GIDTT Technical Assessment of the GEO Infrastructure

This document is submitted to the Executive Committee for information.

GIDTT Technical Assessment of the GEO Infrastructure

Authors: GEO Infrastructure Development Task Team

Working period: 30 September 2023 to 19th March 2024

Table of Contents

Introduction 3

Space based data, Data Cubes, Cloud Providers and Very High-Resolution Data 7

Data Cubes & Tools 8

Data cube services (e.g., STAC, OGC Data Cube API) and Interoperability 8

Capacity development on Data Cubes – Long term sustainability 9

Executable Notebooks 9

Services 9

Enhance the discoverability and accessibility of the CEOS satellite data 9

Cloud providers 10

In situ data and Regional GEOs 10

Enhancing discoverability and accessibility of in situ data 10

Data format and metadata harmonization 11

Hosting of in situ data 12

GEO Work Programme Activities needs and requirements 12

User Interface and domain-oriented approach 14

Quality Check – Integration of GEOSS Platform and GEO Knowledge Hub, and Governance 14

Governance 14

Data Stewardship and quality control 15

Quality Check of EO Derived data 15

Integration of the GEOSS Platform and the GEO Knowledge Hub 15

GEO Infrastructure User Interface 16

Reference documents and presentations 17

Introduction

We read in the Post 2025 Strategy:"There is a clear need for a global partnership where data providers and users from all communities work together, leading to better coordination, greater inclusion, reduced duplication, and faster action" and "Building on the achievements of the past 10 years, GEO reaffirms its commitment to full and open access to Earth observation data, knowledge, products and services. GEO also reaffirms its commitment to promote data and knowledge sharing and the co-development of services that empower users to make sound environmental decisions, enable economic opportunities and promote good governance".

The GEO Infrastructure plays a pivotal role in advancing GEO's role within the international community and demonstrates the importance and impact that Earth Observations have in the decision-making process. It is in this framework that the GEO Infrastructure Development Task team (GIDTT) is working to design a User's friendly, impactful GEO Infrastructure

At its 60th meeting, GEO Executive Committee endorsed the direction for the evolution of the GEO Infrastructure, as proposed by the GEO Secretariat in response to the Expert Advisory Group on GEOSS (EAG) report.

The decision includes re-establishing the GEO Infrastructure Task Team (GIDTT) to assess the technical feasibility and the requirements (institutional and financial resources, and governance model) for the implementation of the endorsed direction.

The GIDTT was expected to start its mandate in September 2023, reporting directly to the Executive Committee (ExCom) and deliver the following deliverables:

Timeline of Engagements and Deliverables:

Document	Due date
Technical assessment of the architecture proposed by GEO Secretariat and GIDTT	Executive Committee April 2024
Finalize the GEO Infrastructure Architecture & Identification of requirements (institutional, financial and governance) needed to implement the architecture	Executive Committee - ~~July~~ November 2024
Evolved GEO Infrastructure Implementation Plan for 2024-2025	GEO Plenary 2025 (April 2025)
Demonstration and start of GEO Infrastructure operations	GEO Plenary 2025 (April 2026)

This document responds to the first deliverable:

Technical assessment of the architecture proposed by GEO Secretariat to make the GEO Infrastructure a User Driven tool to facilitate the discovery, accessibility, usability of Earth Observations data, knowledge and services provided by the large EO community.

This assessment is done considering the role that the GEO infrastructure has in respect to the GEO Community.

The initial draft architecture proposed to EXCOM in March 2023, presented in Figure 1, indicates 3 main elements highlighted by the Expert Advisory Group (EAG), that requires specific attention for the evolved GEO Infrastructure.

