



## GEOSS Platform data content and use

Enrico Boldrini <sup>a</sup>, Stefano Nativi <sup>a</sup>, Jiri Hradec <sup>b\*</sup>, Mattia Santoro <sup>a</sup>, Paolo Mazzetti <sup>a</sup> and Max Craglia <sup>b\*</sup>

<sup>a</sup>National Research Council of Italy, Institute of Atmospheric Pollution Research, Florence, Italy; <sup>b</sup>Joint Research Centre, B6 Unit, Ispra, Italy

### ABSTRACT

The GEOSS Platform is a key contribution to the goal of building the Global Earth Observation System of Systems (GEOSS). It enables a harmonized discovery and access of Earth observation data, shared online by heterogeneous organizations worldwide. This work analyzes both what is made available in the GEOSS Platform by the data providers and how users are utilizing it including multiyear trends, updating a previous analysis published in 2017. The present statistics derive from a 2021 EOValue report funded by the European Commission. The offer of GEOSS Platform data has been the object of various analyses, including data provider characterization, data sharing trends, and data characterization (comprising metadata quality analysis, thematic analysis, responsible party identification, spatial–temporal coverage). GEOSS data demand has also been the object of several analyses, including data consumer characterization, utilization trends, and requested data characterization (comprising thematic analysis, spatial–temporal coverage, and popularity). Among the findings, a large amount of shared data, mostly from satellite sources, emerges with an issue of low metadata quality and related discovery match. Moreover, the trend in usage is decreasing. Therefore, the progressive disconnection of the GEOSS platform from its data Providers and Users and other possible causes are also reported.

### ARTICLE HISTORY

Received 12 September 2022  
Accepted 25 January 2023

### KEYWORDS

Earth observations; GEOSS; System-of-Systems; geospatial data quality; discovery service

## 1. Introduction

In 2003, governments and international organizations committed to a vision of a future wherein decisions and actions for the benefit of humankind are informed by coordinated, comprehensive and sustained Earth observations. In 2005, the first concrete step toward achieving that vision was taken with the establishment of the Group on Earth Observations (GEO). Presently, this inter-governmental initiative comprises more than 100 national governments and more than 100 Participating Organizations (Group on Earth Observation 2022a). A central part of GEO's Mission is to build the Global Earth Observation System of Systems (GEOSS) (GEO 2009). GEOSS is defined as 'a set of coordinated, independent Earth observation, information and processing systems that interact and provide access to diverse information for a broad range of users in both public and private sectors' (GEO 2015b). For about 15 years, GEOSS was conceived and implemented as a global hub

**CONTACT** Enrico Boldrini [enrico.boldrini@cnr.it](mailto:enrico.boldrini@cnr.it) Area di Ricerca di Firenze, Via Madonna del Piano 10, 50019 Sesto Fiorentino, FI, Italy

\*The views expressed are purely those of the authors and may not in any circumstances be regarded as stating an official position of the European Commission.

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

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.

for Earth Observation (EO) data and information (GEO 2009; GEO 2015b; Group on Earth Observation 2015c). The GEOSS platform is a key contribution to GEOSS cause it proves an overarching infrastructure with the main task of easing discovery and access to the many datasets (Boldrini et al. 2021) made available by a plethora of diverse national and international organizations (GEOSS Infrastructure Development Task Team 2017; Boldrini et al. 2021).

In 2017, the Joint Research Centre (JRC) of the European Commission accomplished a first comprehensive study of GEOSS data content and its users (Craglia et al. 2017). The goal of the present work is twofold: to consistently update the JRC analysis, and to understand the main changes occurred in the last few years to GEOSS and its Community. The full analysis and the complete statistical results are included in a dedicated report managed by the JRC and funded by the European Commission DG RTD (Boldrini et al. 2021).

### 1.1. Main GEOSS Platform components, stakeholders, and meta-information

The GEOSS framework is substantially made up of the same components that were present at the time of the last analysis, except for an important addition, that is the GEOSS Knowledge Hub (GKH) initiative (GEO 2022b). GKH aims at being an open-source digital repository of open, authoritative, and reproducible knowledge created by GEO. Its connection to the existing GEOSS platform components is however still under discussion, being GKH in the prototype stage.

This analysis focuses on the GEOSS platform data content and use. The GEOSS platform implements a geoscience digital ecosystem (Nativi and Mazzetti 2021; Nativi, Mazzetti, and Craglia 2021; Nativi and Craglia 2021; Guo et al. 2020), which is realized as a distributed System-of-Systems also by applying a brokering paradigm (Nativi, Craglia, and Pearlman 2013).

The main components and actors of the GEOSS Platform are (Nativi et al. 2015):

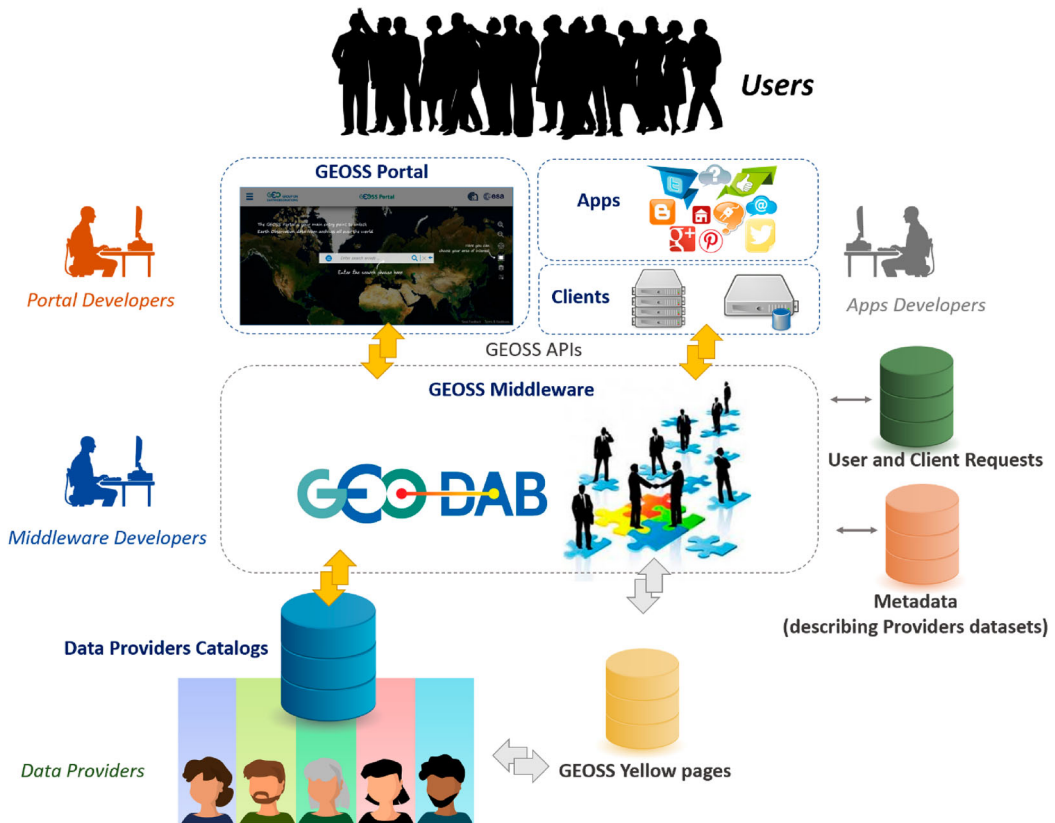
- **GEOSS (data) Providers and Catalog services:** organizations responsible for sharing metadata records on the GEOSS platform through their catalog services
- **GEOSS (data) End-users and the GEOSS Portal:** organizations and individuals that search and collect GEOSS metadata records through the GEOSS Portal or other tools.
- **GEOSS Applications, Developers, and GEOSS APIs:** organizations responsible for implementing end-user tools interacting with the GEOSS Platform.
- **GEOSS Yellow Pages:** a registry component filled by data provider organizations with information about the catalog services for which they are responsible
- **GEO Discovery and Access Broker (GEO-DAB):** the middleware component that connects the GEOSS data Providers to the GEOSS End user tools and Client applications. The middleware implements metadata mediation, harmonization, and profiling services. GEO-DAB periodically harvests the data Providers catalogues registered in the GEOSS Yellow Pages component. During the harvesting procedure the DAB stores their metadata content in the Metadata database. In addition, the GEO-DAB stores and manages information about the received User and Client requests – specifically, in the User and Client Requests database.

The GEOSS Platform's components and their main interactions are depicted in [Figure 1](#).

### 1.2. Data provider catalog service types

It is possible to recognize three main groups of data services, depending on the functionalities, implemented by data Provider, and their related interactions with the GEOSS Platform middleware – i.e. the GEO-DAB. The three types of services published by the GEOSS Providers are:

- **Harvestable services:** they allow the GEO-DAB (i.e. its standard or customized *harvester* components) to periodically collect the entire set of metadata records published by the data



**Figure 1.** GEOSS platform main actors and components.

Providers. These records are stored into the GEO-DAB Metadata database to enable fast user queries.

- **Searchable services:** they do not permit harvesting but allow on-the-fly searches for finalizing data discovery requests. Therefore, once a user/client request is received, the GEO-DAB must forward the query to these data Provider catalogs to get matching results.
- **Combined services:** data Providers allow harvesting of only a subset of metadata records by the GEO-DAB, while the others are searchable at user query time. Commonly, this is the case data collections are harvestable, while collection data components (i.e. Granules) are only searchable on-the-fly (see the case of harvestable satellite data collections and searchable scenes).

## 2. Analysis methodology and limitations

### 2.1. Data pool and the metadata information hyper-cube

For the analysis of GEOSS data content and use, the utilized data population consists of two record types: **shared data descriptions** and **client/user discovery requests**. They stem from the 'Metadata' and the 'User and Client Requests' databases, which are managed by the GEO-DAB. Both information records are described by a set of metadata elements, also known as *properties*. The descriptions of data shared through the GEOSS Platform are contributed by the GEOSS data Providers. The discovery requests' metadata are instead generated by GEOSS users: either individuals (i.e. human users interacting with the GEOSS Portal) or machines (i.e. software clients via the GEOSS APIs).

As for the analysis process, the metadata elements were organized to generate a multidimensional cube: this is an information hyper-cube, whose subject is the information type and whose dimensions are the metadata elements (or the hyper-cube's features). The cube is an information construct that is ready to be further explored by means of analytical tools and addresses analysis demands.

