evolution of the land surface over large areas at short time intervals. Integrating this with in situ data and models allows us to monitor and understand the key processes at work in mountain regions, such as snow cover evolution, vegetation development or landslide movements.
URL of the website institution	https://www.eurac.edu
Geographical coverage to the organization*	Regional, National, Local, European
If you have selected "National" and/or "Regional" in the previous question, please give the names of countries and/or regions	Italy, South Tyrol, Alpine region
GEO affiliation*	Non GEO
What is the name of the Official Focal point?*	OMITTED
What is the email of the Official Focal point?*	OMITTED
What is the name of the Technical focal point?*	OMITTED
What is the email of the Technical focal point?*	OMITTED
Type of online resource?*	Data (including imagery)
Data accessibility*	Yes - without restrictions
Data policy*	Restricted
GEOSS Data Management Principles label	<ul style="list-style-type: none"> ● Discoverable ● Accessible ● Standard encoding using ● Well-documented metadata ● Traceable ● Preserved ● Periodically verified ● Reviewed and refreshed ● Tagged with permanent ID ● Biodiversity and Ecosystem Sustainability ● Disaster Resilience ● Energy and Mineral Resource Management http://earthobservations.org/area.php?a=emrm ● Food Security and Sustainable Agriculture ● Infrastructure and Transportation Management ● Sustainable Urban Development ● Water Resources Management ● Goal 3: Good Health and Well-being ● Goal 8: Decent Work and Economic Growth ● Goal 9: Industry, Innovation, and Infrastructure ● Goal 11: Sustainable Cities and Communities ● Goal 12: Responsible Consumption and Production ● Goal 13: Climate Action ● Goal 15: Life on Land
Relevant SBA	
Relevant SDG	
Other initiative	-
Service endpoint*	OMITTED
Organization Logo URL*	

* Mandatory fields.

context of several international initiatives, such as the WMO Hydrological Observing System (WHOS)¹⁰ (Boldrini et al., 2022), the Ocean Data Interoperability Platform (ODIP)¹¹ and the European Marine Observation and Data Network Partnership for China and Europe (EMOD-PACE)¹² to enable discover and access functionalities across several

heterogeneous Data Provider systems, often in domain-specific scenarios (i.e. hydrology and oceanography).

GEO DAB is able to interconnect the heterogeneous and distributed capacities contributing to the multidisciplinary GEOSS environment, realizing an abstract and harmonized view of the different data/metadata by mapping the remote specific data provider models into the GEO DAB internal model. The GEO DAB internal metadata model is based on ISO 19115 which is a rich and extensible metadata standard defining more than 400 elements to describe geographical datasets in detail. Moreover, ISO 19115 allows an extension mechanism useful to define new concepts and to create additional metadata elements and related attributes. The GEO DAB uses the internal metadata model to harmonize and consolidate the information from the different heterogeneous data provider systems; it periodically harvests data provider services fetching the original metadata records, harmonizing them, and storing the information to a central database. This process allows an efficient discovery of records across various GEOSS Data Providers.

The GEO DAB brokers two different services from the Eurac Research Data Provider: the Geonetwork catalog and the SOS catalog. The Geonetwork catalog exposes several different APIs to interact with the datasets: REST API, OGC Catalogue Services for the Web (CSW), OpenSearch, etc. The GEO DAB uses the CSW interface (version 2.0.2) to communicate with the Eurac Geonetwork catalog.

The SOS standard service defines the interface to interact with sensor observations; like most of the OGC standards, the SOS service is based on the exchange of standard messages. The GEO DAB sends requests to the SOS service through an HTTP GET method specifying the request type (GetCapabilities, DescribeSensor or GetObservation) and the relative allowed parameters. The Eurac SOS service responds with XML file compliant with the specification (versions 1.0 or 2.0 are both supported). All the metadata records present in the Eurac Research data systems are expressed in XML format and are harvested through the HTTP protocol.

Since the Eurac Research Data Provider uses well-known and widely used standard protocols for managing and sharing geo-referenced data, the integration process of the catalogs into the GEOSS Platform was smooth and straightforward. In particular, no major issues were identified during the interoperability tests, and the datasets were published rapidly into the GEOSS Platform.

As of today, the GEO DAB publishes 300 datasets from the Geonetwork catalog and 27 datasets from the SOS service. The GEO DAB has a mechanism that allows to schedule a harvesting frequency period for all the data providers present in the GEOSS Platform. The default harvest period is 30 days, but the frequency (monthly, weekly, daily) can be customized to meet the needs of the different Data Providers.

For the Eurac catalogs, we scheduled a weekly harvest frequency period (i.e. the GEO DAB automatically searches new records from the remote Eurac Research Data Provider once every 7 days).

