GEO Work Programme Activities Data Practices and Needs Analysis Report

This document is submitted by the Data Working Group, the GEOSS Platform Team, and the GEO Secretariat to the Programme Board for discussion.

1. INTRODUCTION

The GEO Work Programme (GWP) is the primary coordination and planning instrument used by GEO to achieve its strategic mission. It is a multi-year plan accepted by the Plenary. The GWP brings together activities (Flagships, Initiatives and Community Activities/Pilot Initiatives[1], all indistinctly later referred to as GWP activities) as well as Foundational Tasks that are respectively coordinated by members of the GEO community and the GEO Secretariat.

One key aspect of the work carried out within GWP activities is the transformation of Earth observations (EO) data into actionable information to be used by GEO Members. This includes collecting, processing, curating, and sharing data as well as products and services in domain specific areas for defined end-users.

This report provides a status report on some of the GWP activities data practices and needs, conducted during the October 2021 to November 2022 period, initially with an online survey that was filled by 22 activities, followed by rounds of consultation with 13 individual activities in dedicated "engagement" calls.

These consultations provided an in-depth understanding of how the GWP activities participants deal with matters related to in situ data, their awareness and implementation of the GEO Data Sharing and Management Principles, practices linked to law and data policies, and finally, their needs regarding the GEOSS Infrastructure, focused on the GEOSS Platform and the GEO Knowledge Hub (GKH).

In the transition period between the 2020-2022 and the 2023-2025 phases of the Work Programme, as well as during the gathering of the Expert Advisory Group on GEOSS (EAG), this report brings attention to technical, legal, and financial challenges related to data and services encountered by the GWP activities to be considered by the GEO leadership.

2. METHODOLOGY

The consultation phase took place during the GWP 2020-2022 implementation period, but some considerations and projects also have potential implications for the 2023-2025 period, since such considerations might have induced some responses that differ from the original implementation plans that were accepted by the XVI Plenary in 2019. The status of each activity is set as they are listed in the 2020-2022 GWP.

[1] The terminology "Community Activity" is referred in the 2020-2022 Work Programme, where the terminology "Pilot Initiative" refers to the 2023-2025 Work Programme.

In its 2020-2022 edition, the GWP consists of 4 Flagships (F), 20 Initiatives (I) and 36 Community activities (CA). These numbers evolved to 5 Flagships, 19 initiatives and 20 Pilot Initiatives in the 2023-2025. The findings described in this report were gathered either via an online survey or via individual calls with the leaders and/or with the technical lead person of each activity.

In total, 28 distinct activities provided inputs to this report, either via the survey, or via individual calls, or both. Annex A1 provides the list of the activities.

2.1 Survey

Using the EUSurvey open-source tool, and with support from the European Environment Agency (EEA), an online form was sent to all GWP activities in November 2021. It comprised 6 sections, related to (i) information on GWP activities, (ii) input data and services, (iii) challenges related to input data, (iv) output data and services, (v) general questions related to your activities and relevant output data, and (vi) questions related to the implementation of GEO Data Management principles.

To capture the variety of used input and output data produced, GWP activities were asked to upload a table listing all datasets they use and produce. For each dataset, questions were asked about discoverability, accessibility through web services, the possibility to directly download it or only upon request, its availability as open data and the attached license, the possibility to discover and access it using the GEOSS Platform and whether it has a Digital Object Identifier (DOI) or not. Beyond data, similar questions were asked about output services and applications. Additional questions were asked about the interest to contribute and make the produced applications available in the GEO Knowledge Hub.

2.2 Engagement calls

Based on the initial survey results, and with the objective to gather more detailed insights about the needs and practices of individual activities, 13 engagement calls were conducted with individual GWP activities. Selection of the activities was initially done at their request to follow up on the survey, and was then extended by prioritizing some of them, given their strong implication in relevant topics, and their availability to engage in the process.

These 90 to 120 minutes-long calls allowed the team to ask detailed questions about the key in situ data providers and networks, their practices related to implementing the GEO Data Sharing and Data Management Principles, aspects covering law and policy, and how they are leveraging the GEOSS Infrastructure components, mainly the GEOSS Platform and the GEO Knowledge Hub (the latter started to operate in July 2021).