## 2.2. The Dataset information type

The fundamental type of information is the *Dataset* object (i.e. resource record), which is shared on the GEOSS Platform by the GEOSS data Providers. These objects were utilized for the analysis of the GEOSS Platform content. According to the ISO standard technical committee (ISO 2003) on geomatics, *Dataset* can be defined as an 'identifiable collection of data' representing an observation of the Earth (e.g. in-situ or remote observation), or a model output (e.g. a simulation or forecast). As depicted in Figure 2, the GEOSS dataset resource type is characterized by a content and a metadata record:

- **Content**, containing the actual collected data, that is to say the values of the physical observation (or digital simulation) encoded using a different approach, depending on the data Provider interoperability policy (e.g. binary, XML, CSV, etc.).
- **Metadata**, a description of the dataset, composed of different metadata elements (i.e. distinct units of metadata, each documenting a specific aspect of the dataset). Presently, the GEOSS Platform provides different metadata elements (in keeping with the ISO 19115 model); for the purpose of this study, only a subset of significant elements was analyzed. The examination focused on the following metadata elements:
  - [Mandatory] **Dataset provider** identifies the GEOSS dataset Provider – i.e. the organization held responsible for dataset sharing in GEOSS.
  - [Optional] **Dataset cited organization** identifies an organization contributing to the dataset, under distinct roles (i.e. originator, publisher, contributor).
  - [Optional] **Dataset spatial extent** identifies the geographic area covered by the dataset 'content' (i.e. the spatial geometry defining the data geographic position and shape).
  - [Optional] **Dataset temporal extent** identifies the temporal period covered by the dataset 'content' (i.e. the temporal geometry defining the data temporal occurrence and duration).
  - [Optional] **Dataset keyword**: identifies the commonly used word(s), formalized word(s), or phrase(s) used to describe the dataset.

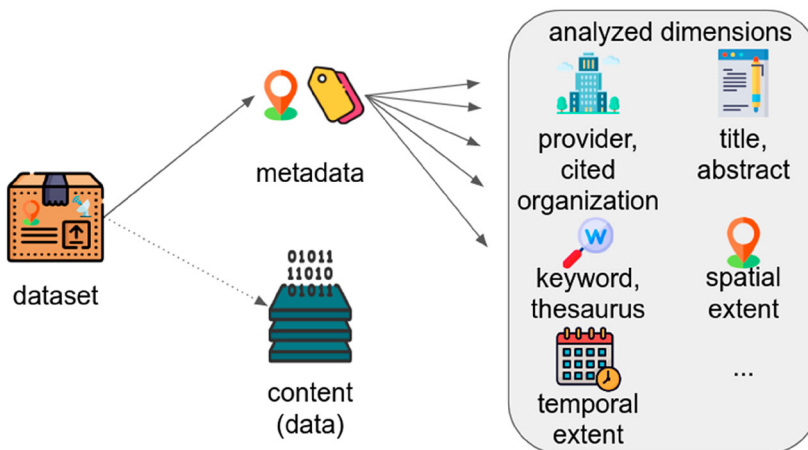


Figure 2. GEOSS dataset main components and choice of analyzed elements.

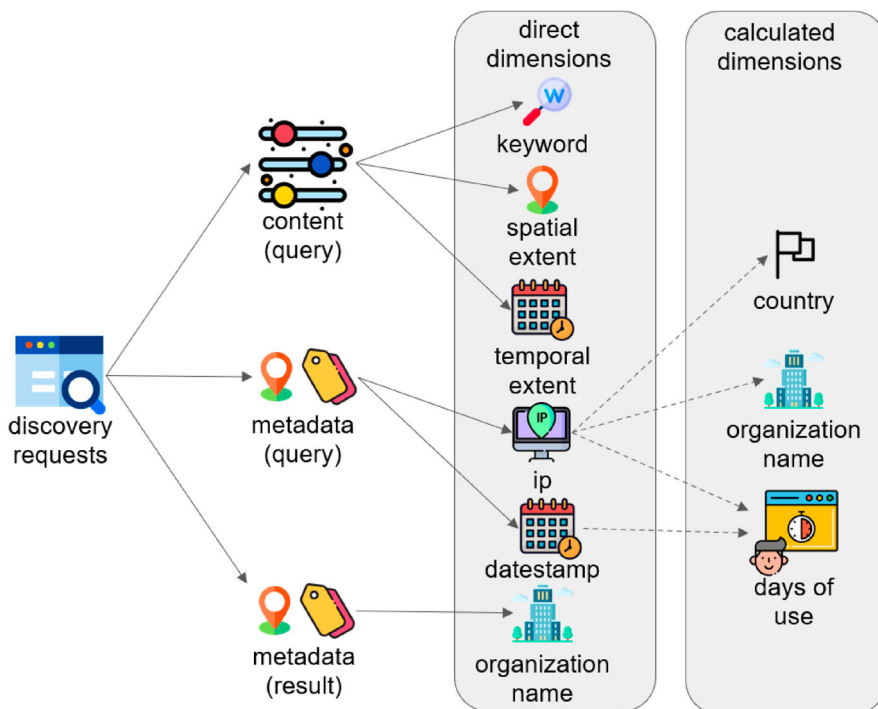
- [Optional] **Dataset keyword thesaurus** identifies the source of utilized keywords.
- [Optional] **Dataset title**: defines the name by which the dataset is known.
- [Optional] **Dataset abstract**: defines a brief narrative summary of the dataset content.

This study focused on the metadata analysis, with the exception of content examination. In keeping with the GEOSS data policy, all the *Dataset* metadata elements are optional, except for the dataset *provider* element, which is automatically generated by the GEO-DAB (GEO 2015a).

### 2.3. The Discovery request information type

The other fundamental information type is the *Discovery request* (i.e. resource record), which was presently utilized for analyzing the GEOSS Platform utilization. *Discovery request* records are generated by the GEOSS users and/or clients (i.e. Machine-to-Machine interactions). These requests aim to search GEOSS for the *Dataset* objects that match a set of well-specified search clauses – i.e. the query constraints. As depicted in Figure 3, a stored GEOSS *Discovery request* is characterized by three query components:

- **Query filters**, constituted by a set of query constraints (i.e. a list of *Dataset* metadata elements and their respective occurrence values).
- **Query metadata**, a set of metadata elements that describe the query itself (e.g. the query time-stamp and its originator IP address).
- **Query results**, a set of metadata elements that describe the query results (e.g. the data provider's name and the title characterizing each matching *Dataset*).



**Figure 3.** Components of a GEOSS *Discovery request* and analyzed metadata elements.

The study considered all three components characterizing a stored *Discovery request*. The content investigation dealt with the following query clauses (i.e. *Dataset* metadata elements):

- **(Dataset) spatial extent:** geographic area (bounding box) specified in the query to restrict the result set to those datasets georeferenced within it
- **(Dataset) temporal extent:** temporal period specified in the query to restrict the result set to those datasets having temporal extent within it
- **(Dataset) keyword:** search word or phrase specified in the query to restrict the result set to those datasets having a matching keyword

The query metadata analysis, on the other hand, focused on the following elements describing a query:

- **Query date stamp:** the query timestamp indicating the precise time instant of query submission
- **Query originator IP address:** numerical label identifying the client originator of the query. This is extracted from the X-Forwarded-for HTTP header. Unfortunately, this important metadata element is available in the request logs only for requests executed after 2018.
- **Query originator country** (calculated): the country of the client originator of the query. This element has been calculated from the IP address through the ‘whos’ Linux utility.
- **Query originator organization name** (calculated): the organization of the client originator of the query. This element has been calculated from the IP address through the ‘whos’ Linux utility.
- **Days of use or number of daily visits** (calculated): this metadata element is calculated starting from the originating IP address and the query date stamps and represents the number of days when the same originator client was active. As an example, if a user posts five requests to the GEOSS portal during day one and one request during day two, two days of use are counted for this single user.
- **Query relative time:** the difference between the user’s request time of submission and the temporal extent requested by the user (e.g. ‘yesterday’, ‘the last five years’, etc.)

Finally, the analysis of the query result metadata considered one element, only:

- **Query result organization name:** the organization(s) providing the matching records that are results of the query.

This limitation leaves space for a future and deeper study on the GEOSS Platform utilization. For example, other result metadata elements could be considered (such as keywords, title, spatial–temporal extent), in order to assess the quality of the match between user searches and obtained results.

## 2.4. Analysis limitations

The authors are aware and acknowledge the presence of the following analysis limitations:

- **Data pool limitation:** the discussed analysis only considered the harvestable resources – i.e. services registered to the data Provider Catalog. As a result, the following two searchable resources were not considered in this study:
  - ArcGIS Online ESRI;
  - USGS Earthquake.

For the combined resources registered to the GEOSS Catalog, this study was restricted only to their harvestable part – e.g. on a satellite collection level.

This study recognized the following combined services:

- GBIF (Free and Open Access to Biodiversity Data);
- NextGEOSS project;

- Ocean Biogeographic Information System;
- Federated EO Gateway (FedEO) – CEOS;
- CEOS WGISS Integrated Catalog (CWIC);
- CEOS International Directory Network (IDN).
- **Metadata quality shortcomings:** the *Dataset* discovery requests logged before 2018 did not include the originating IP address. This issue triggered a limited full comparison with the subsequent discovery requests. Moreover, some of the major data Providers did not provide some of common metadata elements that are generally used to describe *Datasets*. This created a limit towards a harmonized characterization of GEOSS content.

### 3. Analysis results: data offer

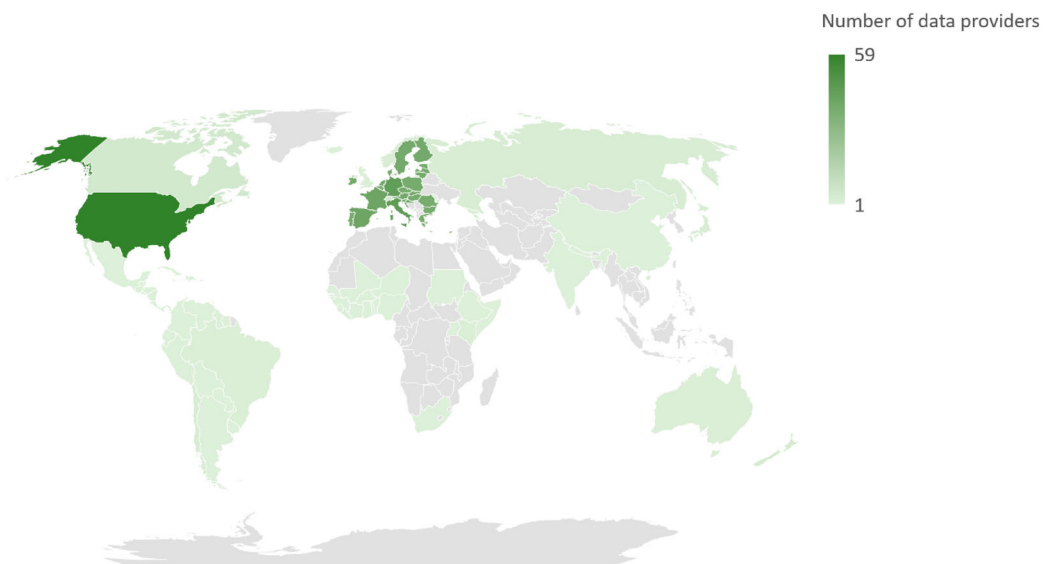
The first part of the analysis focuses on the data provided to GEOSS. The following three sections describe who are the data providers, the amount of data they shared in the last few years, and finally a characterization of the shared metadata and its quality.

