GEO In-Situ Data Strategy Post-2025

This document is presented to the Programme Board for discussion

EXECUTIVE SUMMARY

GEO's Members, Participating Organizations and Associates work together to develop and implement Earth Observation (EO) projects that help decision-makers understand and act on global environmental and societal challenges such as climate change, biodiversity loss, and pollution. EO data, both in-situ and satellite-based, are fundamental components of these actions and essential for GEO to achieve its mission of Earth Intelligence for All.

This document responds to the need for an In-Situ Data Strategy that has been acknowledged within successive GEO Ministerial Declarations e.g. Canberra Declaration (2019). Building on a significant body of work by GEO Members (especially those supporting the In-situ Data subgroup and the Data & Knowledge Working Group) spanning several years, the In-Situ Data Strategy aims to provide a framework for GEO that addresses the specific challenges associated with the acquisition and use of in-situ measurements and observations. It also provides a set of actions and potential roles that will enhance integration of all forms of EO data to support GEO's ongoing mission.

1. INTRODUCTION

Purpose

In-situ data refers to Earth observations and measurements collected directly at or near the Earth's surface, as opposed to remote sensing (RS) data that is gathered from spaced-based systems e.g. satellites. In-situ data encompasses, but is not limited to, observations and measurements obtained from ground-based monitoring systems, maritime platforms, airborne sensors (excluding satellites), field surveys and citizen science (Strobl, 2024).).

In contrast to satellite-based observations, in-situ data are usually direct measurements of selected variables, often in specific and fixed locations, which are likely to be more precise and therefore considered as the "ground truth". Satellite-based observations and measurements tend to be more indirect but systematically cover the Earth's surface.

Together, in-situ and RS data are complementary, providing a comprehensive suite of observations and measurements of the planet. For this reason, both satellite and in-situ Earth observations are equally important in order for GEO to deliver its stated mission, especially with respect to data, knowledge and Earth intelligence for all.

The value of in-situ data is substantial because it:

- is essential for the calibration and validation of remote sensing products and Earth system models (including training Artificial Intelligence and as input for Machine Learning modules).
- provides measurements in difficult to access locations e.g. within the global oceans, in areas affected by clouds, smoke or aerosols, or other areas that are not accessible to space-based observing systems.
- often measure significant variables on the Earth surface that are directly impacted by human activities.
- has a temporal and spatial distribution that allows more accurate and frequent measurements. In-situ measurements tend to be tied to a specific location and taken at a greater frequency, which provides higher resolution data than can be achieved by space-based systems.
- potentially provides long time series based on repeated historical measurements that can, in some cases, be over a period of more than 100 years.
- is the only legally admissible form of Earth observation data that can be used to prove illegal activities in some Member States e.g. destruction of protected natural habitats.

Due to the diversity and nature of the various in-situ data acquisition activities spread across a wide range of different disciplines, e.g. hydrology, climatology, oceanography, biodiversity etc., the current associated data landscape is highly heterogeneous. Despite some limited efforts to standardize monitoring protocols and aggregate in-situ datasets to generate a more general coverage, the landscape remains fragmented in many sectors with significant gaps in some areas. A

situation that is further exacerbated by inconsistent data policies at the national and local level, including those related to data archiving, preservation and reuse, which lead to loss of critical data and knowledge.

In-situ data and remote sensing data require different data management approaches due to their distinct characteristics, especially those associated with the volume, variety and velocity of data acquisition (Vaduval, 2020). The key challenge for satellite-based observations is the large volumes of data that require specific storage and processing solutions. In contrast, the variety of in-situ measurements produced by a wide range of sensors generating different types of data in varying formats, units, and scales creates a very heterogeneous landscape. For example, compiling a global dataset of ground temperature involves a huge amount of data providers that are utilizing various sensors to generate data at potentially different temporal and spatial resolutions and in inconsistent units.

The diverse nature of in-situ data necessitates significant investment of resources in its curation and archiving to ensure its usability and accuracy for specific user applications. Significant data management efforts in terms of standardization, harmonization, and interoperability are also critical to effectively consolidate different datasets to fulfil the needs of users.

Achieving the necessary level of in-situ data integration requires new and existing in-situ data systems and networks to implement common data frameworks, metadata standards, and quality control processes to ensure consistency and reliability across different sources. This should include encouraging wider adoption of recognised data management practices, e.g. the FAIR, CARE and TRUST Principles, as well as the GEO Data Management and Data Sharing Principles and other relevant best practices, including those associated with consistent data licensing.