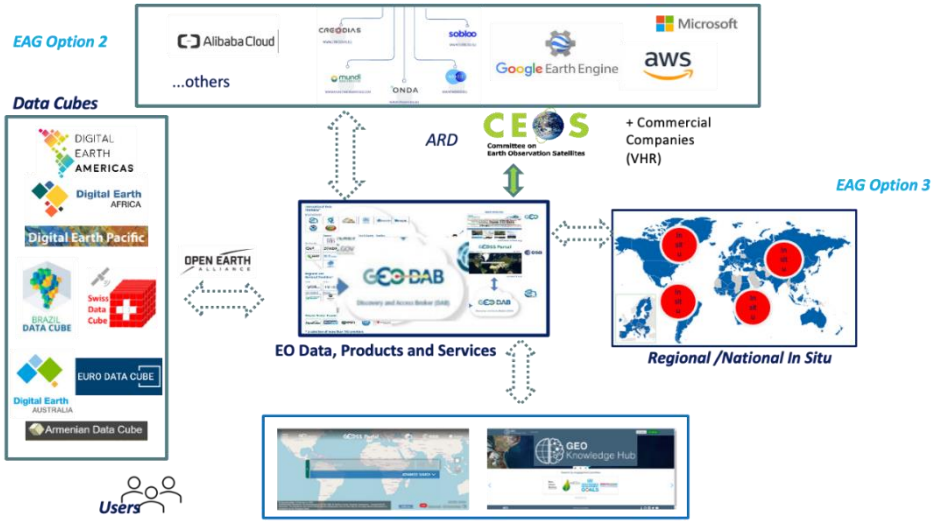


Figure. 1 Architecture proposed by the GEO Secretariat in EXCOM March 2023

Based on the topics highlighted below the GIDTT has developed the 1st deliverable outputs:

Discovery and Access of Space based Data:

1) Establishing discoverability and accessibility of the Data cubes ecosystem, at the moment totally lacking within the actual infrastructure
2) Improving discoverability, accessibility and utilization of Space based data provided by CEOS and other important Space data providers.
3) Including Services and Analysis Ready Data[1] provided by key EO players.
4) Establishing Partnership with Cloud Providers dealing with Space based data
5) Exploring provision of Very High-Resolution data for the GEO Community.

[1] In the context of GIDTT, the term Analysis Ready Data is used as defined by CEOS and OGC.

In-situ data and Regional GEOs:

1) Enhancing the discoverability of existing In-situ data networks (In situ data networks defined as in situ data observing networks owned and operated by institutions and agencies at national level, including, for example, national meteorological and oceanographic services, research institutions or mapping and cadastral agencies, source: https://insitu.copernicus.eu/state-of-play/understanding-in-situ-data)
2) Exploring long-term hosting capabilities for in-situ data collected by time-limited funded projects
3) Enhancing the role of the Regional and National GEOs with respect to the In-Situ data sharing

Quality Check, Integration of GEOSS Platform and GEO Knowledge Hub:

1) Assessing the discoverability and accessibility of already brokered EO data by the GEOSS Platform
2) Assessing the typology of already brokered EO data by the GEOSS Platform with specific attention to existing in-situ data
3) Integrating the two existing modules of the GEO Infrastructure: GEOSS Platform and GEO Knowledge Hub
4) Establishing the GEO Infrastructure Governance

Also, together with the deliverables listed below, GIDTT revised Figure 1, which was proposed by the GEO Secretariat, to clarify the relations and the high-level organization of the GEO Infrastructure. In this revised version, presented in Figure 2, the links between the GEO Infrastructure and other components and partners, such as cloud provider and Regional GEOs, are highlighted. Technical details and connections will be addressed in future deliverables from GIDTT.