The same base structure was used for all calls, allowing to compare answers between GWP activities, with additional questions when estimated necessary. After each call, minutes were written and sent to the activities for validation. A summary matrix was also created, listing all answers to each question, in order to allow for intercomparison.

An analysis of the survey results and the engagement calls is provided in Annex A2.

3 PRELIMINARY FINDINGS

3.1 General findings

While some of the GWP activities have a clear mandate from international conventions to provide timely, relevant, and actionable information from EO, many of them struggle to brand themselves as authoritative consortia to perform such as task.

This results in limited authority to interact with data providers from GEO Members and request for more data openness, and creates a disaggregated knowledge of observation gaps, limited to sectorial activities. High level advocacy and accountability of GEO Members should therefore be increased based on the GWP activities needs to access input data and disseminate their outputs. Moreover, the current way GEO interacts with its Members strongly limits the ability to address cases where data is provided by another institution than the one responsible for liaising with GEO. The lack of national coordination poses barriers in terms of putting the open data in practice, where a ministry does not even know about GEO and the benefits of open data and has no awareness about its data being possibly used for EO.

Despite data gaps, GWP activities often deal with what they can find as open data but could significantly improve their products and services should those gaps be filled. Increasing data sharing and filling gaps would require incentives that they are however not able to provide themselves.

3.2 In Situ Data

Whether it comes from automatic sensors, field surveys or from different sources that fall under the in situ terminology (data that is collected in its immediate or surrounding environment in opposition to remote sensing), the survey and engagement calls highlighted the complexity of GWP activities in accessing data with sufficient coverage (temporal and geographic). Sustainability of observations is of prime concern for many activities. As data consumers, they rely on what is available and the quality of what they produce is often at risk of diminishing because of insufficient funding of observing networks.

Challenges related to usability mainly relates to the input data format, that is not originally provided in a way that allows immediate use and requires time consuming pre-processing and harmonization efforts. This lack of usability is not only found in the technical aspects but also in legal considerations about use and dissemination of products derived from data.

Many activities also face the challenge of disseminating in situ data, either if the data is collected within the scope of the activity, or as an aggregated product compiling many sources. A common challenge relates to having access to an infrastructure to store and share such data. This is typically the case for domains where in situ data is collected by people on the ground, with or without a common methodology.

Much effort is already put in the data curation steps, leaving little resources to establish and maintain an infrastructure for the data. In this regard, support from a common place, under the neutral GEO umbrella and branding could prove useful.

Thematic domains where in situ data is critically needed was identified both in the survey and in the engagement calls and concern mainly: (i) weather data from automatic sensors (including hydrology and atmospheric measurements,), temperature and precipitation (ii) biodiversity data from field observations, (iii) terrestrial field observations on land use and land cover, mainly for agriculture, biomass, forestry and land degradation monitoring purposes.

3.3 Data Sharing and Data Management Principles

Findings from the Survey as well from the Engagement calls revealed the need of the GEO Work Programme Activities to receive additional guidance on learning and implementing the Data Sharing and Data Management Principles. At this purpose the Data Sharing and Data Management Principles Subgroup has focused over the 2022 period in establishing tools to meet the GWP needs. The tools that have been developed are:

- The DMPs Implementation Guidelines
- The Data Sharing and Data Management principles Dialogue Series
- The Data Management Principles Self-Assessment Tool.

The implementation guides for both the GEO Data Sharing and Data Management Principles have been reviewed by the GEO DWG to improve their currency and relevance to the GEO community and other communities that could benefit from their content. The review of the GEO Data Management Principles Implementation Guide has been completed and the resulting revisions have been incorporated into the guide. The revised guide has been published in the GEO Knowledge Hub, as described within that section of this report.

3.4 Law and Policy

Findings from the Survey and from the engagement calls revealed that the GWP activities need support in better understanding of the Open data licenses and related implications of their usage.

At this purpose the Law and Policy Subgroup has been working on releasing the Open Data Licensing Guidance document.