In terms of discoverability, the GEOSS Platform allows users to retrieve datasets through queries based on geographical coverage, temporal extent, and others. [Figure 5](#) shows the Eurac Research datasets available in the GEOSS Portal.

The GEO DAB provides thumbnails/preview information and Web Map Service (WMS) links for some datasets in the Eurac catalog. This allows the online visualization of the discovered records. In terms of accessibility, the GEOSS Platform allows users to download

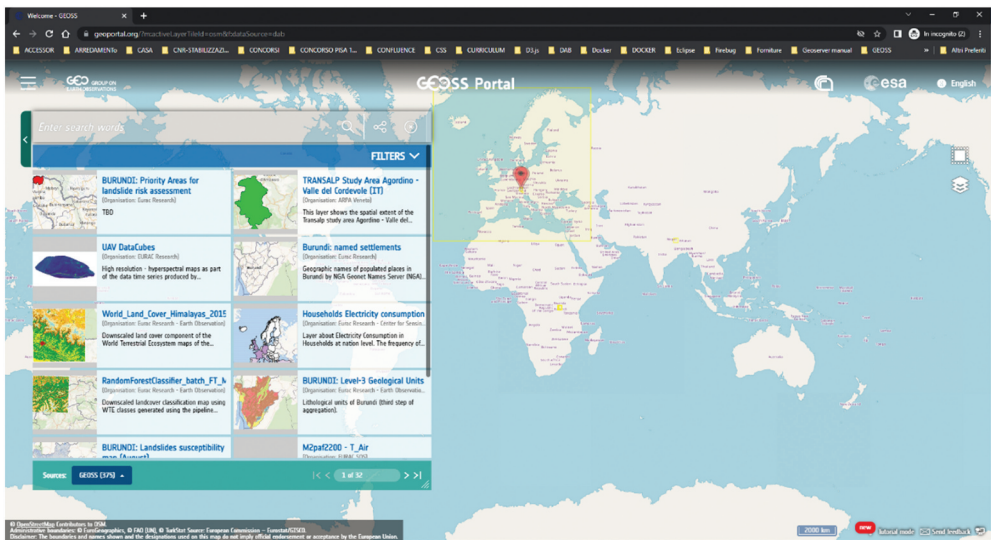


Figure 5. Eurac research datasets in the GEOSS portal.

the Eurac Research data products through the HTTP/HTTPS protocol. In some cases, due to policy agreements, direct data download from the GEOSS Platform may not be possible and the GEOSS Portal forwards the user's requests to the Provider's data system. Sometimes the Eurac Research data system requires the use of credentials (username and password) to download products. Some products (those coming from OpenEO) can be directly downloaded as Network Common Data Form (NetCDF) files, other resources (those coming from Geonode) can be downloaded in Shapefile or GeoTIFF formats.

4. Discussion and considerations

In this section, we analyze the metadata content information for Eurac Research datasets and the end-user query requests to the GEOSS Platform for retrieving these datasets. The discussed study provides statistics information about the main metadata elements available and the use of Eurac Research datasets considering human interactions (users making query requests to the GEOSS Portal) as well as machine interactions; some organizations, such as the WMO WIS GISC and the Korea Meteorological Administration (KMA), are allowed to make regular harvesting requests with automatic tools to periodically collect the metadata content of the GEOSS Platform, including Eurac Research datasets.

4.1. Metadata content analysis

An analysis considering each shared metadata record has been performed, to evaluate the overall information content provided by Eurac Research data. A set of common metadata elements have been selected to test their presence in the records: (dataset) "provider", "title", "spatial extent", "temporal extent", "responsible party", "abstract", "keywords", "data policy". Table 2 reports the percentages of their occurrences:

Table 2. Metadata elements percentage occurrences.

Metadata element	Occurrence percentage
Provider:	100% ¹ ¹ The dataset <i>Provider</i> metadata element is generated by the DAB, automatically
Title:	100%
Spatial extent:	96.33% of which: <i>Rectangular area</i> 64.22% of all the dataset records <i>Point</i> 32.11% of all the dataset records
Temporal extent:	25.38%
Cited Organization and role	91.74% of the records cite an organization 91.74% of the records indicate a role for the organization <i>role = Originator</i> 59.63% of all the dataset records provide an originator organization
Keywords:	100%
Keyword thesaurus [conditional]	31.80% ⁴ of records have utilized keywords from thesaurus (or controlled vocabulary). The only thesaurus cited is <i>GEMET – INSPIRE themes, version 1.0</i>
Abstract:	100%
Data policy:	31.80%

Many important metadata elements are well present, in particular: (dataset) provider, title, keywords and abstract; other elements can be improved (i.e. cited organization and spatial extent); temporal extent and data policy are largely missing. Use of keywords from thesaurus can also be improved.