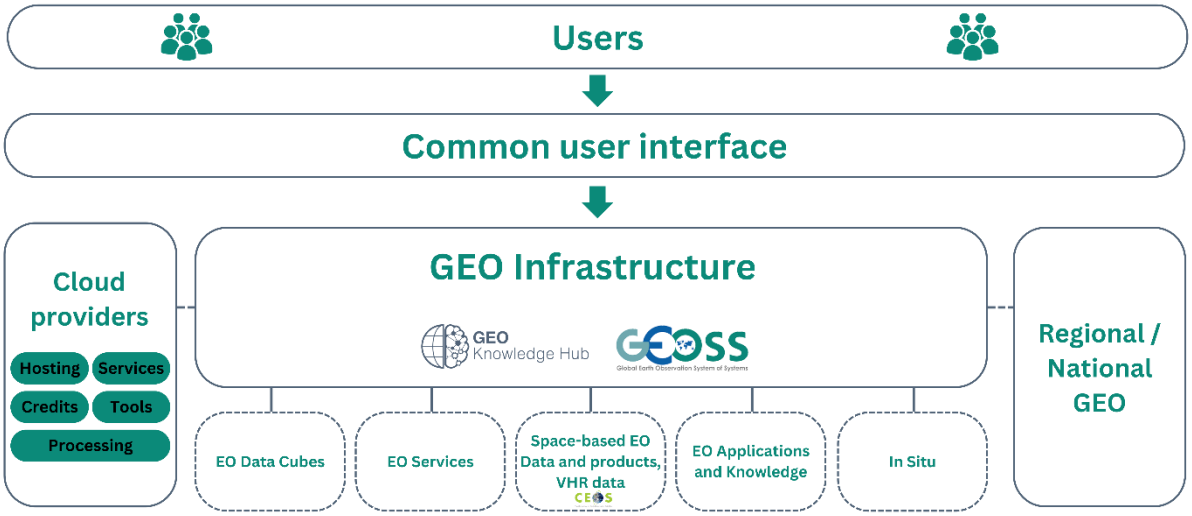


Figure. 2 Revised figure of the Architecture proposed by the GEO Secretariat in EXCOM March 2023

To respond to the 1st deliverable the GEO Secretariat with the agreement and guidance from the EXCOM has worked with the GIDTT nominated members (GIDTT Membership table).

The GIDTT Membership table indicates the categories necessary for the successful operation of the GIDTT. These categories represent essential roles and components of the proposed GEO Infrastructure Architecture.

The work has been distributed over 3 subgroups

- Space based data, Data Cubes, Cloud Providers and Very High-Resolution Data
- In situ Data and Regional GEOs
- Quality check of non-space-based data in the GEOSS Platform/Integration GKH and GEOSS Platform

Each subgroup had a total of 4 monthly meetings (starting from 30 September 2023), and the entire GIDTT had 3 meetings in total. Geographical attendance is shown in the annex.

The common goal of the 3 subgroups and the entire GIDTT is building a simplified, user-friendly GEO Infrastructure with enhanced functionalities to support the GEO community in the usage of EO for the decision-making as presented in Figure 3.