Context

The specific challenges associated with the in-situ data landscape such as its heterogeneity, lack of coordination, geographical variations, large and diverse range of stakeholders, and barriers to data access, have been recognized in successive Declarations from GEO Ministerial Summits (Cape Town 2007, Beijing 2010, Geneva 2014, Mexico City 2015). Reinforcing this message, the Canberra Declaration (November 2019) states that the Ministers:

- Recognise the critical role that data collected from the atmosphere, land, and water (in-situ data) plays in achieving GEO's mission;
- Note that sharing of such data is limited and significant gaps remain for a global observing system;
- Note that such data has specificities in its nature, governance, technical management, financing,
- Call upon the GEO community to develop a strategy to address the challenges in this area and to demonstrate progress in implementation.

GEO's Member governments, Participating Organizations and Associates work together to develop and implement Earth observations projects and initiatives that address global environmental and societal challenges and help decision-makers understand and act on the global challenges of climate change, biodiversity loss, and pollution. To address these challenges and their impacts at different geographical scales, GEO is connected to major thematic global, regional and national stakeholders that share, manage, use, in-situ data such as WMO, EEA, NOAA, FAO, GCOS etc.

GEO has a unique convening power to coordinate and foster best practices for management and sharing of all forms of EO data, which are a fundamental for the creating the synergies necessary for GEO to achieve its mission of Earth Intelligence for All. A critical element of putting into practice is the development of an in-situ data strategy for GEO that responds to the Canberra Declaration (2019) and seeks to address the specific challenges associated with in-situ data, which are identified below.

Challenges associated with in-situ data:

- highly heterogeneous data landscape that lacks coordination in some areas.
- inconsistent spatial and temporal distribution, including critical gaps in some geographical and thematic areas.
- barriers to sharing / reuse caused by variances in data policies across Member States.
- concerns about sharing unvalidated data resulting in low observation data latency.
- lack of adoption of common standards, vocabularies and access protocols.
- available data has inadequate metadata, unsuitable format, lacks quality indicators and / or suitable user licencing.
- absence of adequate long-term storage and preservation provision leading to loss of data, especially from short term projects or research field campaigns.
- sustainability, including consistent funding for key observing systems, leading to uncertainties for ongoing data availability.

Acquisition, management and preservation of in-situ data is significantly different to that required for space-based data. For this reason, GEO requires a targeted and distinct strategy that acknowledges and addresses the specific challenges associated with in-situ data. Furthermore, the highly heterogeneous landscape and diverse range of stakeholders requires that GEO takes a very specific approach to in-situ data that differs from that adopted for space-based Earth observation data.

The GEO strategy should build on the existing in-situ data landscape and the experiences of current activities and initiatives e.g. GEO Working Groups and Work Programme (GWP) activities. Several examples already exist that demonstrate critical importance of in-situ data for GEO initiatives (e.g., GEOGLAM, GEO Mountains, TWIGA-TAHMO, GEO Aqua Watch, and others). Efforts should also be made to engage the wider GEO community, e.g. regional GEOs, Participating Organizations, including private sector partners, that maintain and promote in-situ data networks that are potential data providers in GEO. Many of these initiatives are coordinating existing in-situ data

networks on a regional or global scale, which can provide key insights into the challenges associated with in-situ data systems and how these can potentially be resolved.

Potential roles for GEO within the in-situ data landscape:

- **Coordinator**: actively engaging with relevant stakeholders to encourage greater (re)use of in-situ data, reduce duplication of effort and address specific data gaps. Foster adoption of consistent data management practices including the GEO Data Sharing and Data Management Principles, and implementation of the FAIR, CARE and TRUST Principles.
- **Advocate**: actively seeking opportunities for implementation of new in-situ observing systems and ensuring the sustainability of existing key monitoring networks
- **Capacity builder**: supporting user upskilling to enhance exploitation of in-situ data. GEO tools such as the Knowledge Hub (GKH) can be used to deliver the necessary materials to train and educate users, including the next generation of researchers.
- **Promoter**: highlight availability, quality and value of in-situ data, products, services and systems
- **Facilitator**: initiating community dialogue to address specific issues associated with in-situ data e.g. prioritisation of the creation or continuation of selected in-situ observing systems
- **Custodian**: providing guidance for archiving and preservation of in-situ data, especially "orphan" data that might otherwise be lost.
- **Redistributor**: supporting discovery and access of in-situ data though the GEO infrastructure.

This strategy document provides the outcomes that GEO can seek to achieve in any or all those roles listed above.