Use of controlled vocabularies is in general suggested also for other metadata elements where possible (e.g. organizations, observed parameters, data policies, formats, units), as the resulting records would be more harmonized. For example, the following different spellings can be currently found in the records (amongst others), showing the heterogeneity of citations for a specific organization:

- eurac-research – Institute for Earth Observation (230 occurrences)
- Eurac Research (195 occurrences)
- Eurac Research – Earth Observation Institute (72 occurrences)
- Institute for Earth Observation (26 occurrences)
- Eurac Research – Earth Observation (43 occurrences)
- Institute for Earth Observation (26 occurrences)
- eurac research (1 occurrence)

Table 3 shows the top originator organizations, indicating Eurac Research (often specified with the indication of the actual involved institute) as the main originator. Be aware that

Table 3. Relative percentages of top originator organizations after syntactic harmonization had taken place.

Organization	Percentage
Eurac Research	94.88% of which: <i>No institute specified</i> 39.52% <i>Institute for Earth Observation</i> 25.14% <i>Center for Sensing Solutions</i> 17.96% <i>Institute for Renewable Energy</i> 9.58% <i>Institute for Regional Development</i> 7.78%
Ca' Foscari University of Venice	2.27 %
RMIT University	1.70 %
Regional Environmental Protection Agency of Italy, Veneto region	1.13 %

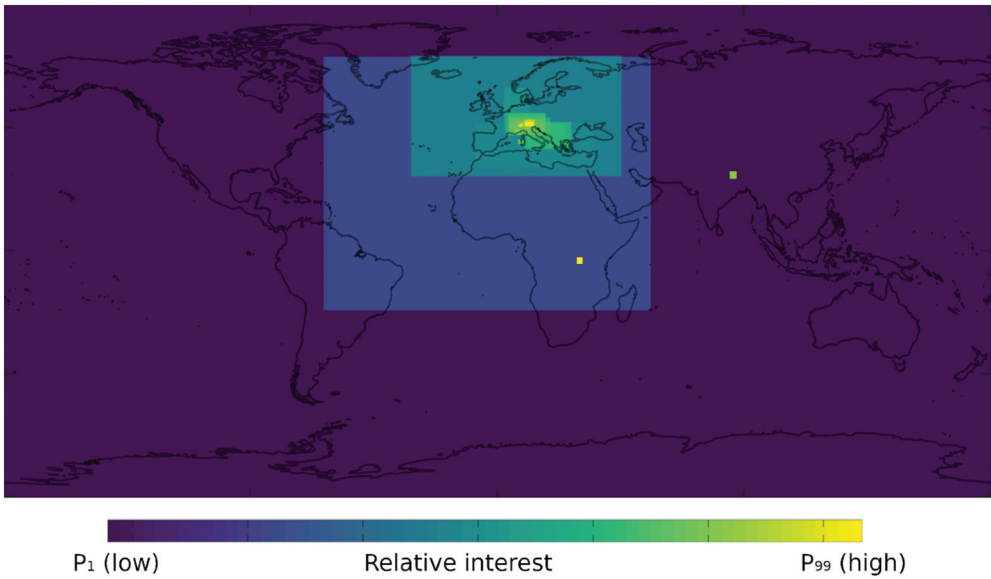


Figure 7. Spatial coverage of Eurac Research datasets. The contribution on a pixel of each dataset is weighted according to the dataset extent. The relative interest is calculated on each pixel as the combined contributions of all the datasets divided by the total. It is finally plotted on a world coastline map according to a logarithmic scale in the range between the 1st and the 99th percentiles of its distribution.

Figure 7 shows the spatial coverage of Eurac Research datasets: Europe, in particular, Tyrol and the Alps are well represented. Burundi and Himalayas are also represented. Figure 8 shows the temporal coverage of the Eurac Research datasets, spanning the period 2007–2023 plus a forecast dataset spanning the period 2041–2070. The bulk of