#### 3.1. Who are the GEOSS data providers?

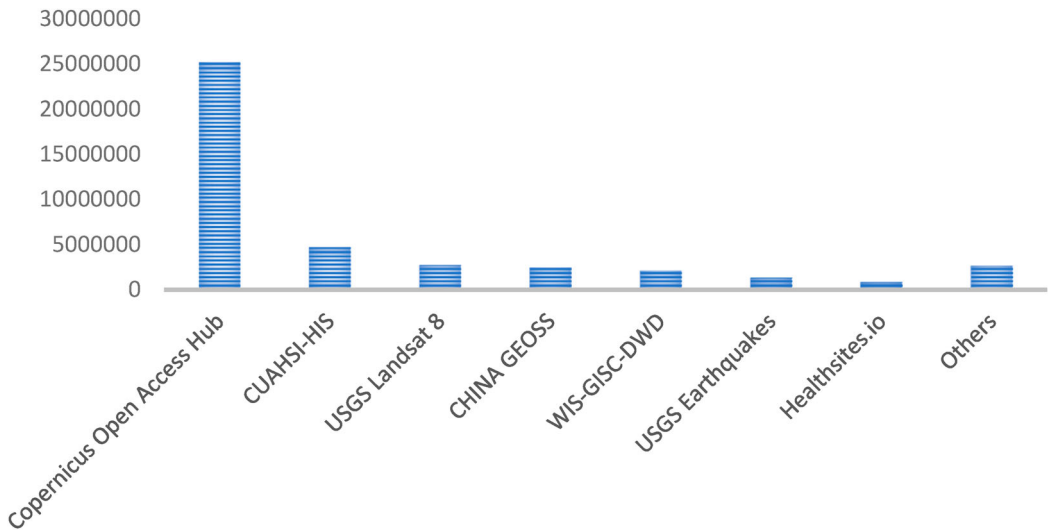
GEOSS data Providers oversee the publication and online sharing of datasets, for example through catalog, registry, proxy, and listing services. They can be different from the originators of dataset content: a valuable instance is represented by PANGAEA, a data provider publishing data from various originators. Data providers are the organizations directly targeted by the DAB, by establishing a connection with each of them through their data publishing web services. According to this study, **the DAB recognizes 196 different GEOSS data providers.**

These organizations mostly pertain to the public, government, and academia sectors. However, also private organizations are present, to a lesser extent.

The map in [Figure 4](#) visualizes the number of data provider organizations for each country. Darker countries contribute with a higher number of data providers, with a maximum of 59 organizations from the United States of America; European countries are also well represented (the



**Figure 4.** Data provider originating countries – darker countries contribute with a higher number of data providers. Contributions of worldwide organizations such as UNEP have been excluded from the map to improve readability.



**Figure 5.** GEOSS platform metadata records by provider.

contribution of worldwide organizations such as UNEP has been filtered out of the map, as it would have had an indistinct impact on all UN countries).

The pie chart in [Figure 5](#) shows the ten data *providers* that contribute with the most records to the DAB metadata database.

Notably, the Copernicus Open Access Hub publishes about the 60% of the datasets shared through the GEOSS Platform. Four of the top ten providers share (also) in-situ datasets – i.e. CUAHSI-HIS, USGS Earthquakes, WIS-GISC-DWD, Healthsites.io. Considering the major provider percentages and their expected typology of dataset contributions, it is possible to estimate that GEOSS Platform in-situ data are between 20% and 30% of the total.

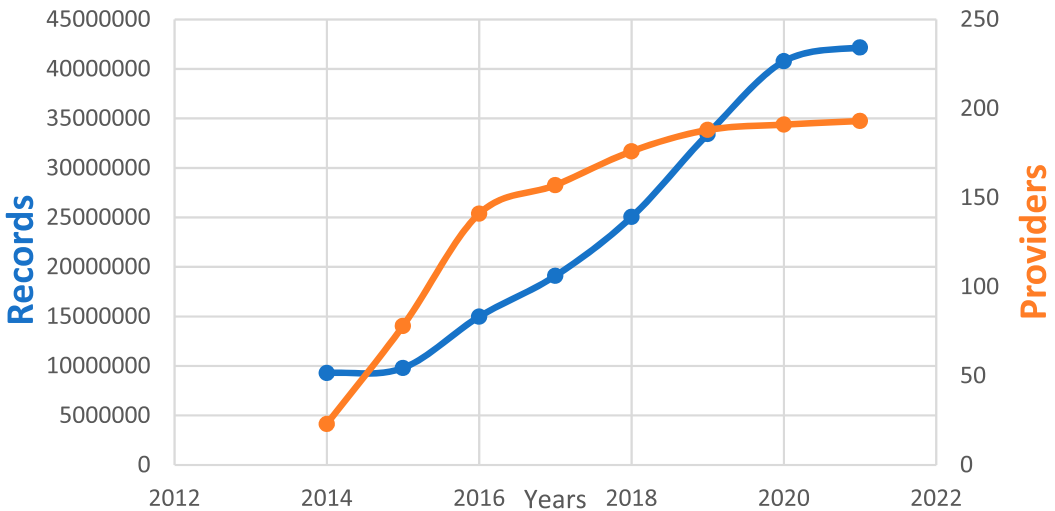
Some remarkable percentage differences can be explained because of the diverse granularity level characterizing the dataset content being shared. For example, the considered satellite sources (e.g. Copernicus Open Access Hub, USGS Landsat 8, and China GEOSS) share records at the granularity level of a single observed scene.

### 3.2. Dataset sharing trends

**More than 42 million records** are present in the GEOSS platform database. [Figure 6](#) shows the progressive increase of the number of harvested metadata records (in blue) along with the progressive increase in the number of their data providers (in orange). The number of retrieved records is remarkably greater than the one counted in the 2017, which was equal to 1.8 million. The main reasons for this significant increase can be explained as follows:

- A physiological increment of the datasets managed by the sources recognized in 2017. In particular, the EU Copernicus Open Access Hub has substantially expanded the number of shared datasets.
- A significant increase of data providers (an increment of about 30%) who share their datasets through harvestable data sources.

However, as represented in [Figure 6](#), the annual increase of data providers has been slowing its pace each year since 2018, because most data providers already take part in GEOSS and we are witnessing a decrease in interest. The overall great increase of data records is to be attributed to



**Figure 6.** Progressive increase of the number of harvested metadata records from about ten million in 2014 to more than 42 million in 2021, compared to the progressive increase of the number of data providers from about 25 in 2014 to about 200 in 2021.

the big data production pace of the lately added data providers, in particular remote sensing data providers.

### 3.3. What data is being shared?

#### 3.3.1. Low metadata quality issue

For each metadata record, the analyzed elements are: (dataset) ‘provider’, ‘title’, ‘spatial extent’, ‘temporal extent’, ‘originator’, ‘abstract’, ‘keywords’. Details of their occurrences can be found in the EOValue GEOSS content exploration report (Boldrini et al. 2021).

Some metadata elements are well present, in particular: (dataset) *source*, *title*, *spatial* and *temporal extent*, and *cited organization*; while other elements (or sub-elements) are largely missing (i.e. *originator* and *keywords*); finally, the *abstract* element is commonly missing.

The very low percentage occurrences of three key elements (*originator*, *keyword* and *abstract*) represent a major gap in GEOSS metadata quality and discovery match. This shortcoming should be addressed by GEO in order to understand, by engaging with the data Providers, the reasons for this drawback and act accordingly on a case-by-case basis. In order to improve metadata completion, some actions can be suggested to support the GEOSS data Providers, including the following: providing feedbacks, extracting the missing content automatically from the other present metadata elements, and/or checking on a possible interoperability issue between the data Provider catalog and the GEOSS Platform.

If GEO were able to help the major data sources (e.g. EU Copernicus Open Access Hub, CUAHSI HIS, USGS Landsat 8, CHINA GEOSS, USGS Earthquakes, Healthsites.io) in fixing their present metadata shortcomings, the occurrence of all the metadata elements would reach more than 90%.

However, it is also important not only to check the presence of metadata values, but also their semantic effectiveness. In fact, in some cases the values published by the data providers act only as placeholders, without conveying useful information about the dataset.

A more general approach to improve the metadata quality (hence, to query matching effectiveness) might also be to set a policy for the analyzed elements to be mandatory and not optional – see for example the adoption of the GEOSS Data management (GEO 2018) and FAIR





Significantly, about the 80% of found keywords were reported to follow a thesaurus or a controlled vocabulary. The use of such tools is key to enable semantic interoperability and query results effectiveness, particularly in multidisciplinary contexts such as GEOSS (Craglia et al. 2011).

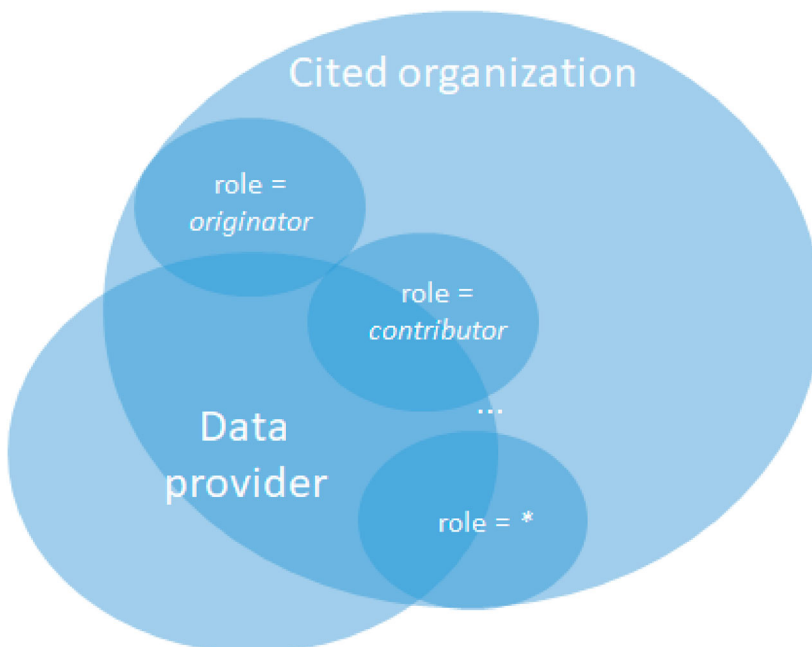
The most used thesauri (or controlled vocabularies) are: **CUAHSI Controlled Vocabularies** (i.e. ‘Data Type CV’, ‘Value Type CV’, ‘Quality Control Levels CV’), **PANGAEA Project List**, **GEMET – EU INSPIRE themes**, **WMO codes** (i.e. ‘CategoryCode’, ‘DistributionScopeCode’, ‘Radiosondes’), **Synoptic hours conventions**, **UN Sustainable Development Goals Taxonomy**, **Mapbender controlled vocabulary**, **Global Change Master Directory (GCMD)** (i.e. ‘Science Keywords’, ‘Data Center Keywords’, ‘Instrument Keywords’), **Australian and New Zealand Standard Research Classification (ANZSRC)**, **SeaDataNet directories (EDMO, EDMED, EDMERP, CSR and EDIOS)**.