2. GOALS AND OBJECTIVES

The **goal** of this strategy is to coordinate in-situ data providers, mobilize GEO Members and Participating Organizations, engage in-situ data users, and advocate for in-situ open knowledge. It aims to foster a change in the mindset of the GEO community by ensuring in-situ data sharing and management is an integral part of all relevant activities and support this approach by delivering the tools needed for in-situ data sharing and integration.

The strategic objectives are as follows:

a. Identify and monitor data users (as defined by the GIDTT second deliverable) requirements for discovering, accessing and re-using in-situ data. Determine the barriers to delivering in-situ data that respect the relevant data management principles e.g. FAIR, CARE etc. as open access.
b. Coordinate with existing national and regional in-situ data networks across different geographical areas and by domains with the aim of:

- Defining and describing the existing in-situ data landscape
- Promoting sustainability and improve observing networks in data sparse regions and focus areas;
- Identifying existing gaps in the observation systems, support data providers that can close those gaps and promote new harmonized observing networks.
- Supporting implementation of a GEO infrastructure that can serve as a gateway for in-situ data and stimulate the adoption of appropriate tools for data management.
- Incorporating novel sources of in-situ data such as from citizen science, Internet of Things (IoT), and drones (UAVs).

c. Advocate for and monitor the implementation of the GEO Data Sharing and Data Management Principles for in-situ data applications at national and regional scale with the goal of:
   - Taking a leading role in promoting good data stewardship (e.g. with the adoption of DSDMP, FAIR, CARE, TRUST principles)
   - Encouraging the use of flexible and common data licenses.
   - Illustrating the use of in-situ for increasing environmental knowledge.
   - Minimizing and avoiding loss of in-situ data, including through adoption of relevant archiving and preservation best practices.

3. OVERALL IMPLEMENTATION APPROACH

The GEO community is in a unique position by having a variety of stakeholders that can effectively implement an in-situ data strategy through existing partners (see Figure 2) and, in particular, via:

- Member States and National GEOs (including official in-situ collections)
- Regional GEOs
- Participating Organizations (including the research infrastructures)
- GEO Associates and private sector organisations (including those associated with the Internet of Things (IoT))

GEO bodies that will be supervising and facilitating the implementation of this strategy will be:

- GEO Data and Knowledge Working Group
- GEO Infrastructure Development Task Team (GIDTT)
- GEO Secretariat

The Post-2025 GEO Work Programme is re-organized into Focus areas, which will also be reflected in the revised GEO Infrastructure as shown in Figure 1 below.

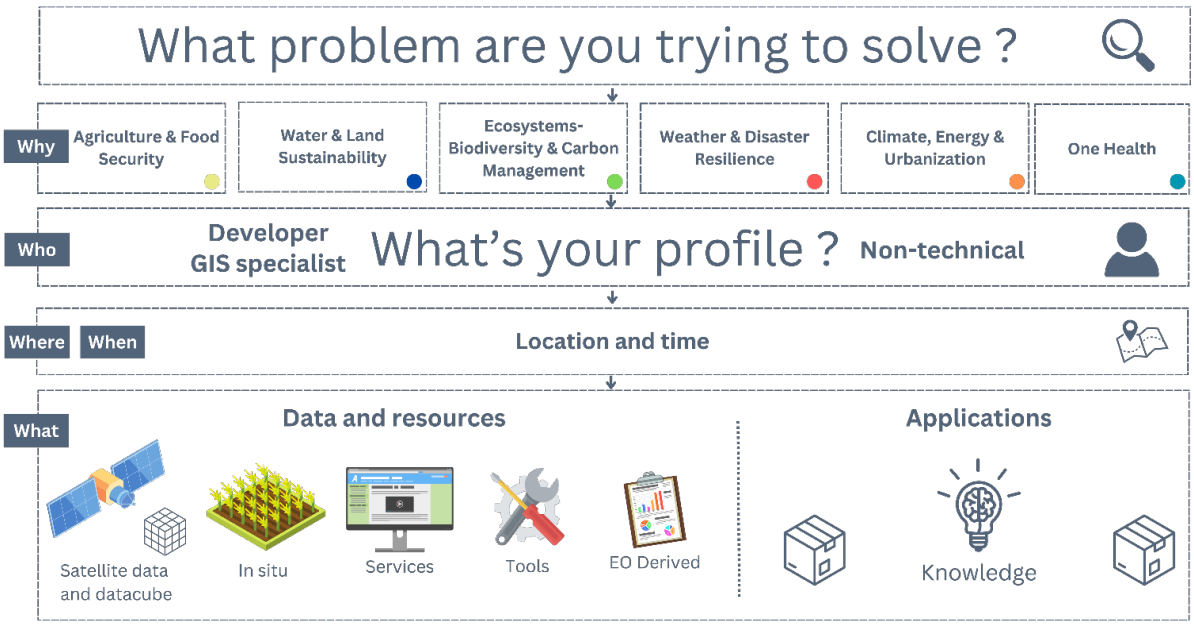

Figure 1: Roadmap for the revised GEO Infrastructure to reflect the GEO Post-2025 Strategy