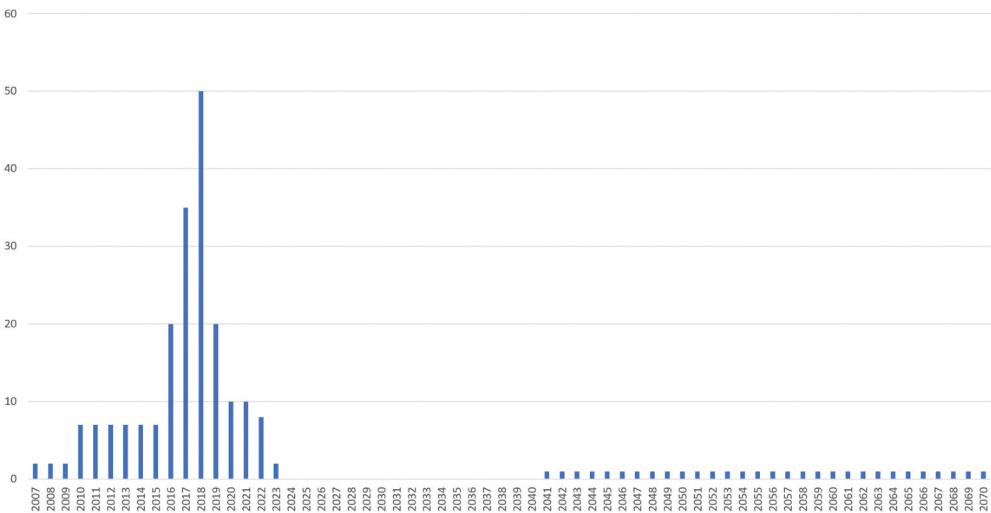


Figure 8. Temporal coverage of Eurac Research datasets. The value for each year is calculated as the count of all the datasets documented to contain data on the given year.

the data is in the period 2010–2022, with the peak reached on year 2018, having 50 datasets covering that specific year. The results of this analysis must be, however, viewed with due caution, as the temporal coverage metadata element was documented only in the 25.38% of the records, as previously noted.

4.2. User requests analysis

All the following indicators hold information over the last 21 months (from 1 January 2021 to 30 September 2022). In this range period, 411 queries made by GEOSS Portal users matched one or more Eurac Research datasets generating an average number of about 20 matching queries per month. [Figure 9](#) shows the most occurring keywords in requests matching the Eurac Research datasets.

It is worth mentioning that a large part of matching requests (about 38%) have no keyword indicated by the users. More specifically:

- in 54.7% of queries, the user only indicated the Eurac Geonetwork catalog as the target catalog for the search;
- in 4.45% of queries, the user only indicated the Eurac SOS catalog as the target catalog for the search;
- in 38.85% of queries, the user indicated both the Eurac Geonetwork and the Eurac SOS catalogs as the target catalog for the search;
- 1.91% queries matched Eurac datasets only because of the bounding box search constraint.

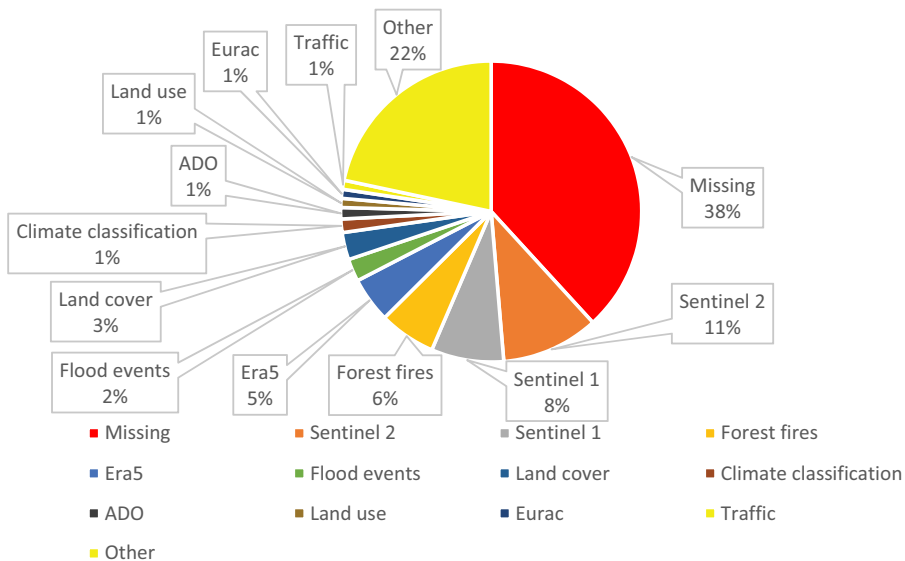


Figure 9. Relative percentages of the most searched keywords to retrieve Eurac Research data.