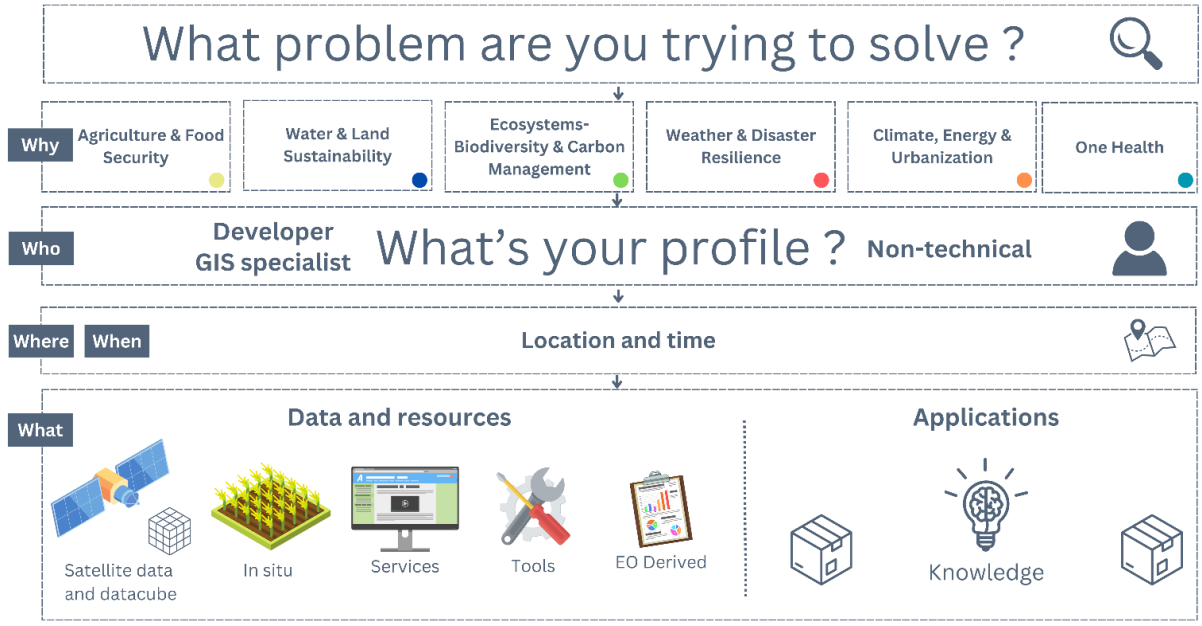


Figure. 3 Simplified schema of the user-oriented GEO Infrastructure

GIDTT considers important to redesign the GEO Infrastructure following the 5 W approach:

Why: Any user in need of Earth Observations will have a scope or specific problematic to solve; therefore, it is essential that the user can be guided, in the search of EO data/ knowledge by community agreed thematic areas.

Who: The GEO Infrastructure should respond to the needs of several profiles of users, ranging from highly technical to not technical experts; this will influence the typology of results provided by the search.

Where and **When**: The user will be able to define the place and time for which he/she needs EO data /knowledge

What: Based on the choices previously selected the user will be prompted with EO data/knowledge responding to his/her needs

Based on this 5W approach the 3 subgroups have worked on:

Space based data, Data Cubes, Cloud Providers and Very High-Resolution Data

Data Cubes & Tools

Data Cube technology has been widely used within the GEO community and by some Members in the last few years, with great advantages to easily produce valuable analysis of Space based data over time.

Data Cubes, despite representing an important effort within the space-based data ecosystem, are not yet discoverable and accessible by the large majority of EO users.

Among the various tasks of the GIDTT is to explore integration of the Data Cubes within the GEO Infrastructure to enhance their discoverability, accessibility and usability by a broader user community rather than few researchers.

The actual assessment is that Data Cubes ecosystem present a sparse landscape and therefore the subgroup has started the task on collecting the existing Data Cubes with specific attention to their geographical focus, technology used, stakeholders, status and reference. and created an initial inventory available here..

This catalog reveals a diversity of Data Cubes design (ODC or others), implementation, and technologies used.

Data cube services (e.g., STAC, OGC Data Cube API) and Interoperability

In a very varied landscape, the GIDTT initially discussed the registry of Data cubes to make them available in the GEO Infrastructure. This registry will act as a list, presenting all relevant data cubes for the GEO Community. Similar tools, like the STAC Index, were discussed.

The subgroup has also discussed the importance of establishing interoperability to the Data Cubes, leveraging the existing Data Cubes APIs that the OGC is working on Data Cube Interoperability - Open Geospatial Consortium (ogc.org) An OGC Geodatacube Standard Working Group is already established and active. Since the discussion paper summarizing the 2 days "Towards Data Cube Interoperability" workshop co-organized by GEO and OGC in April 2021 (https://portal.ogc.org/files/?artifact_id=99131) , OGC has establish this working group and several Engineering Reports resulting from Interoperability initiatives have been published or are currently under review to be published soon (OGC Testbed 17: Geo Data Cube API Engineering Report https://docs.ogc.org/per/21-027.html)Data Cube Interoperability - Open Geospatial Consortium (ogc.org) To understand and test the interoperability scenario for Data Cubes, the GIDTT members are planning to do interoperability tests with existing Data Cubes initiatives. To understand and test the interoperability scenario for Data Cubes, the GIDTT members are planning to do interoperability tests with existing Data Cubes initiatives.

Capacity development on Data Cubes – Long term sustainability

It is critical to mention that countries who have started the establishment of Data Cubes expressed their needs to enhance the capacity development to address long-term sustainability and national sovereignty issues.

Executable Notebooks

Beside Data Cubes, the subgroup has also been discussing that the GEO Infrastructure should be the place to provide access to the variety of Executable Notebooks differentiated by thematic areas as the following example:

- TrainHub Portal (eumetsat.int)

Executable notebooks (e.g., R Markdown, Jupyter Notebook), have the capacity to merge narrative (e.g., text, images, videos) with programming language code (e.g., Python, R, JavaScript), which enhance the reproducibility and usability of EO Applications. Therefore, executable notebooks represent an important component of accessing and analyzing EO.