### 3.3.3. The community of the GEOSS data providers

**3.3.3.1. What about the GEOSS Platform data community?** GEOSS data community is composed by all the organizations contributing to dataset sharing through the GEOSS Platform. These organizations can play several roles, including data observation, metadata definition, dataset online publication, etc.

There are two metadata elements that refer to these organizations, as depicted in [Figure 9](#):

- *data Providers*, as already introduced;
- *cited Organizations*, distinguished according to their role played in dataset sharing process. In the context of this analysis, the *data Originator* role was of particular interest. Therefore, cited Organizations were divided into the following roles:
  - data Originator (the organization who created the dataset)
  - any other role



**Figure 9.** GEOSS platform data community.

**3.3.3.2. Who are the cited organizations?.** Different organizations may play a role for the implementation of the dataset sharing workflow – e.g. to acquire, describe, contribute, and publish. To calculate the size of this Community, the content of all the metadata elements citing one or more organizations (regardless of their role) was analyzed. The analysis recognized about 18 thousand unique organizations (exactly 18,738 ones).

At least one organization is generally documented for each GEOSS dataset. This helps understanding who the GEOSS dataset stakeholders are: satellite data Providers (i.e. European Commission, China GEOSS, USGS, and NASA GES DISC) are the top cited organizations.

Nevertheless, about 10% of all datasets does not contain any reference to an organization. This issue should be fixed by GEO as one of the necessary actions to improve the GEOSS platform effectiveness.

The following section will investigate the extension of those cited organizations that play the significant role of data *originator*.

**3.3.3.3. Who are the data originators?.** The total number of recognized different *originator* organizations is 13,895. However, some of them are different ways to name the same organization (e.g. ‘NASA/JPL’, ‘NASA-JPL’, ‘NASA JPL Jet Propulsion Laboratory’, ‘NASAJPL (NASA)’, ‘NASA/JPL > JPL National Space Administration’). Once cleaned, the total number of unique *originator* organizations is 7906 – naturally, the same issue and cleansing task was also performed for the cited organization analysis.

The *originators* analysis shows that the actual GEOSS community providing datasets is by far larger (i.e. one order of magnitude) than the number of sources identified by the data *Provider* metadata element – i.e. 196 data *Provider* values.

As reported in Boldrini et al. (2021), the *originator* element is often missing (65.62% of all the datasets), somehow limiting the insights provided by the current analysis and possibly indicating that the GEOSS community providing data is larger. Remote sensing data originators are underrepresented. European Commission appears to be the major originator, due to the data shared by the Copernicus Open Access Hub.

To address the issue of utilizing non-unique organization names, it is suggested the use of organization yellow pages (e.g. GCMD) and advancing the DAB capabilities to implement names harmonization.

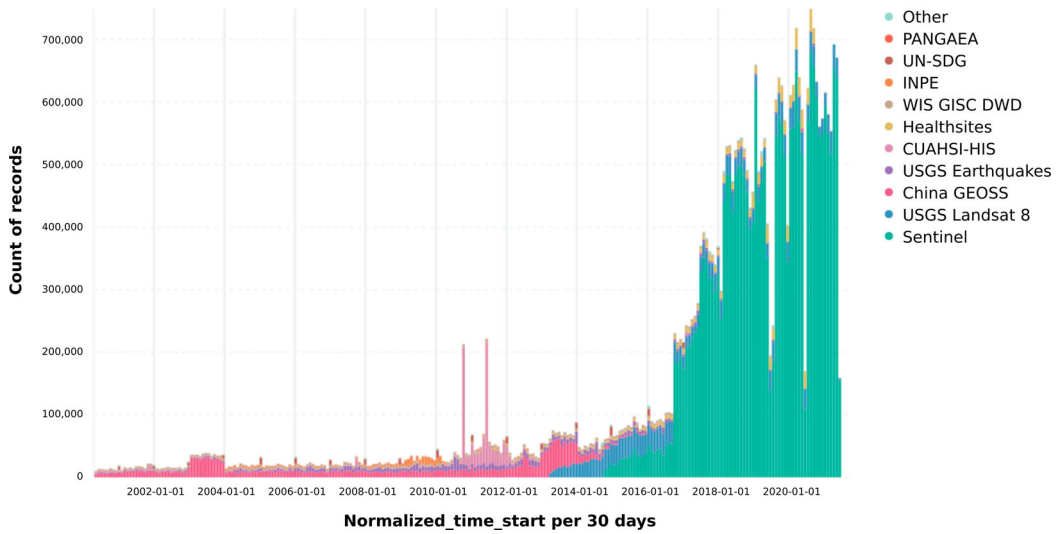
### 3.3.4. Temporal coverage

**3.3.4.1. What is the dataset temporal extension?.** For those datasets having a documented **temporal extent**, the plot of Figure 10 shows their occurrences by time. Temporal extent is the temporal information attached to the dataset by the data *Provider/Organator*; this metadata element represents the time range covered by the dataset (i.e. the observation). The plot is centered on the temporal period that represents the majority of GEOSS data (i.e. the last 20 years). It is useful to note the existence of a long tail of data that have been acquired in the past (including historical records recently digitized), although not shown. There is also a very small amount of synthetic data covering the next future although not appreciable from the figure (e.g. climate model forecasts). These data provide a good example of knowledge sharing as the result of an analytics process.

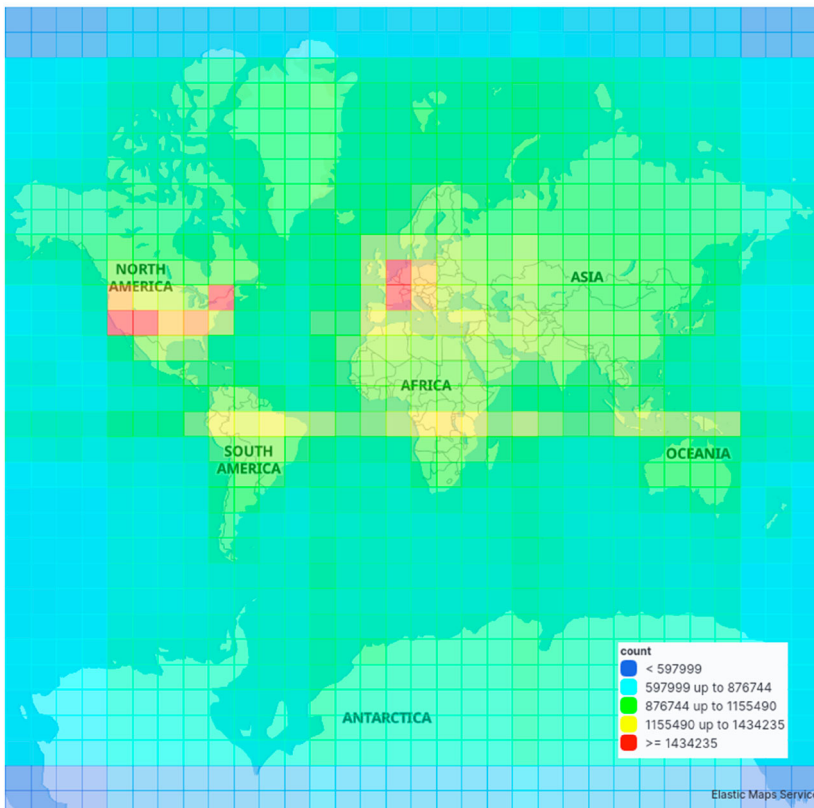
The analyzed temporal coverages show an exponential growth of observations in the last few years. That agrees with the increased availability of both in-situ and (predominantly) remote sensing datasets. As already discussed, most of these observations stem from the Sentinel sensors, whose granules (i.e. single scenes) are harvested by the DAB, continuously.

### 3.3.5. Geographic coverage

**3.3.5.1. Where is the dataset spatial extension?.** Almost all the records are georeferenced (see [Boldrini et al. 2021]); Figure 11 shows the heat map of the world areas, which are the most covered by the datasets accessible through the GEOSS Platform.



**Figure 10.** GEOSS dataset occurrences by their monthly aggregated temporal coverage. Sentinel datasets are the responsible for the data explosion on the right side starting around 2017.



**Figure 11.** GEOSS records by spatial coverage.

The most observed areas are Europe and the USA. This appears in accordance with the actual geographical distribution of data providers. The GEOSS Platform provides datasets on the polar areas (i.e. Svalbard and the Antarctic region), too. In general, we can notice that inland areas are better covered than oceans.

As expected, the Copernicus Open Access Hub is the major contributor of datasets over Europe. CUAHSI-HIS is also well-represented, although this US data Provider mostly shares data related to hydrological in-situ measurements. This stems from the fact that CUAHSI-HIS also shares satellite data with a global extent – such as NASA GLDAS.

The Copernicus Open Access Hub is the major datasets contributor (i.e. Sentinel global extent) for USA, as well. The datasets percentage of CUAHSI-HIS is also very high, and the USGS Earthquakes Community datasets is the third provider in terms of contribution. The fourth and fifth data providers are GISC WIS DWD and USGS Landsat 8.

About 20% of the registered datasets have a ‘point’ spatial extension, and they are most likely in-situ observations. On the other hand, the datasets characterized by a ‘rectangular area’ spatial extension include both regular grids (notably, satellite observations) and trajectories. Therefore, it is possible to infer that in-situ data are more than 20% of all the datasets accessible via the GEOSS Platform.

## 4. Analysis results: data demand

The second part of the analysis focuses on requests for GEOSS data. Since the majority of GEOSS traffic is contributed by harvesters (93%), it is very important to investigate the portion of regular users. This information is collected by looking at the log files – updating and completing some previous analysis (Craglia et al. 2017; Xia et al. 2014). The following three sections describe who are the GEOSS users focusing on human users, the number of requests they submitted in the last few years, and finally, a characterization of the request constraints and results is provided.

### 4.1. Who are the GEOSS users?

#### 4.1.1. Interaction type

Most of the registered traffic (about 9 billion requests, representing the 93.0% of the total) was originated from machine-to-machine interactions (i.e. clients). Only a relatively small portion of the traffic to the GEOSS platform (about half a million requests, representing the 5.56% of the total) is traceable to the discovery requests made by the users of the GEOSS portal (<https://www.geoportal.org/>).

Most clients are harvesters (i.e. **automatic software agents**) that periodically request to download the entire metadata content of the GEOSS platform (or a relevant subset of it), for various purposes (such as its analytics). The main sources generating by automatic software agents are traceable to KMA (i.e. WMO WIS GISC) and China GEOSS. The other organizations generating harvesting traffic (however to a much lesser extent) are USGS EROS and DLR (LOD-GEOSS). The other requests originate from other interaction types (e.g. retrieval of metadata details, and access of discovered datasets).