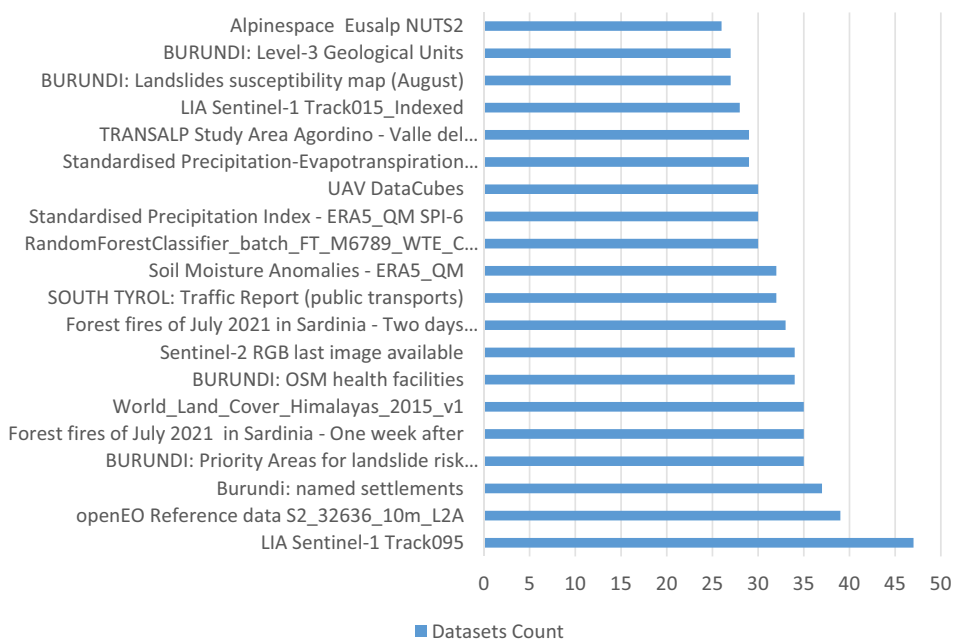


Figure 10. Number of the first 20 popular collections from Eurac Research datasets.

Some of the users seem to be interested in satellite data as proved by the most popular keywords (*Sentinel 2* and *Sentinel 1*) that together represent about the 20% of the matching requests.

Figure 10 reports the most popular datasets retrieved from Eurac Research resources. In accordance with the most popular keywords analysis, the remote sensing resources are the most returned ones (*LIA Sentinel-1 Track095* and *openEO Reference data S2_32636_10m_L2A*).

The following analysis regards the organizations that searched for Eurac Research datasets. Studying the request originators using the “whois” program on each IP addresses associated to the queries, we found out that the main users come from academia and research centers (e.g. University of Padua, National Observatory of Athens, University of Liege). Unfortunately, this specific identification mechanism may not always lead to relevant results, because in many cases, the IP addresses are associated to Internet Service Providers (ISPs), which assign dynamic IP addresses to their users, thus preventing to find out the possible originating organization. Table 5 shows the full list of countries with the greatest number of requests for discovering Eurac Research datasets.

4.3. Considerations

The Eurac Research Data Provider is an important Data Provider of the GEOSS Platform because (a) contains several heterogenous datasets from different platforms such as

Table 5. Countries with the greatest number of requests for discovering Eurac Research datasets.

Country	Number of requests
Italy	214
Switzerland	18
Spain	15
Netherlands	12
Greece	11
Austria	11
United Kingdom of Great Britain and Northern Ireland	11
France	10
Germany	9
Serbia	8
Kenya	8
United States of America	8
Belgium	7
China	7
Australia	7
Portugal	4
Pakistan	4
Finland	3
Denmark	3
Ireland	3
India	3
Poland	3
Romania	3
Norway	2
Nepal	2
Albania	2
Philippines	2
Russian Federation	1
Bulgaria	1
Japan	1
Burundi	1
Luxembourg	1
France	1
Hungary	1
Saudi Arabia	1
Uganda	1
Algeria	1
Israel	1
Republic of Korea	1
Mexico	1
Colombia	1
Ethiopia	1
Peru	1
Lebanon	1
Turkey	1

satellite data as well as in situ observations; (b) follows and implements the FAIR principles; (c) is aligned with the major international directives and guidelines on metadata and data, although some suggestions have been provided in the metadata analysis section to further improve its metadata quality.

In this series of manuscripts describing popular datasets published on the GEOSS Platform, the main focus is to discuss the process to share well-used open datasets through a system of systems like GEOSS providing visibility on the datasets itself and helping to promote open Earth Observation data. During the analysis of these datasets, some lessons learned are here highlighted:

- The importance of interoperability tests to better understand the valuable usage of geospatial open standards and to establish a virtuous feedback cycle with the data provider aimed at improving the publication quality. For example, metadata quality can be improved if data provider receives suggestions in this direction (e.g. assuring the presence of the most important metadata elements, use of thesauri) at the same time, the GEO-DAB can benefit from data providers' feedback.
- The importance of a sustainable technical governance to establish and guarantee the creation of a digital ecosystem such as GEOSS.
- The importance of communication, dissemination, and exploitation to promote the contribution and visibility of GEOSS data providers. For example, the active participation of data providers to GEOSS-related workshops, such as the next Open Data and Open Knowledge Workshop¹³ would be beneficial to re-establish a strong connection and positive feedback with the user community.