OGC testbed 16 Engineering Report: Earth Observations Application Packages with Jupyter Notebooks provides lessons learned on the use of Jupyter Notebooks for EO. (https://docs.ogc.org/per/20-035.html).

Services

GEO Infrastructure should also be the place where users will discover and access Analysis Ready Data (based on emerging OGC/ISO ARD standards informed by CEOS-ARD efforts). and EO Services by connecting existing efforts (e.g., Copernicus Space Data Ecosystem platform, STAC Index)

Enhance the discoverability and accessibility of the CEOS satellite data

The subgroup has also been discussing the importance of making existing space based data missions more discoverable and accessible and therefore enhancing their usability.

The main source for the space-based data is CEOS Database accessible through the following links:

https://database.eohandbook.com/database/instrumenttable.aspx

https://search.earthdata.nasa.gov/search?portal=idn&ac=true&lat=-0.0703125

Defined as the Gateway to the world of Earth Science data. The CEOS IDN is an international effort developed to assist researchers in locating information on available datasets.

To respond to the user driven approach, the subgroup consider that the Space based data offered through the CEOS Database should be discoverable through a thematic areas-oriented interface allowing the user to search for open Space based data that are best fit for use for the specific usage.

Cloud providers

GEO has been collaborating with specific Cloud providers specifically for the cloud credits programme. Cloud Providers represent critical and essential actors in the Earth Observations domain. None of the Cloud Provider is officially part, yet, of the GEO Membership Community (Members, Participating Organizations, GEO Associates) and therefore could not be an integral part of the GIDTT discussion.

The GIDTT is working to define the landscape of existing Cloud Providers dealing with Earth Observations, with specific attention to the Regional inclusiveness, available here.

Following this landscape analysis, GIDTT has been working to describe and define requirements, description of services and functionalities that the GEO Community would need and benefit from a partnership with Cloud Providers. The document will constitute the basis for official engagement with cloud providers and for inviting them to take part in the GEO Cloud Providers consortium.

The importance of geographical inclusion related to the Cloud Providers consortium has a crucial role for their collaboration with Regional and National GEOs.

In situ data and Regional GEOs

Following the key findings from by the Expert Advisor group on in-situ data, specifically

a. Reduced public investment in in-situ data collection.

b. Lack of open data policies in many countries;

c. Time-limited nature of funding for in-situ data collection in many projects;

d. Lack of adoption of the FAIR principles, and limited availability of in-situ data as part of the published research e.g journals.

e. Lack of the in situ data networks global collections (From e-shape findings)

The subgroup has been focusing on analyzing the existing in situ global and regional landscape to address and explore essential functionalities needed for the revised GEO Infrastructure.

Enhancing discoverability and accessibility of in situ data

Many of the GEO Participating Organizations are also in situ data providers and they represent part of the existing in situ data networks not yet discoverable and accessible by the actual GEOSS Platform. So, similarly for the landscape analysis done for the existing Data cubes on the space-based data, the GIDTT has started to do research of the existing in situ data networks, organized

by domain and geographical location, relation to GEO, Standards and GEO Interoperability arrangements.

This list, In Situ Data Networks.xlsx,can be considered a living document. It is not yet exhaustive and constitutes the basis for future actions to enhance the discoverability and/or accessibility of existing in situ data networks. It is advised by the GIDTT that the GEO Infrastructure should also valorize already existing in situ data networks provided by the authoritative actors, some of which are already within the GEO Membership (Participating Organizations), for example GCOS (e.g., Data sources from GCOS) and WMO.

The barriers to integration into the GEO Infrastructure need to be reviewed and understood not only from the GEO perspective but also as perceived by key data providers that could potentially be integrated into the GEO Infrastructure e.g data infrastructures These different perspectives on the barriers to integration can potentially provide a roadmap for connecting requirements for data, as identified through the G-Reqs database, with the relevant data providers. By lowering the barriers for in situ data, the GEO Infrastructure can clearly show benefits of enhanced data sharing for users and data providers.