Importantly, a preliminary filtering process has been carried out to remove unwanted data from the analysis – such as test queries performed by developers of the GEO DAB and GEOSS portal. These were identified through their originating IP addresses. Queries performed by automatic tools were removed from the analysis of the human interactions with the GEOSS portal: this latter filtering process was not easy, because it seems that automatic tools were used to access the same service interface used by the GEOSS portal, thus making it difficult to distinguish between humans and clients. For example, many queries matching datasets from the China GEOSS data provider were identified as originating from automatic tools because of their characteristics (i.e. high volume of requests in a limited time range, repetition of exactly equal requests at regular, short, intervals)

and therefore removed from the human analysis. These requests were probably performed for harvesting or monitoring purposes.

The following sections will focus on the users (i.e. human) interactions with the GEOSS portal, trying to identify their profile, their search targets and how their searches are expressed.

#### **4.1.2. Originating organizations**

According to the analysis of the GEOSS Platform logs, the main Platform users belong to the academia and research domain. In addition, there is a consistent user activity, which seems to lead to private and commercial organizations making use of internet services providers (Boldrini et al. 2021).

Analyzing the number of days of use, it is possible to identify the most 'loyal' users (i.e. the users who utilized the GEOSS platform more frequently).

Research and academia organizations have been the most loyal along with a consistent activity interfaced by internet and cloud services providers. Presumably, these visits are generated by private and commercial users.

The introduction of a mandatory login on the GEOSS Portal could constitute a possible improvement for the definition of/ to profile the actual GEO users, for example, by using single sign-on mechanisms. On the other hand, single sign-on could discourage some potential users to utilize the GEOSS Portal. Perhaps, a valuable compromise would be implementing anonymization techniques with the only scope of statistics production in order to process user data.

#### **4.1.3. Originating countries**

Figure 12 illustrates the user's requests top countries of origin. They are ordered by count and are highlighted in increasing dark red colors. The first five top countries are: Germany, USA, Italy, Albania, and China.

Summing up the contribution coming from each of the 27 countries of the European Union, it is possible to see that the GEOSS platform usage from the EU region is predominant, as shown by the blue area in the chart of Figure 12, confirming the relevant role of EU in GEOSS.

## **4.2. Utilization trends**

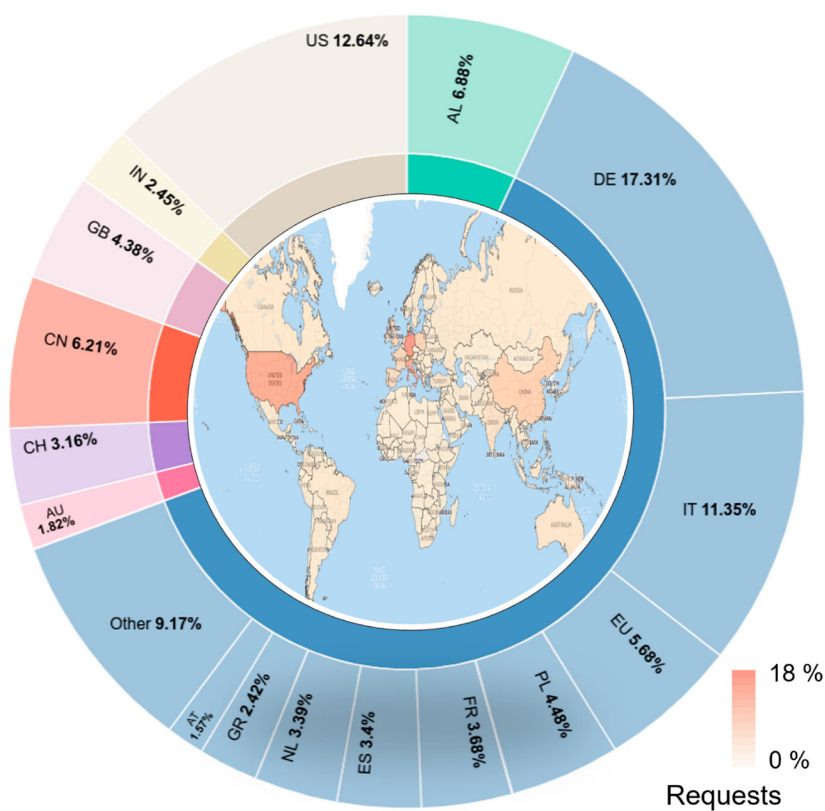
Figure 13's plot shows the histogram of the user query requests submitted, via the GEOSS portal, in the last four years. The average number of human requests is about 190 per day.

Referring to Figure 13, a general decrease of the human requests can be noticed starting from 2018: numbers in 2017 reached about 78 thousand requests, decreasing to an approximate number of 60 thousand requests in 2021. The factors that have likely caused the decrease of about 23% of requests in four years can be multiple, including: a minor visibility of the GEOSS platform in the recent GEO activities and events; the discontinuation of the GEOSS Data Providers workshops; the launch of the GEOSS Knowledge Hub initiative with the lack of information on interoperability with the GEOSS platform; the new interest of the GEO Flagships in using public cloud platforms that provide satellite data access.

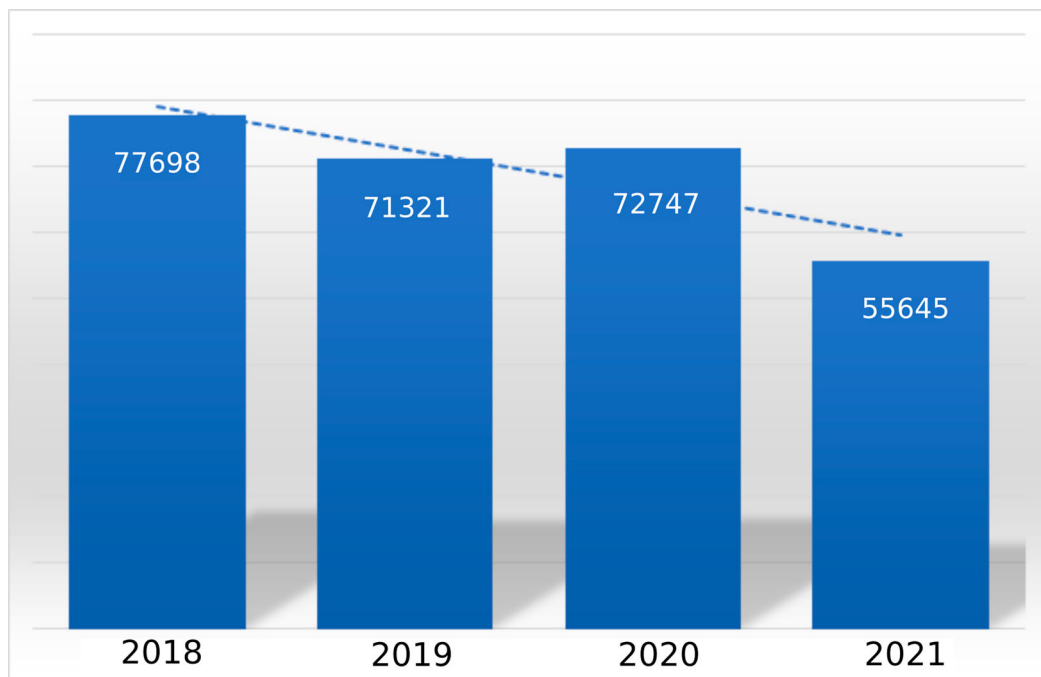
It is interesting to note that the utilization trend for the harvester portion is instead stable with the major regular harvesting being made by KMA (i.e. WMO WIS GISC). Additionally, other organizations have also performed spot harvestings during the last few years (most importantly China GEOSS in 2020 and 2021).

Almost 68 thousand visits have been logged by the GEOSS platform, each visit collecting all the requests made by a human user in the same day. These are started by around 56 thousand unique users. On average, a daily visit consisted of 3.84 requests – with a median of 2 requests.

A returning user can be defined as a user who already utilized the GEOSS portal in the past (e.g. by having submitted a discovery request) and returned to the portal for a new request. This piece of information is important because it provides hints about the user's satisfaction. In the past few



**Figure 12.** Countries originating the greatest number of requests to the GEOSS Platform. EU countries overall account for the highest percentage.



**Figure 13.** Number of requests per year submitted to the GEOSS portal.

years, the users who returned at least one time to the GEOSS Platform, were about 11,800, representing the 20.87% of all the unique users recognized by this study.

Figures 14 and 15, respectively, show the variation in the number of returning visits and their percentage over the last four years. The number and percentage have diminished in the last two

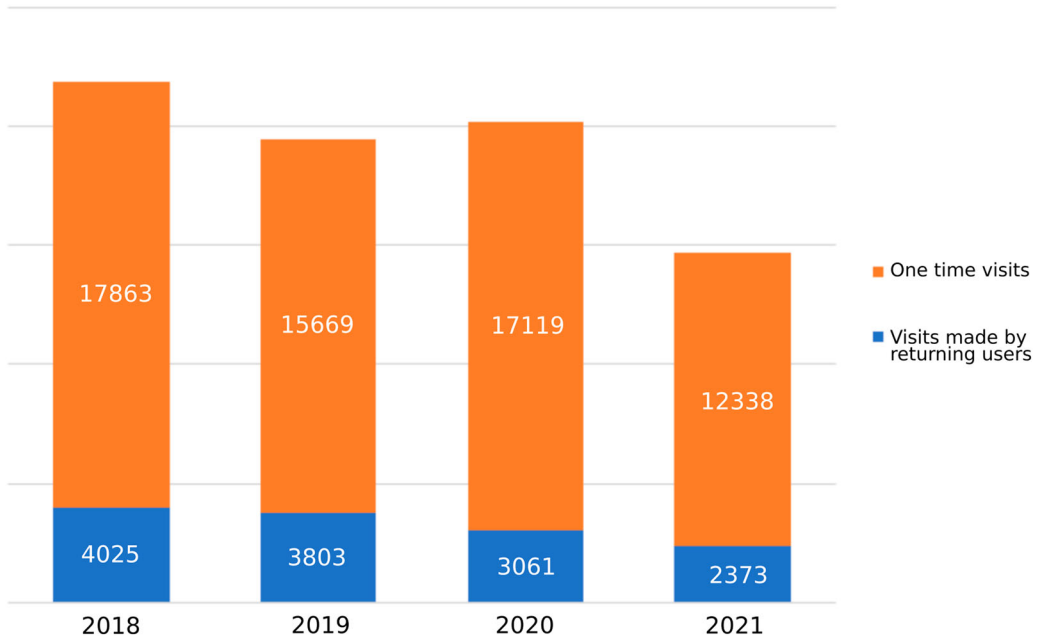


Figure 14. Users visit count from 2018 to 2021.