Notes

1. <https://earthobservations.org/index.php>.
2. <https://openeo.org/>.
3. <https://geonode.org/>.
4. <https://52north.org/software/software-projects/sos/>.
5. <https://www.geonetwork-opensource.org/>.
6. <https://stacspec.org/en>.
7. <http://rasdaman.com/>.
8. <https://www.w3.org/TR/1999/REC-xslt-19991116>.
9. <https://github.com/ESSI-Lab/DAB>.
10. <https://hydrohub.wmo.int/en/whos>.
11. <http://www.odip.org/>.
12. <https://emodnet.ec.europa.eu/en/emod-pace>.
13. <https://www.earthobservations.org/meetings.php>.

Acknowledgements

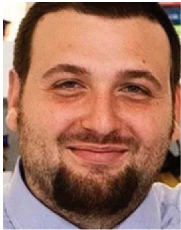
The research leading to these results has received funding from the European Space Agency through the DAB4EDGE (GEO-DAB Support for European Direction in GEOSS Common Infrastructure Enhancements; 2018–2020; ESA Contract No. 4000123005/18/IT/CGD) project and from Horizon 2020 research and innovation program under grant agreement N. 776136 (EDGE – European Direction in GEOSS Common Infrastructure Enhancements) and N. 101039118 (GPP – GEOSS Platform Plus).

The authors would like to thank Fabrizio Papeschi (CNR-IIA) for the GEO DAB component development, Massimiliano Olivieri (CNR-IIA) for the administration of cloud services, Roberto Salzano for useful discussion about metadata analysis techniques and Lena Rettori (CNR-IIA) for language revision.

Disclosure statement

No potential conflict of interest was reported by the author(s).

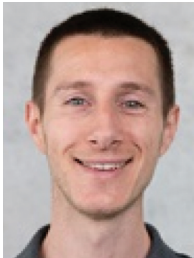
Notes on contributors



Roberto Roncella is a technologist at the Florence Division of the Institute of Atmospheric Pollution Research of the National Research Council of Italy (CNR-IIA). He has a Master degree in Information Technology achieved at the University of Florence, Italy. He joined the Earth and Space Science Informatics Laboratory (ESSI-Lab) of the National Research Council of Italy (CNR) in 2011. His work is focused on the design and development of geo-spatial services for environmental models and their interoperability. He participated in several research projects and initiatives funded by EC (FP7, H2020, CIP). He cooperated to design and develop the GEO Discovery and Access Broker (GEO DAB). He is responsible for the brokering process of Data Providers in the GEOSS Platform and a member of GEOSS Platform Operations Team.



Bartolomeo Ventura received a Ph.D. in Physics from the University of Bari (Italy) in 2008. He is presently working at the EURAC-Institute for Earth Observation (Bolzano, Italy). Within the same Institute, he is responsible for the research line “Scientific Data Management & Processing” with the main goals to provide access and solutions to the processing environment for research, to foster the uptake of new technologies, and bridge to IT infrastructure and service providers. He is also responsible for the management and organization of all the data collected by the Institute. The definition of the best strategy for the creation, organization, and maintenance of the data structure as well as the definition of access policies and the long data preservation represents the main goal for this task. He is also in charge of the creation of well-structured metadata (ISO and OGC standards-compliant) and easy data access.



Andrea Vianello received his BSc in Geographical Information System in 2003 at the University IUAV of Venice and a second BSc in Forestry Science at the University of Padua in 2009. He received his MSc in Geographical Information System and Remote Sensing in 2014 at the University IUAV of Venice. He has been working at ISMAR-CNR in Venice from 2003 to 2016 as spatial data expert with a focus on data sharing on the web. He has experience also in underground water and GPS measurements. From 2016, he is working at Eurac Research as senior technician to manage the Spatial Data Infrastructure and web services.



Enrico Boldrini is a researcher at the Florence Division of the Institute of Atmospheric Pollution Research of the National Research Council of Italy (CNR-IIA). He is collaborating to the activities of the Earth and Space Science Informatics Laboratory (ESSI-Lab) since 2007, participating to several projects by different bodies (e.g. WMO and GEO intergovernmental organizations, European Union H2020/FP7/EASME programs, US NSF, Italy MIUR/ARPA-ER). His work focuses on enabling information sharing and systems interoperability in the multidisciplinary context of Earth Sciences. He coordinates the development and management of the Discovery and Access Broker framework (DAB), which is a key technology in the architecture of both the Global Earth Observation System of Systems (GEOSS) and the WMO Hydrological Observing System (WHOS).