Within the GEO Infrastructure Regional and National GEOs will play an integrated role with respect to in situ data sharing and hosting, to be further defined within the evolution of the GEO Infrastructure.

Assessing in situ data brokered by the GEOSS Platform

The subgroup has considered it essential to perform an initial review of the existing in situ data providers already discoverable and accessible by the actual GEOSS Platform.

An initial inventory of such providers is available here In Situ Data Networks.xlsx (Tab GEOSS Portal Resources).

This initial review has highlighted an important gap related to the metadata to support an appropriate typology for properly identifying in situ data that is searchable and discoverable via the GEOSS Platform and suggested that it is essential to reconcile existing and future providers to improve their discoverability and accessibility.

Data format and metadata harmonization

- The subgroup has identified an essential action with respect to the adherence to the FAIR principles by the in-situ data providers: It is critical that actual and future in situ data providers perform a FAIR self-assessment.
- This initial analysis would maximize discoverability and accessibility and establish interoperability making in situ data Analysis Ready Data like space-based data.
- The subgroup has built over previous work achieved by the GO FAIR and enhanced the FAIR analysis matrix available here
- The FAIR analysis can benefit from crosswalk work developed by the Data Sharing and Data Management Principles subgroup of the Data working group, which allows easy comparison

between the GEO Data Management Principles, as well as the wider FAIR, TRUST, and CARE principles.

Hosting of in situ data

Nowadays there are many short-term funded projects that produce in situ data. These in situ data often at the end of the project are not valued, discoverable and accessible, but lost. These projects should have the choice of either hosting their newly produced in situ data in domain specific repositories or if not, obvious repository available GEO Infrastructure could provide the long-term hosting solution. This action can confirm GEO's commitment to Open EO Data.

GIDTT considers that the GEO Infrastructure should be playing a role in minimizing in situ data lost by either exploring hosting functionalities and/or partnering with domain specific regional/national repositories.

The GEO Infrastructure should be playing the central role for in situ data across the various thematic and geographic based repositories.

GEO Work Programme Activities needs and requirements

GEO has the necessity to engage with the user community represented by the GEO Work Programme activities members with the aim of knowing which in situ datasets they require. While the informal dialogue with GWP activities is regular, sometimes it is not that easy to extract their requirements in a structured way that is easy to compare and evaluate. To cover this gap, the EEA has funded the creation of a geospatial in-situ requirements system, called G-reqs.

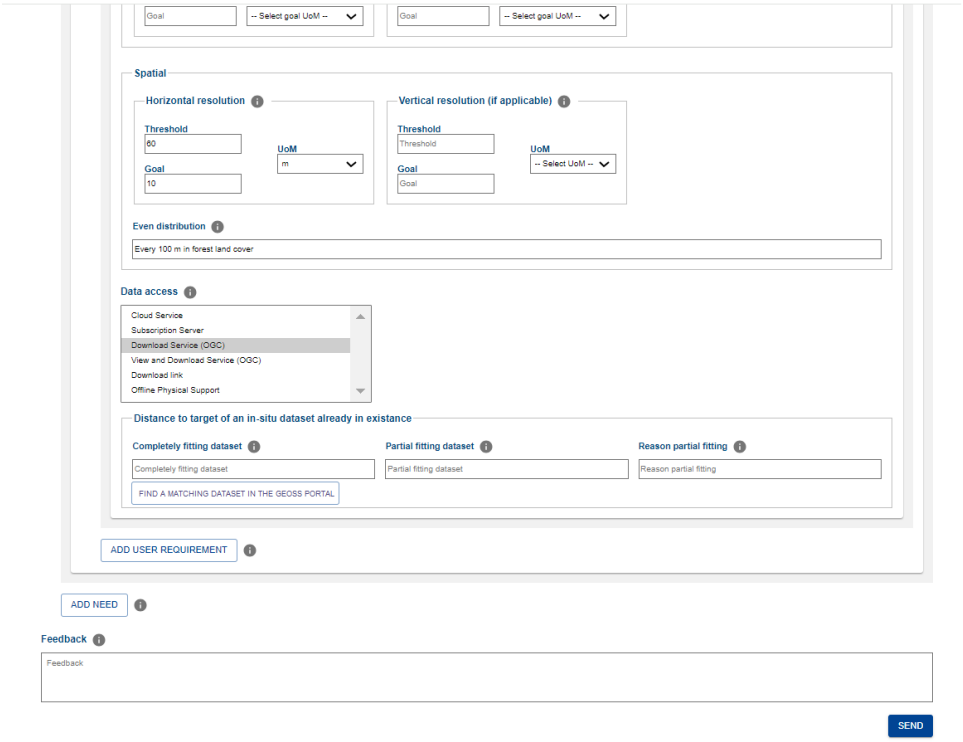
G-reqs provides a FAIR database that can be browsed and analyzed, available at: https://discomap.eea.europa.eu/EuroGEO/dev/