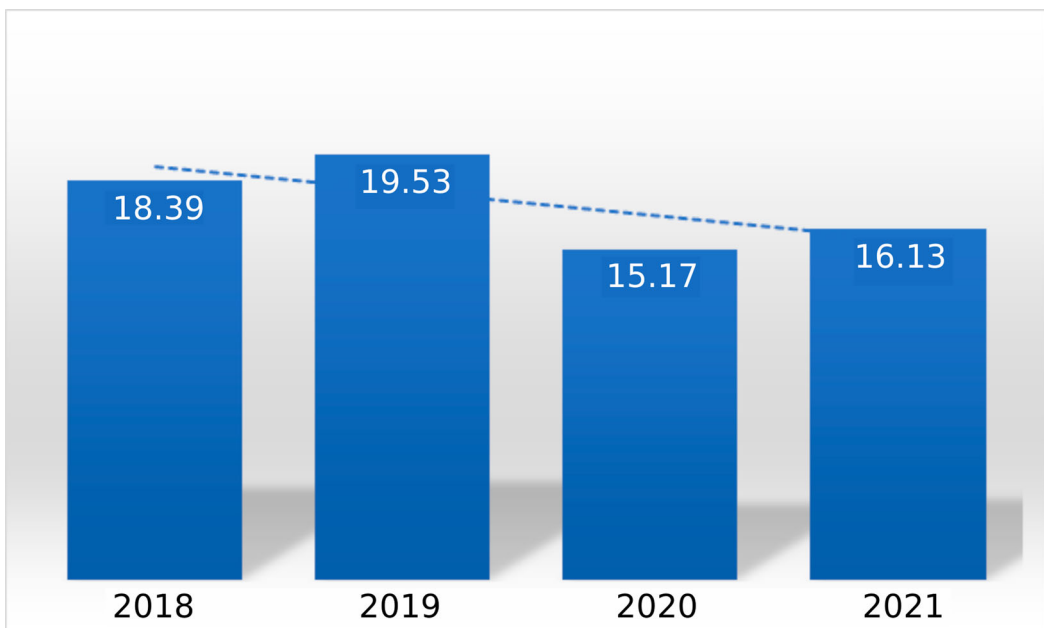


Figure 15. Percentage of returning users out of the total number of users.

years (i.e. 2020 and 2021). In addition, consistently with the decrease of the number of requests (see [Figure 13](#)), the total number of visits significantly dropped in 2021.

However, the statistics on the number of returning users must be taken with caution because they are based on the recognition of the user's IP, and this approach does not allow to identify a returning user if he/she makes use of a dynamic IP allocation. This is common for the case of users who worked from home during the COVID pandemic.

As already discussed, a mandatory login mechanism to the GEOSS portal would enhance an actual GEO user profiling.

More generally, GEO should further investigate the trends about returning users and, more generally, their experience with the portal. Likely, the cancelation of the GEOSS Data Provider workshops created a certain disaffection in some users, thus limiting the collection of their feedbacks and requirements.

Comparing the number of utilization days between 2018 and 2021, for the most active users it is possible to infer the following trends: (a) most active UK organizations seem to have stopped using the platform; (b) ISPs significantly increased their activity, likely caused by remote working; (c) new users appeared, probably for accessing in-situ datasets (Boldrini et al. 2021).

Notably, the total number of utilization days significantly dropped from approximately 800 in 2018 to about 200 in 2021.

### 4.3. What data is requested?

#### 4.3.1. Thematic coverage

This study analyzed the requests submitted by users and finalized by the GEOSS platform from January 2016 to December 2021, in order to recognize the most utilized query clauses. The most searched keywords (i.e. the 'keyword' clause of user queries) are: *chlorophyll*, *water*, *sentinel*, *earthquake*, *temperature*, *WMS*, *pollution*, *soil*, *wind*, *precipitation*, *sulfur dioxide*, *river*, *Landsat*, *flood*, *climate*, *deforestation*, *sea state*.

The top ten keyword occurrences (expressed in relative percentages) are depicted in [Figure 16](#).

The keywords used in searches are rather general, mostly indicating either a theme (e.g. *pollution*), a parameter (e.g. *temperature*), or a natural element (e.g. *water*), with the exception of satellite names (i.e. *Sentinel* and *Landsat*) that indicate an observation source. In this latter case, the GEOSS platform is used as a data hub for satellite datasets.

Notably, a number of top keywords are expected to characterize the datasets shared by some of the major harvested GEOSS data Providers: for example, 'Sentinel' and Copernicus Open Access Hub; 'water' and CUAHSI-HIS; 'earthquake' and USGS Earthquakes.

In general, it is possible to say that user interest has changed variably over the last years. [Table 1](#) reports the most searched keywords per year.

#### 4.3.2. Temporal coverage

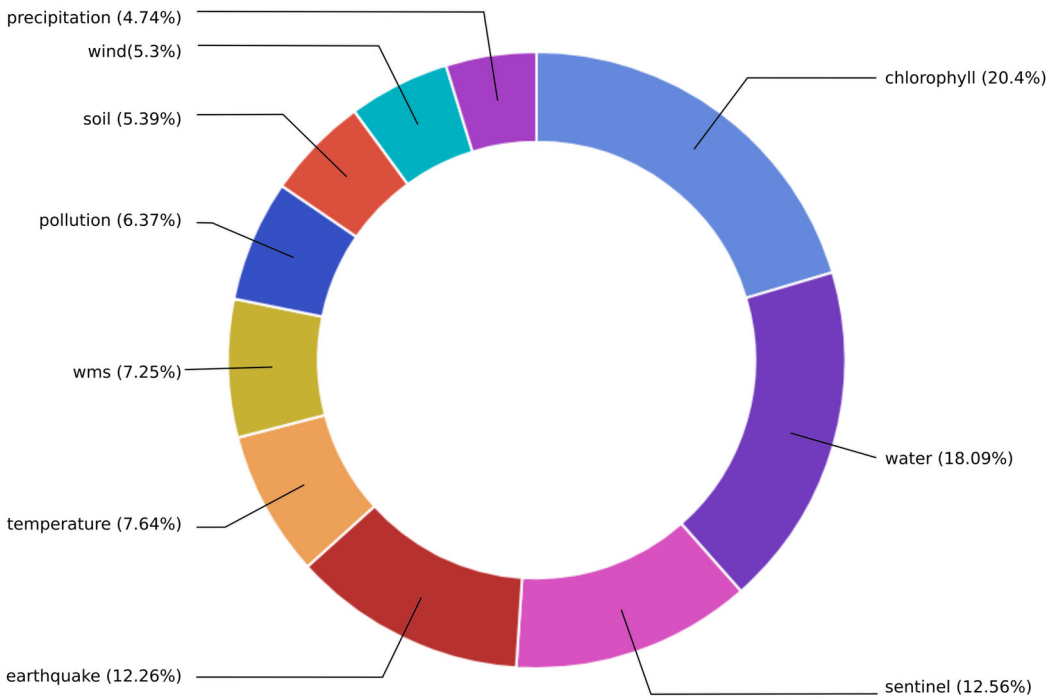
[Figure 17](#) shows the most searched temporal extents (for simplicity, aggregated by year) by the GEOSS platform user, over the past few years.

The most searched year was the 2017, and a large majority of requests were generated by users in the 2017 itself and the following year. This could be linked to the publication of the Sentinel datasets on the platform.

More generally, GEOSS platform users are interested in the more recent observation – i.e. datasets observed in the same year of the request.

#### 4.3.3. Area coverage

Many discovery requests made by users have a bounding box constraint. [Figure 18](#) shows the all-time most searched areas by the GEOSS Platform users (period 2017–2021).



**Figure 16.** Relative percentages of the most searched keywords by GEOSS users.

**Table 1.** Top searched keywords in the last few years.

Year	1st keyword	2nd keyword	3rd keyword	4th keyword	5th keyword
2016	WMS	Sulfur dioxide	Precipitation	Sulfur	Land cover
2017	Water	Earthquake	Sentinel	Precipitation	River
2018	Chlorophyll	Earthquake	Water	Sentinel	Soil
2019	Chlorophyll	Water	Earthquake	Solar energy	Land degradation
2020	Water	Temperature	Sentinel	Soil	Meteorology
2021	Soil	Water	Sentinel	Earthquake	Temperature

Note: The keywords *sentinel*, *water*, *earthquake*, and *soil* are always present over the past six years.

According to [Figure 18](#), Europe (especially Central Europe), Asia (in particular the Middle East, Nepal, South India and Southeast Asia), Central America and parts of Africa (especially Uganda) are amongst the most searched areas as from the setup of the GEOSS Platform.

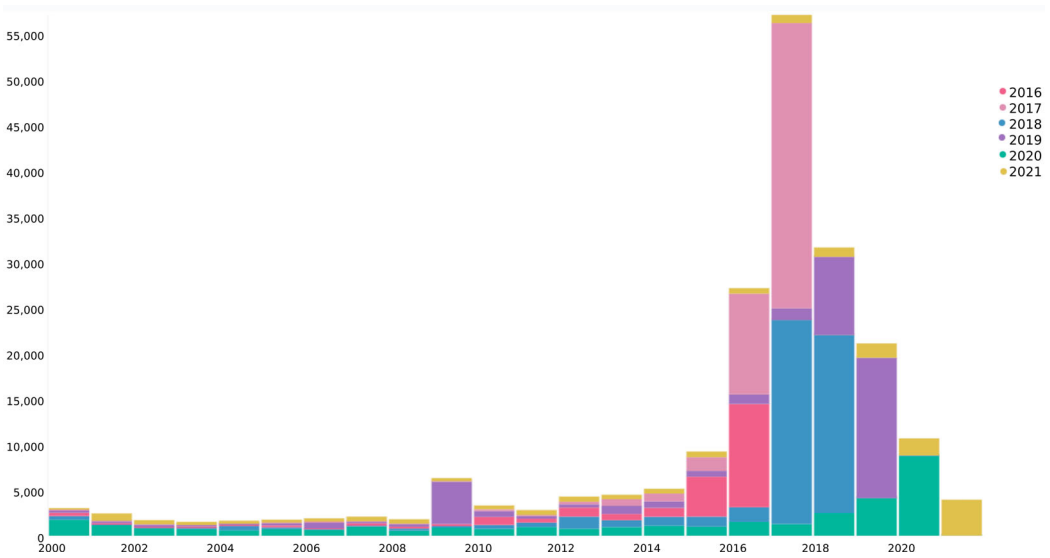
[Figure 19](#) shows the search evolution in the period 2017–2021. The map shows some changes in the last years: for example, the most requested areas changed from Atlantic Ocean, Africa and South America to Mediterranean Sea, Central America, Southeast Asia and Arctic polar region including the Svalbard.

Filtering data by user countries, it is also possible to visualize the most searched areas, depending on their country of origin.