Mattia Santoro is a researcher at the Division of Florence of CNR-IIA. He has a Ph.D. in Methods and Technologies for Environmental Monitoring at the University of Basilicata, Italy. He obtained a degree in computer science at the University of Florence, Italy. He has been working at the Earth and Space Science Informatics Laboratory (ESSI-lab) of the National Research Council of Italy (CNR) since 2007. His research interests deal with multidisciplinary interoperability, designing, and developing infrastructures and services for geo-spatial resources, with particular focus on semantic discovery and environmental models interoperability. He is responsible for the Geospatial Artificial Intelligence and Information Sharing (GAINS) Working Group of the CNR-IIA. He participated in several research projects and initiatives funded by EC (FP7, H2020), US NSF, and National R&I frameworks. He coordinates the design and development of the VLab framework. He is responsible for the GEO Discovery and Access Broker (GEO DAB) operational environment, member of GEOSS Common Infrastructure (GCI) Operations Foundational Task, GEOSS Platform Operations Team and of the GEOSS Infrastructure Development Task Team.



Paolo Mazzetti is Head of the Division of Florence of CNR-IIA. He holds a degree in Electronic Engineering and he taught “Telematics” at the University of Florence in Prato for the degree in Information Engineering for 7 years. He has more than 15 years of experience in the design and development of infrastructures and services for geo-spatial data sharing in the context of national, European (FP7, CIP, H2020), and global initiatives. He is Principal Representative of Italy in the GEO Programme Board. He is the Coordinator of the GEO DAB activities in the GEOSS Common Infrastructure (GCI) Operations Foundational Task, and member of the GEOSS Infrastructure Development Task Team. He has been a member of the GEO Secretariat Expert Advisory Group (EAG) and of the GEO Institutions Development Implementation Board (IDIB). He is a member of the EuroGEO Coordination Group representing Italy. He is a member of the National Working Group on Land Degradation Neutrality. He is the project coordinator of the NewLife4Drylands LIFE Preparatory Project.



Stefano Nativi co-chairs the “GEOSS Development Task” team and the “Data Ethics” team for GEO. He is also the vice-chair of the International Advisory Board of CBAS (International Research Center of Big Data for Sustainable Development Goals). He is a member of the ISDE (International Society of Digital Earth) council. He funded and chaired the Earth and Space Sciences Informatics Division of the European Geosciences Union. He is an Editor-in-Chief of the *Big Earth Data* journal and co-editor of the “AI section” of the *Remote Sensing* journal. He was the Big Data Lead Scientist of the JRC of the European Commission. He received the EGU Ian McHarg medal, the Geospatial Innovation Award of GWF, and the Meritorious Service Medal of the ICEO.

ORCID

Roberto Roncella  <http://orcid.org/0000-0002-6825-8345>

Bartolomeo Ventura  <http://orcid.org/0000-0002-8898-4205>

Andrea Vianello  <http://orcid.org/0000-0002-3017-6045>

Enrico Boldrini  <http://orcid.org/0000-0003-2075-4742>

Mattia Santoro  <http://orcid.org/0000-0003-0401-3100>

Paolo Mazzetti  <http://orcid.org/0000-0002-8291-1128>

Stefano Nativi  <http://orcid.org/0000-0003-3185-8539>

Data availability statement

The data that support the findings of this study are available in the GEOSS Platform at <https://www.geoportal.org/>.