3.5 Relations with the GEOSS Infrastructure

3.5.1 GEOSS Platform

The GEOSS Platform is the instrument for implementing GEOSS. GEOSS presently links more than 178 open data catalogs and information systems, comprising over 414 million data and information resources. GEOSS, in the last two years, served about 15 thousand unique users finalizing about 150 thousand searches. In this framework The GEOSS Platform is a Data Discovery & Access Ecosystem providing harmonization and transformation built on a single map-based Internet point of access (the GEOSS Portal), providing community customizable GUIs and Views on different resources and exposing Machine to Machine features to directly connect to the GEOSS Platforms functionalities to support developers in implementing new or enhancing functionalities.

From the engagement calls was highlighted the lack of awareness about the GEOSS Platform and its functionalities. A better communication and involvement of the different communities with the GEOSS Platform Team is recommended and taken into consideration within the GPP Project. Another important aspect that came out from the calls was that communities were interested in sharing data and services and make them discoverable and where possible freely accessible through the GEOSS Platform, but they had no information on what was feasible and how to integrate their services/applications and data with the GEOSS Platform. Another aspect is the interest in a dedicated Community Portal to be customized to the community needs.

For these reasons the GEOSS Platform team will engage again the engaged communities in order to identify technical issues and solutions and to investigate and advocate the communities on how to integrate desired services/application to be interfaced with the GEOSS Platform.

One important aspect to take into consideration within the GWP is that some of the communities don't have budget or expertise to allocate to the integration of the services or to the implementation of the portal or for the data storage, and this is a limiting aspect of the GEOSS Infrastructure to be addressed.

3.5.2 GEO Knowledge Hub

In the survey, 20 out of 22 GWP activities responded positively when asked if they were interested in sharing their applications/services within the GKH. In 2022, progress has been made with integration of almost 500 knowledge resources and involved many GWP activities.

The GEO Knowledge Hub offers capabilities for the GEO community to share resources that can contribute to understanding of relevant resources and issues. The GEO Data Management Principles Implementation Guide, which was recently revised by the GEO DWG, has been published on the GKH, enabling interested stakeholders to familiarize themselves with these Principles and how they can be adopted to improve the stewardship of data and other information resources of interest to the community. Similarly, the Dialogue Series and the DMPS Self-Assessment tool are available in the GEO Knowledge Hub.

3.6 Output Data and Services

From the indicated outputs data and services, dissemination as fully open products are still made difficult because of restrictions imposed by partners, lack of resources to make these available and maintenance of an infrastructure to support it.

The way outputs are made accessible is often minimal and could be enhanced, for instance by using more APIs and facilitating direct access to data rather than only through a web portal and download of the data. Knowledge exchange on best practices within the community would also need to be facilitated and could increase the capacity of GWP activities to disseminate their products through standards (e.g. OGC WFS, WMS, and STAC as it is becoming more and more widely used) in order to stay competitive in the EO landscape.

4 RECOMMENDATIONS

A. **Account for high-level advocacy on open data implementation.** GWP activities struggle to have access to enough data to fulfill their objectives, despite GEO being committed to provide open data for everyone. In the run up to the 2023 Ministerial Summit, Programme Board an ExCom Members could set the example by announcing progress of opening new datasets concerning priority areas as indicated in section 3.2.
B. **Strengthen national coordination, especially for in situ data sharing.** The lack of awareness about the GEO data principles between different national agencies often prove to be a difficulty when trying to access data. Point of contacts in every relevant agency of a GEO Member could help advance this issue and not be only limited to one or two persons. The Data Working Group could establish a process to interact with GEO Members and advocate the data openness critical to advance the work of the GWP activities.
C. **Reinforce technical support to GWP.** Many activities would benefit from enhancements to the current infrastructure offered by GEO, mainly for what regards in situ data storage (currently not supported) and sharing their output data products.

D. **Increase technical knowledge sharing for the benefit of GWP activities.** Resources mentioned in this document are being developed by the Data Working Group and members of the community, but they need to be disseminated more widely. Examples include the Dialogue Series, guidelines on implementing the GEO Data Management Principles, and licensing guidelines.
E. **Connect GWP activities with data providers**. In order to needs to have data providers being involved more closely with those who use their data, including GWP activities.
F. **Events around Open Data and Open Knowledge:** GWP Activities reported on the importance of re-establishing technical events such as the Data providers workshops and Open Data and Open Knowledge workshop.