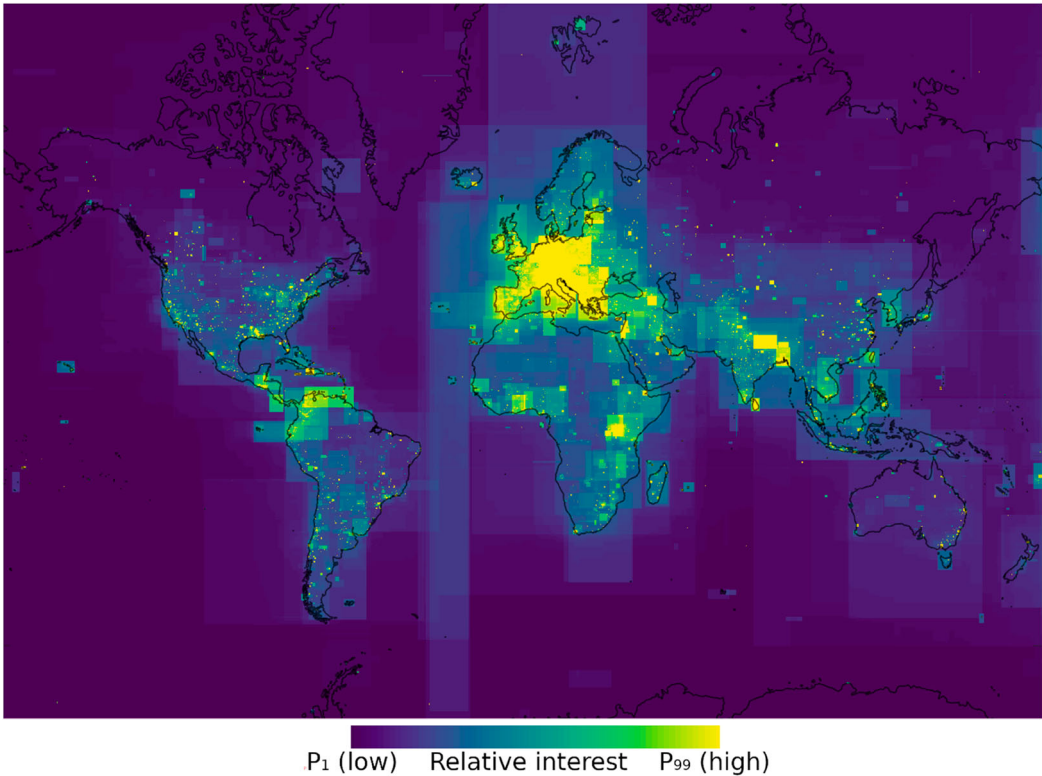
Not surprisingly, in general, the most searched area matches with the country where the searches originated. However, in some cases other areas are deemed of interest for social and/or economic reasons.

#### 4.3.4. Popular data providers

The plot in [Figure 19](#) shows the 15 most popular data providers appearing in the result set of queries issued from January 2021. Each time a query is executed, the GEO-DAB takes note of matching



**Figure 17.** Most searched temporal extents (with a yearly resolution). The different colors indicate the year when the query was issued.



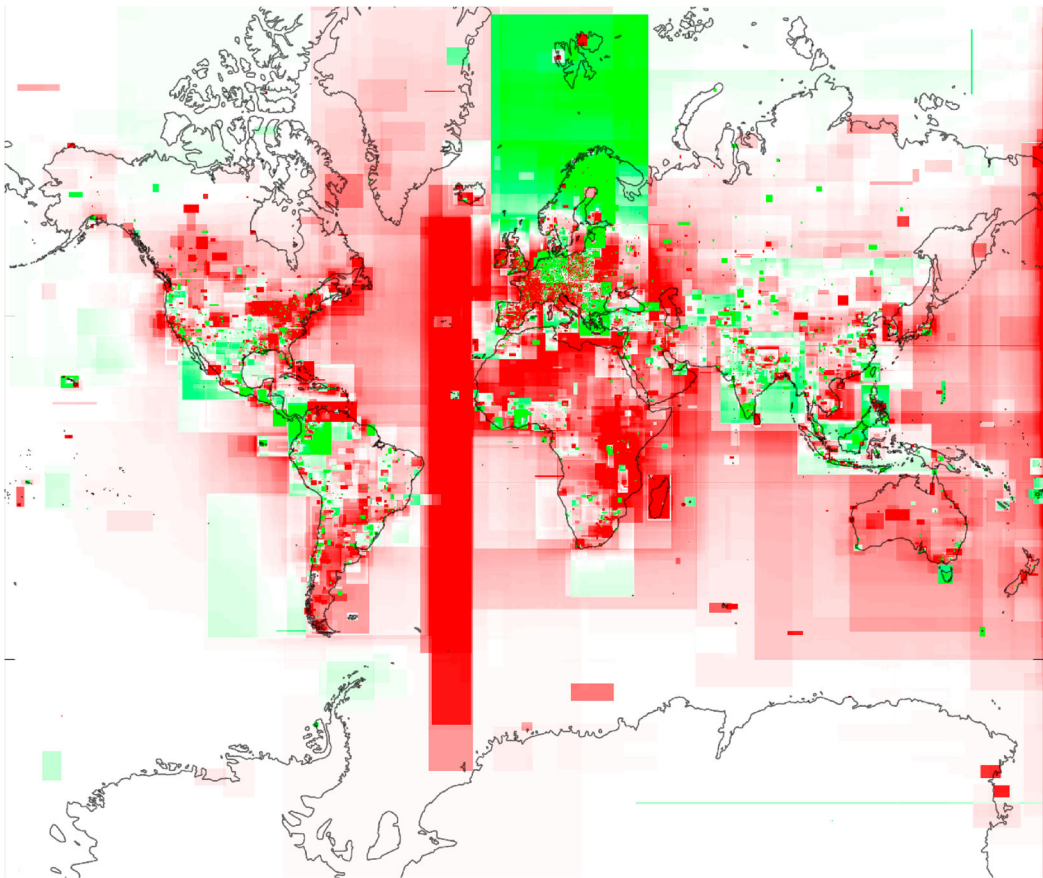
**Figure 18.** Most searched spatial extents. The contribution on a pixel of each query is weighted according to the extent of the query bounding box. The relative interest is calculated on each pixel as the combined contributions of all the queries divided by the total. It is finally plotted on a world map according to a logarithmic scale in the range between the 1st and the 99th percentiles of its distribution.

records and their relative data providers. The plot shows the number of matching queries for each data provider, hence highlighting the most popular ones.

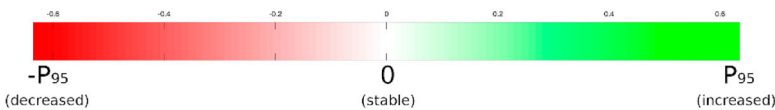
The most popular data providers are Copernicus Open Access Hub, USGS Landsat 8, China GEOSS and INPE. They have dataset collections that cover the entire globe, greatly contributing to their popularity, they are providers of satellite imagery (Figure 20).

Three factors contribute in general to the matching of records and hence to the popularity of a data provider:

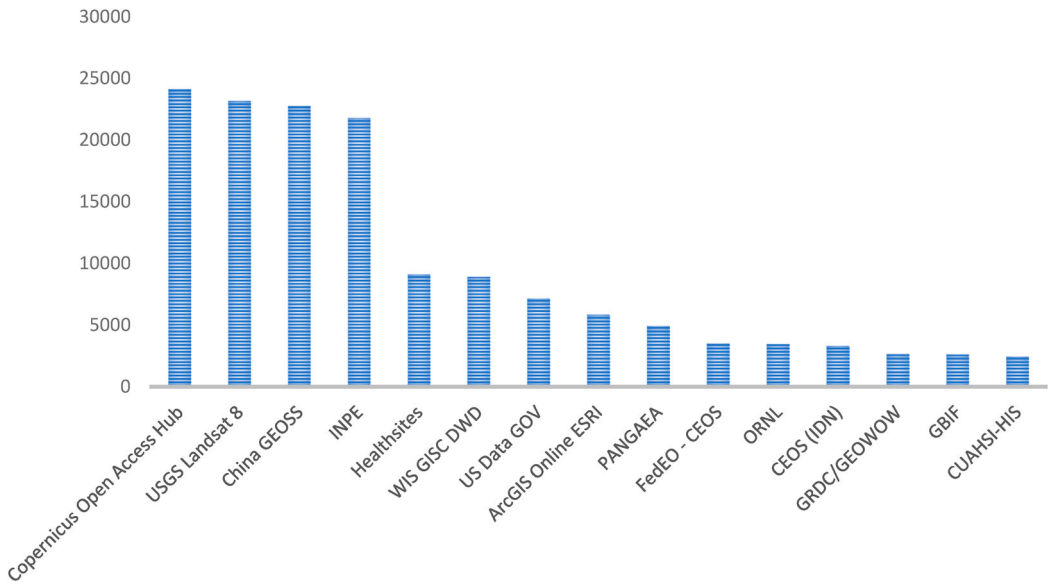
- (1) Users formulate queries on the exact data provider content (e.g. on the provider spatial-temporal coverage or using keywords appearing in the data provider metadata records).
- (2) The data provider records are described in great detail (for example, by using more keywords), having thus more chances of matching user queries.
- (3) The data provider records cover a large spatial and or temporal extent, having thus more chances of matching user queries.



Changes in relative interest from first to second half of 2017-2021 period



**Figure 19.** Changes in searched spatial extents from the first half to the second half of the 2017–2021 period. The plot shows the ratio between the relative interest of the two periods, expressed in base 2 logarithmic scale between the – 95th and the 95th percentile of the positive distribution. Areas such as Arctic polar region including Svalbard, Mediterranean Sea, Central America and SouthEast Asia have an increased relative interest, while areas such as Atlantic Ocean, Africa and South America have a decreased relative interest.



**Figure 20.** Most popular data providers – most of the queries from January 2021 returned data from Copernicus Open Access Hub, USGS Landsat 8, China GEOSS and INPE.

## 5. Conclusions and recommendations

Presently, looking at the platform content and user requests, GEOSS is a popular satellite imagery access platform. Notably, the Copernicus Open Access Hub publishes about the 60% of the datasets shared through the GEOSS Platform. The analyzed temporal coverages show an exponential growth of observations, in the last few years. That agrees with the increased availability of both in-situ and (predominantly) remote sensing datasets (e.g. Sentinel sensors).

The most observed areas are Europe and the USA. This appears in accordance with the actual geographical distribution of data providers. It is possible to estimate that the GEOSS Platform in-situ data are between 20% and 30% of the total. The overall metadata quality of the datasets registered in the DAB database is low, because some major data providers do not specify a set of recommended metadata elements to describe their shared data – e.g. ‘keywords’, ‘abstract’, and ‘originating organization’. This issue represents a major gap in GEOSS metadata quality and discovery match. The acknowledged low level of quality of dataset descriptions stems from the difficulties encountered by the data Providers in the implementation of the GEOSS data management and FAIR principles and guidelines. Likely, it significantly contributed to the already-mentioned disconnection, recently experienced, between GEOSS Community and the platform.

In the last few years, the GEOSS Platform was mainly used by clients for machine-to-machine interactions. Harvesting requests have been, by far, the most relevant, in terms of traffic activity engaging the platform. These requests came from organizations such as – e.g. WMO, China GEOSS, USGS EROS, and DLR. The majority of human requests submitted to the GEOSS Platform are originated by users from EU, USA, and China, respectively.