References

- Boldrini, E., Hradec, J., Craglia, M., & Nativi, S. (2021). *GEOSS content exploration*. European Commission.
- Boldrini, E., Nativi, S., Pecora, S., Chernov, I., & Mazzetti, P. (2022). Multi-scale hydrological system-of-systems realized through WHOS: The brokering framework. *International Journal of Digital Earth*, 15(1), 1259–1289. <https://doi.org/10.1080/17538947.2022.2099591>
- Craglia, M., Hradec, J., Nativi, S., & Santoro, M. (2017). Exploring the depths of the global earth observation system of systems. *Big Earth Data*, 1(1–2), 21–46. <https://doi.org/10.1080/20964471.2017.1401284>
- Craglia, M., Nativi, S., Santoro, M., Vaccari, L., & Fugazza, C. (2011). Inter-disciplinary interoperability for global sustainability research. *Fourth International Conference on Geospatial Semantics* (pp. 1–15). May 12–13 Brest France.
- GEO. (2022a, June 21–22). *23rd Programme Board Meeting*. https://www.earthobservations.org/documents/pb/me_202206/PB-2311_Revised%20GEO%20Data%20Management%20Principles%20Implementation%20Guidelines.pdf
- GEO. (2022b, Mar). *About us*. Tratto da Group on Earth Observations: <https://earthobservations.org/index.php>
- GEO. (2015, November 11–12). *GEO Strategic Plan 2016–2025: Implementing GEOSS*. Tratto da earth-observations.org: https://www.earthobservations.org/documents/open_eo_data/GEO_Strategic_Plan_2016_2025_Implementing_GEOSS_Reference_Document.pdf
- GEOSS Infrastructure Development Task Team. (2017). *The GEOSS Platform: All you need to know to become a GEO data provider*. GEO.
- Guo, H., Nativi, S., Liang, D., Craglia, M., Wang, L., Schade, S.,... Corban, C., He, G., Pesaresi, M., Li, J., Shirazi, Z., Liu, J., & Annoni, A. (2020). Big Earth Data science: An information framework for a sustainable planet. *International Journal of Digital Earth*, 13(7), 743–767. <https://doi.org/10.1080/17538947.2020.1743785>
- Nativi, S., & Craglia, M. (2021). Digital twins of the earth. In Daya Sagar, B. S., Cheng, Q., McKinley, J., & Agterberg, F. (Eds.), *Encyclopedia of mathematical geosciences. Encyclopedia of earth sciences series* (pp. 1–4). Springer, Cham. https://doi.org/10.1007/978-3-030-26050-7_457-1
- Nativi, S., Craglia, M., & Pearlman, J. (2013). Earth science infrastructures interoperability: The brokering approach. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, (Vol. 6, pp. 1118–1129). Munich, Germany. <https://doi.org/10.1109/JSTARS.2013.2243113>
- Nativi, S., & Mazzetti, P. (2021). Geosciences digital ecosystems. In Daya Sagar, B.S., Cheng, Q., McKinley, J., & Agterberg, F. (Eds.), *Encyclopedia of mathematical geosciences. Encyclopedia of earth sciences series* (pp. 1–6). Springer, Cham. https://doi.org/10.1007/978-3-030-26050-7_458-1
- Nativi, S., Mazzetti, P., & Craglia, M. (2021). Digital ecosystems for developing digital twins of the earth: The destination earth case. *Remote Sensing*, 13(11), 2119–2144. <https://doi.org/10.3390/rs13112119>
- Nativi, S., Mazzetti, P., Santoro, M., Papeschi, F., Craglia, M., & Ochiai, O. (2015). Big data challenges in building the global earth observation system of systems. *Environmental Modelling & Software*, 68, 1–26. <https://doi.org/10.1016/j.envsoft.2015.01.017>
- Ollier, G. (2019, November). *GEO XVI plenary: session 5: Broadening the impact of earth observation and GEO - Eurogeo*. Tratto da. <https://youtu.be/W-JQc3rjC7g?t=3017>
- Roncella, R., Boldrini, E., Santoro, P., Mazzetti, M., Andrade, J., Catarino, N., & Nativi, S. (2022). Publishing NextGEOSS data on the GEOSS Platform. *Big Earth Data*, 1–15. <https://doi.org/10.1080/20964471.2022.2135234>
- Roncella, R., Zhang, L., Boldrini, E., Santoro, M., Mazzetti, P., & Nativi, S. (2022, May). Publishing China satellite data on the GEOSS Platform. *Big Earth Data*, 1–15. <https://doi.org/10.1080/20964471.2022.2107420>

- Santoro, M., Mazzetti, P., & Nativi, S. (2020). The VLab framework: An orchestrator component to support data to knowledge transition. *Remote Sensing*, 12(11), 1795. <https://doi.org/10.3390/rs12111795>
- Santoro, M., Nativi, S., & Mazzetti, P. (2016). Contributing to the GEO model web implementation: A brokering service for business processes. *Environmental Modelling & Software*, 84, 1364–8152. <https://doi.org/10.1016/j.envsoft.2016.06.010>
- Ventura, B., Vianello, A., Frisinghelli, D., Rossi, M., Monsorno, R., & Costa, A. (2019). A methodology for heterogeneous sensor data organization and near real-time data sharing by adopting OGC SWE standards. *ISPRS International Journal of Geo-Information*, 8(4), 167. <https://doi.org/10.3390/ijgi8040167>
- Wilkinson, M., Dumontier, M., Aalbersberg, I., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. -W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R. 2... Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3(1). <https://doi.org/10.1038/sdata.2016.18>