Annex A1: Table of consulted GWP activities

Table 1: Summary of GWP activities and how each of them was consulted.

	Activity name	Acronym	Status	Survey	Engagement call
1	Arctic GEOSS	ARCTIC-GEOSS	CA	Yes	No
2	Digital Earth Africa	DE-AFRICA	I	Yes	No
3	Digital Earth Pacific	DE-PACIFIC	CA	Yes	No
4	Earth Observations for Health	EO4Health	I	Yes	Yes
5	Earth Observations for the Atlantic Region	ATLANTIC-EO	CA	Yes	No
6	Enhancing food security in African AgriCULTUral Systems with the support of Remote Sensing	AFRICULTURES	CA	Yes	No
7	Forest Biomass Reference System from Tree-by-Tree Inventory Data	GEO-TREES	CA	Yes	Yes
8	GEO Global Water Sustainability[2]	GEOGLOWS	I	Yes	Yes
9	GEO Land Degradation Neutrality	GEO-LDN	I	Yes	No
10	GEO Vision for Energy	GEO-VENER	I	Yes	Yes
11	Global Agricultural Drought Monitoring	AGRI-DROUGHT	CA	Yes	No
12	Global Crop Pest and Disease Habitat Monitoring and Risk Forecasting	CROP-PEST-MONITORING	CA	Yes	Yes
13	Global Drought Information System	GDIS	I	Yes	No
14	Global Ecosystems and Environment Observation Analysis Research Cooperation	GEOARC	CA	Yes	No
15	Global Network for Observations and	GEO-MOUNTAINS	I	Yes	Yes

[2] The data coming from this activity only concerns the "Transforming Water, Weather and Climate Information through In Situ Observations for Geo-Services in Africa (TWIGA) project, which is participating in GEOGLOWS.

	Information in Mountain Environment				
16	Global Observation System for Mercury	GOS4M	F	Yes	Yes
17	Global Wildfire Information System	GWIS	I	Yes	No
18	In-Situ Observations and Applications for Ecosystems Status of China and Central Asia	IN-SITU-ESC	CA	Yes	No
19	Night-Time Light Remote Sensing for Sustainable Development Goals	NIGHT-LIGHT	CA	Yes	No
20	Space and Security	SPACE-SECURITY	CA	Yes	No
21	The International Grand Global Ensemble	TIGGE	CA	Yes	No
22	Urban Heritage Climate Observatory	UHCO	CA	Yes	No
23	GEO Global Agriculture Monitoring	GEOGLAM	F	No	Yes
24	AquaWatch[3]	AQUAWATCH	I	No	Yes
25	Oceans and Society: Blue Planet[3]	BLUE-PLANET	I	No	Yes
26	Global Observation System for Persistent Organic Pollutants	GOS4POPS	I	No	Yes
27	Global Forest Observation Initiative	GFOI	F	No	Yes
28	Geohazard Supersites and Natural Laboratories	GSNL	I	No	Yes

[3] The engagement call with AquaWatch was done jointly with Blue Planet.

Annex A2: Survey and Engagement Calls Results Summary

A dashboard summarizing survey responses was created with support from the EEA and may be accessed here: Survey Results Dashboard

Input data

From survey responses, 258 input datasets were mentioned (some of them with redundancy between several GWP activities). 123 were listed as satellite, 43 as in situ, 67 as other (e.g. derived from in situ and satellite) and 8 as statistical (e.g. socioeconomic). 17 were not specified.

About the input data used by the GWP activities, mentioned providers are mainly the USA (61 occurrences), China (59), the EC (36), and South Africa (11), followed by European Members like France, Germany, Spain and Belgium, and European organizaions like ESA, EUMETSAT and ECMWF. Other mentioned GEO Members include Ghana, Japan, Kenya, Mexico, Norway the UK, Vietnam. International organizations like FAO, IIASA, the World Bank, the Pacific Community are mentioned. A couple of private sector companies such as Google and ESRI are also listed.

Overall, 83% of input datasets are listed as discoverable, and 74% are available through webservices. On the accessibility of the input data, 49% are listed as directly downloadable, where 18% are available upon request and 33% did not give an answer. When asked about openness, 70% of the listed input datasets are made available as open data, 12% were not, 4% responses were "I don't know" and 14% did not give an answer.