The GIDTT proposes some ways of integration between the G-reqs and the current GEO Infrastructure. The proposed ways can be fully or gradually implemented.

A simplest approach consists in giving access to the G-reqs from the GEOSS portal as part of the main menu offerings, as seen in the following mock-up:

The requirements gathering part of the G-reqs could also be recommended in the results page bottom part, as seen in the following mock-up:

In the opposite direction, a user documenting requirement in G-reqs (https://discomap.eea.europa.eu/EuroGEO/dev/#/form) could have the possibility to verify if some matching dataset is available from the GEOSS portal. This can be implemented with a dedicated button in the "Distance to target" section of the requirements forms, as seen in the following mock-up:



The GEO Infrastructure should play the central role in gathering requirements for in-situ datasets.

User Interface and domain-oriented approach

Similarly to space-based data, in situ data should be integrated and made available using the same thematic areas, to facilitate the user experience in discovering and/or accessing EO data for a specific purpose, this data being space based or in situ.

Quality Check – Integration of GEOSS Platform and GEO Knowledge Hub, and Governance

Governance

Following the recommendations of the Expert Advisory Group, the GIDTT has been discussing on the crucial importance of having a GEO Infrastructure governance document that provides the definition of the functions and form of the services to be provided as well as the necessary set of actors, beneficiaries, resources, timelines, and coordination mechanism to execute the task.

- Initial discussions related to the Governance document are captured in the following document

Data Stewardship and quality control

The GIDTT also discussed the governance document as the basis for data stewardship and quality control in the GEO Infrastructure. For instance, regarding data stewardship, the governance must define the rules that enable it. For quality control, all the aspects should also be specified, for instance

1) What are the quality requirements (e.g., data accessibility and discoverability)
2) What are the procedures used to quality check
3) What are the rules applied when quality does not fill the requirements from the GEO Infrastructure

Quality Check of EO Derived data

The GEO Infrastructure, playing a pivotal and central role for the entire GEO Community for the discovery, access and use of EO data and applications, will have to address the quality aspect of the content available in the infrastructure.

Data and Knowledge Providers, being the only owners of the resources shared, are called to provide high quality fully documented resources, compliant with Data Management Principles and specification of the fit for use aspect.

On the other side the GEO Infrastructure will use automatic mechanisms to check their accessibility through the existing links.

The GEO Infrastructure will leverage two existing tools to check the links of data and knowledge resources available to make sure that there are accessible. These tools are the Status Checker and the Link Checker.

Regarding the usability, the actual status of the content of the GEOSS Platform reveals the urgency to check the quality of its content. The following actions have been discussed

1) Discoverability and Accessibility of current data providers
    a. Categorize them by Space based Data, in situ data and EO Derived
    b. Reconcile the Data Providers with the Yellow Pages
    c. Ensure a full reliable list of Data Providers with authoritative and trusted data
    d. Validate the discoverability and accessibility of existing EO Derived data available via the GEOSS Platform.

Integration of the GEOSS Platform and the GEO Knowledge Hub

At the moment there are two components of the GEO Infrastructure that are not yet integrated, and this constitutes an element of confusion for the end user.