The most popular data providers are satellite imagery data sources, with Copernicus Open Access Hub at the first place. Being characterized by a global coverage, the matching of these datasets is very likely (e.g. against queries with a single bounding box constraint), thus greatly contributing to their popularity. The keywords *sentinel*, *water*, *earthquake*, and *soil* have always been part of the most used keywords over the past six years. Though the in-situ dataset requests seem to be increasingly popular, still, the search for satellite time series imagery (i.e. Sentinel) are the most frequent, by far.

Despite a significant growth in published datasets (especially satellite ones), the acknowledgement of an important decrease of user requests (i.e. human users), both in terms of absolute value and platform utilization days is assessed. However, a decreased pace of insertion of new data providers is also to be noted. The most active users of the platform have changed over time (with some UK organizations no longer being active), whereas new research and academia organizations have become users of the platform, likely due to their interest in in-situ observations. Although the main users belong to the academia and research domain, after 2020 a notable increase of ISP activity was raised, possibly due to the pandemic situation. To improve the understanding of who are the actual GEO users, a possible recommendation is to introduce a mandatory login on the GEOSS Portal, in order to provide a detailed user profiling.

The possible causes for the drop in the number of user requests are interconnected, and likely stem from both GEO internal and external factors, in particular:

- (a) the progressive disconnection of the GEOSS platform from its data Providers and Users, a couple of valuable examples are the discontinuation of the GEOSS data Providers (and Users) workshop, and the lack of the platform dissemination and training at the most recent GEO and GEOSS events;
- (b) the partial ineffectiveness, related to the previous factor, of the current discovery and/or subsequent data access services – a valuable example is the search for Sentinel data;
- (c) the launch of the GKH initiative, lacking a clear message about its synergy and interoperability with the GEOSS Platform.

As for the external factors, the following issues are to be deemed as a possible cause of a moderate number of user requests:

- (a) the publication of satellite long time series datasets on other public and commercial clouds – e.g. AWS, Google EE, and Microsoft Azure;
- (b) the COVID effect, with the advent of remote working and a decisive push of computing virtualization;
- (c) the advent of the cyber-physical paradigm that requires high computing scalability.

To revamp users' interest, possible recommendations are:

- (a) to re-establish a strong connection with the GEOSS Community working closely with each data provider as undertaken by some recent efforts (Roncella Zhang, et al. 2022; Roncella, Boldrini, et al. 2022), rescheduling regular GEOSS data Providers and Users workshops, also to the aim of formally identifying gaps and requirements to be subsequently implemented by the GEOSS platform and by the data providers (i.e. to tackle the low metadata quality issue); where useful, to jointly define community tailored views and portals;
- (b) to differentiate the needs of platform clients versus users, responding with increasingly specialized services;
- (c) to present, in a clear manner, an overall design where the GKH is part of and advances the GEOSS Platform; a possible synergy could be achieved by enabling an interoperable data workflow that, starting from observations published to the GEOSS platform, produces the products available on the GKH (e.g. exploiting knowledge generation frameworks such as the VLab [Santoro, Mazzetti, and Nativi 2020; Santoro, Mazzetti, and Nativi 2023]).
- (d) to reinforce the role of the GEOSS Platform for in-situ data sharing, for example establishing connections with complementary in-situ data sharing initiatives such as the WMO Hydrological Observing System (WHOS) (Boldrini et al. 2022);

- (e) to advance the GEOSS platform capabilities by adding: computing scalability, relying on existing and distributed infrastructures; knowledge generation processing
- (f) in keeping with the original system-of-systems (multi-lateral) philosophy, to evolve the GEOSS platform into a geoscience digital ecosystem.

Noticeably, the thrive of cyber-physical interactions could also be the main reason also for another significant change observed in this analysis: the predominant utilization of the GEOSS platform through machine-to-machine requests, as opposed to the decreased number of human requests. More generally, it is possible to argue about a progressive evolution in the way geospatial information is nowadays discovered and accessed. In other words, software clients (which understand and mediate actual user needs) are replacing human experts, who perform manual searches and download datasets.

## Acknowledgements

The authors would like to thank Fabrizio Papeschi, Roberto Roncella, Massimiliano Olivieri and Roberto Salzano from CNR-IIA for their contributions to the GEOSS cloud platform and statistics collection, Lena Rettori from CNR-IIA for English text revision; Guido Colangeli and Joost Van Bemmelen from ESA for their contributions to the GEOSS cloud platform and statistics collection.

## Disclosure statement

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

## Funding

This research was funded by EOValue project funds from European Commission Directorate-General for Research and Innovation; DAB4EDGE project funds from European Space Agency [ESA grant agreement 4000123005/18/IT/CGD]; DAB4GPP project funds from European Space Agency [ESA grant agreement 4000138128/22/I/AG].

## Data availability statement

The data that support the findings of this study are available from the Group on Earth Observations (GEO). GEOSS Data Sharing Principles apply to the availability of these data, which were used for this study. More information about GEOSS Data Sharing Principles can be found at: <https://www.earthobservations.org/dswg.php>.

## ORCID

Enrico Boldrini  <http://orcid.org/0000-0003-2075-4742>  
Stefano Nativi  <http://orcid.org/0000-0003-3185-8539>  
Mattia Santoro  <http://orcid.org/0000-0003-0401-3100>  
Paolo Mazzetti  <http://orcid.org/0000-0002-8291-1128>  
Max Craglia  <http://orcid.org/0000-0001-6244-6428>

## References

- Boldrini, Enrico, Jiri Hradec, Max Craglia, and Stefano Nativi. 2021. *EOValue 1.1 v.2 GEOSS Content Exploration*. Ispra: Joint Research Center.
- Boldrini, Enrico, Stefano Nativi, Silvano Pecora, Igor Chernov, and Paolo Mazzetti. 2022. "Multi-Scale Hydrological System-of-Systems Realized through WHOS: The Brokering Framework." *International Journal of Digital Earth* 15 (1): 1259–1289. doi:10.1080/17538947.2022.2099591.
- Craglia, M., J. Hradec, S. Nativi, and M. Santoro. 2017. "Exploring the Depths of the Global Earth Observation System of Systems." *Big Earth Data* 1: 21–46. doi:10.1080/20964471.2017.1401284.

- Craglia, Massimo, Stefano Nativi, Mattia Santoro, and Lorenzino Vaccari. 2011. "Inter-Disciplinary Interoperability for Global Sustainability Research." *Lecture Notes in Computer Science* 6631: 1–15. doi:10.1007/978-3-642-20630-6\_1.
- GEO. 2009. *GEOSS Strategic Targets*. Geneva: GEO-WMO.
- GEO. 2015a. *GEO Discovery and Access Broker*. <http://www.geodab.net>.
- GEO. 2015b. *GEO Strategic Plan 2016–2025: Implementing GEOSS*. Geneva: GEO-WMO.
- GEO. 2015c. "The Group on Earth Observations (GEO) Mexico City Declaration." *Earthobservations.org*, November 13. [https://earthobservations.org/documents/ministerial/mexico\\_city/MS3\\_Mexico\\_City\\_Declaration.pdf](https://earthobservations.org/documents/ministerial/mexico_city/MS3_Mexico_City_Declaration.pdf).
- GEO. 2018. "3rd GEO Data Providers Workshop." [https://sedac.ciesin.columbia.edu/binaries/web/global/news/2018/gdpw3\\_desherbini\\_et\\_al\\_day1\\_plenary-1700.pdf](https://sedac.ciesin.columbia.edu/binaries/web/global/news/2018/gdpw3_desherbini_et_al_day1_plenary-1700.pdf).
- GEO. 2022a. "About Us." March. <https://earthobservations.org/index.php>.
- GEO. 2022b. "GEO Knowledge Hub." <https://gkhub.earthobservations.org/>.
- GEOSS Infrastructure Development Task Team. 2017. *The GEOSS Platform: All You Need to Know to Become a GEO Data Provider*. Geneva: GEO.
- GO FAIR. 2011. *Fair Principles*. <https://www.go-fair.org/fair-principles/>.
- Guo, H., S. Nativi, D. Liang, M. Craglia, L. Wang, S. Schade, C. Corban, et al. 2020. "Big Earth Data Science: An Information Framework for a Sustainable Planet." *International Journal of Digital Earth* 13 (7): 743–767. doi:10.1080/17538947.2020.1743785.
- ISO. 2003. "ISO 19115:2003 – Geographic information – Metadata." <https://www.iso.org/standard/26020.html>.
- Nativi, Stefano, and Max Craglia. 2021. *Destination Earth: Ecosystem Architecture Description*. Luxembourg: European Union/JRC, European Commission. doi:10.2760/08093.
- Nativi, S., M. Craglia, and J. Pearlman. 2013. "Earth Science Infrastructures Interoperability: The Brokering Approach." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 6: 1118–1129. doi:10.1109/JSTARS.2013.2243113.
- Nativi, S., and P. Mazzetti. 2021. "Geosciences Digital Ecosystems." In *Encyclopedia of Mathematical Geosciences*, edited by Q. Cheng, J. McKinley, F. Agterberg, and Sagar B. Daya, 1–6. Cham: Springer.
- Nativi, S., P. Mazzetti, and M. Craglia. 2021. "Digital Ecosystems for Developing Digital Twins of the Earth: The Destination Earth Case." *Remote Sensing* 13 (11): 2119–2144. doi:10.3390/rs13112119.
- Nativi, S., P. Mazzetti, M. Santoro, F. Papeschi, M. Craglia, and O. Ochiai. 2015. "Big Data Challenges in Building the Global Earth Observation System of Systems." *Environmental Modelling & Software* 68: 1–26. doi:10.1016/j.envsoft.2015.01.017.
- Roncella, Roberto, Enrico Boldrini, Mattia Santoro, João Andrade, Nuno Catarino, and Stefano Nativi. 2022. "Publishing NextGEOSS Data on the GEOSS Platform." *Big Earth Data*, 1–15. doi:10.1080/20964471.2022.2135234.
- Roncella, Roberto, Lianchong Zhang, Enrico Boldrini, Mattia Santoro, Paolo Mazzetti, and Stefano Nativi. 2022. "Publishing China Satellite Data on the GEOSS Platform." *Big Earth Data*, 1–15. doi:10.1080/20964471.2022.2107420.
- Santoro, Mattia, Paolo Mazzetti, and Stefano Nativi. 2020. "The VLab Framework: An Orchestrator Component to Support Data to Knowledge Transition." *Remote Sensing* 12 (11): 1795. doi:10.3390/rs12111795.
- Santoro, Mattia, Paolo Mazzetti, and Stefano Nativi. 2023. "Virtual Earth Cloud: A Multi-Cloud Framework for Enabling Geosciences Digital Ecosystems." *International Journal of Digital Earth* 16 (1): 43–65. doi:10.1080/17538947.2022.2162986.
- Xia, J., C. Yang, Z. Gui, K. Liu, and Z. Li. 2014. "Optimizing an Index with Spatiotemporal Patterns to Support GEOSS Clearinghouse." *International Journal of Geographical Information Science* 28 (7): 1459–1481. doi:10.1080/13658816.2014.894195.