The overall discoverability through the GEOSS Platform is 52% yes, 23% no, 8% of "I don't know" and 17% where no answers were given. Accessibility through the GEOSS Platform is 40% yes, 33% no, 10% of "I don't know" and 17% where no answers were given. Please note that "discoverability" means the dataset can be searched through the GEOSS Platform, where accessibility means that the user can access the original dataset location using the GEOSS Platform. When asked if the input data has a Digital Object Identifier (DOI), only 13% of overall datasets were indicated to have one.

For what regards in situ only datasets, proportions are similar about general discoverability (84% yes), with slightly lower availability through webservices (60%) and only about 45% of availability as open data. Discoverability and accessibility through the GEOSS Platform are significantly lower, with respectively 28 %and 5% positive answer (against 52% and 40% for all data types). 37% of in situ datasets have a DOI.

On the barriers faced by GWP activities, they mainly relate to licensing of input datasets that is too restrictive, and about the coverage that is not sufficient and inadequate timeliness. Survey responses about the type of datasets that were critically needed to reduce to barriers are mostly linked with weather and climate data (precipitation, streamflow, atmospheric, hydrology).

On the role of GEO to help fill the existing data gaps, survey respondents were mostly in favour of increasing awareness raising and communication activities (13 and 12 votes respectively). 13 were also in favor of more standards and guidelines development, and 12 voted for more support to projects funding proposals. The proposal that GEO acts as a data custodian was supported by 4 votes in the survey, but was supported during discussions in the engagement calls.

Engagement calls allowed to provide some clarifications with regards to how GWP activities access data, especially in situ. In many cases, it was highlighted that many of them use any input data they can find that fits their needs and has sufficient quality. Responses from the survey only

highlighted the data that was already being used, but significant gaps outside Europe and North America remain in terms of in situ observations.

Output data

68 output data products were provided in the survey results, from which 59% are discoverable and 57% are accessible through web services. The same proportion (57%) of outputs are directly downloadable, with 6% available upon request only and 37% where no answer was given. Overall, only 61% of outputs are made available as open data, 19% are not and 19% did not give an answer.

The connection to the GEOSS Platform is quite limited, with only 15% and 12% of outputs that are respectively discoverable and accessible. 37% of outputs are not discoverable and 40% are not accessible through the Platform. Interestingly, almost half of the answers were either not given (34%), or the respondent did not know (15%) for both questions. In similar proportions as input data, only 13% of output data have a DOI.

Some GWP activities produce harmonized, global datasets as output, that aggregate different sources of data. Some of them face challenges to share these outputs in a fashionable standard way and would benefit from support by GEO. This could be done by developing guidelines and examples on how to share via web services for instance.

GEO could also provide technical support to store in situ data where GWP activities have the expertise and are well placed in the community to coordinate future data sharing efforts. This could enhance the availability of such data including via the GEOSS Platform through standards and APIs.

Output services (and applications)

51 services were listed from the survey results, from which 92% are described as discoverable, including 82% from the GEOSS Platform. Only 29% are however freely accessible (either because of an access restriction given to certain users or because they are not free).

Data Sharing and Data Management principles implementation

Annex A3: Supplementary material

The complete minutes from each engagement call may be accessed with the following links:

1. GEOGLOWS (TWIGA) – 4th February 2022
2. EO4Health – 28th February 2022
3. CROP-PEST-MONITORING – 1st March 2022
4. GEO-TREES – 8th March 2022
5. GOS4M – 22nd March 2022
6. GEOGLAM – 13th June 2022
7. AQUAWATCH & BLUE-PLANET – 16th June 2022
8. GEO-VENER – 17th June 2022
9. GEO-MOUNTAINS – 24th June 2022
10. GOS4POPS – 14th September 2022
11. GFOI – 3rd October 2022
12. GSNL – 30th November 2022

The matrix summarizing all answers from all engagement calls in a single place may be accessed here: Data Analysis Survey Engagement Calls summary matrix.xlsx

A table providing cleaned information about input data, output data and services, and the barriers identified by GWP activities from the survey may be accessed here: Data Analysis Survey Input, Outputs and Services Summary.xlsx

Raw inputs from the online survey may be accessed here: Raw Survey Results