The two modules are: GEOSS Platform for EO Data and the GEO Knowledge Hub for EO Knowledge, which need to be integrated to provide a unified service for the end user.

As part of this discussion, it was identified that both modules have REST APIs, which enable machine-to-machine communication via HTTP operations. This capacity from both modules can be explored as the base for their integration.

GEO Infrastructure User Interface

GIDTT agrees that the GEO Infrastructure will have to provide a user friendly, efficient and reliable service to its end users, and besides the quality check and curation of its data and knowledge resources, the User Interface plays a very important role.

The user interface should guide the user to discover, access data and knowledge resources in a few steps, using the 5 W approach, as shown in the picture below.

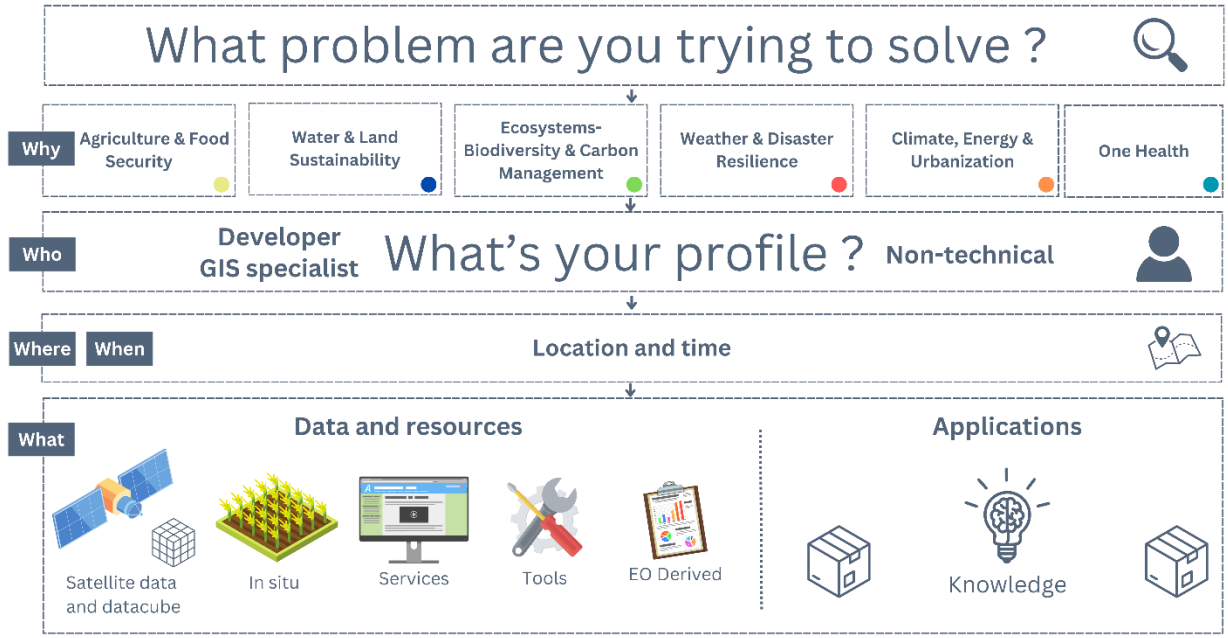


Figure. 4 Simplified schema of the user-oriented GEO Infrastructure

Thematic areas will be the entry point for the user to discover and access data and knowledge resources.

To define User Interface, the GIDTT will work on a few (3) real scenarios identified within the GEO community.

Reference documents and presentations

- **GIDTT Attendance document:** GIDTT - Attendees.xlsx
- **EAG 2022**: https://geosecretariatgeneva.sharepoint.com/:b:/s/GEOSecretariat/Ed6V9hIVvaNDjynIDvbD750B4NNCK_zO5JUQDOiYQYnpmg?e=vQM2Oh
- **EXCOM March 2023:** ExCom 60_Session 2.2_EAG_followup_v2.pptx
- **EXCOM July 2023:** ExCom 61_3.3_GIDTT TORS.pptx
- **GIDTT TORs:** ExCom61-12_GEO Infrastructure Development Task Team Terms of Reference_310823.docx