Implementation of the In-situ Data Strategy will require close collaboration with all GEO community actors including Work Programme Activities, Member countries, POs, Associates, existing research infrastructures and other in-situ data providers, as well as private sector partners. The GEO Infrastructure should also be leveraged as a tool to enhance the discoverability and accessibility of in-situ data and used to identify any critical gaps in essential observing systems.

This approach will require definition of:

- Requirements
  - Identification of user needs (starting with the GEO activities), which also demonstrate the inherent value of in-situ data;
  - Recognition of gaps in-situ data that are not covered by current in-situ measurements and observations;
  - Definition and promotion of good data management practices and tools such as implementing Essential Variables (EVs) to structure observations and integrate in-situ and remote sensing data systems.
- Involvement
  - Establish a discussion forum where users, producers and custodians of in-situ data regularly meet and address known challenges and gaps;
  - Secure the involvement of the GEO Work Program (GWP) activities and existing regional and national networks;
  - Engagement with early career researchers to promote use of in-situ data;
  - Embrace new technologies (e.g. UAVs, AI, ML) through collaboration with partners including those in the private sector;

    - Organize promotional events and awareness/communication activities;
    - Provide recommendations to funding agencies with respect to in-situ data.
- Principles
    - Elaborate, share and promote the adoption of good / best practices which could be transferred from one discipline or region to another (e.g. enabling semantic interoperability, adopting data licenses);
    - Produce specific guidelines for in-situ observation systems and associated data management practices;
    - Encourage the uptake of relevant standards and principles by data providers;
    - Update the Data Sharing and Data Management Principles and illustrate how providers can implement them for in-situ data.
- Data and infrastructure
    - Utilize the GEO Infrastructure as a federating agent for existing network infrastructures to support multi-thematic and multiscale applications (structured by variable, area and access conditions) and organized by GEO Focus Areas;
    - Curate the content of existing in-situ data providers already brokered by the GEO-DAB to enhance the quality of their contribution;
    - Assess the existing in-situ data networks by region and domain, and seek to integrate more in-situ data services / providers into the GEO infrastructure, especially in key areas and where there are identified coverage gaps; in-situ Data Networks.xlsx;
    - Involve the Regional, National GEOs to fill the emerging in-situ data gaps;
    - Identify appropriate repositories for in-situ data, including use of data spaces and data lakes. Provide guidance to users and data providers to ensure in-situ measurements and observations, especially orphan data, is archived with a suitable repository. The GEO Knowledge Hub can be used for this purpose in exceptional cases;
    - Secure the sustainability of existing in-situ data systems;
    - Integrate new sources of in-situ data such as IoT, citizen science and UAVs.
- Knowledge
    - Collect, organize and expose knowledge extracted from in-situ data;
    - Extend the Dialogue Series on aspects of in-situ data and the associated Focus Areas;
    - Provide and augment relevant information for data users and providers available via the GEO Knowledge Hub.

4. OPERATIONALIZATION OF THE STRATEGY

To implement this strategy, some operational aspects should be considered:

- GEO Members will appoint technical principals responsible to promote the GEO in-situ data strategy in their countries and provide funding and resources to implement the aspects relevant to them.

- GEO Work Programme Activities are actors in identifying needs and gaps as well as demonstrating the value of the in-situ data in their domain specific activities. They link to the in-situ data networks that act as data providers.
- GEO Participating Organisations and Associates (that may include several in-situ data providers, and also private sector partners that can provide derived data or data services) who will be strongly involved in the advancement of this strategy.
- Regional and National GEOs are actively encouraged to include elements of this strategy in their respective activity plans.
- The GEO Data and Knowledge Working Group will update the Data Sharing and Management Principles for in-situ data providers to enhance interoperability. Capacity development events around in-situ data management will also be organized.
- The GEO Infrastructure Development Task Team (GIDTT) will be responsible for monitoring this strategy and customizing the functionalities needed to support implementation of the in-situ data strategy.
- The GEO Secretariat will communicate the relevance of in-situ data and support the actions that GEO develops under the framework of this strategy across all GEO community actors.

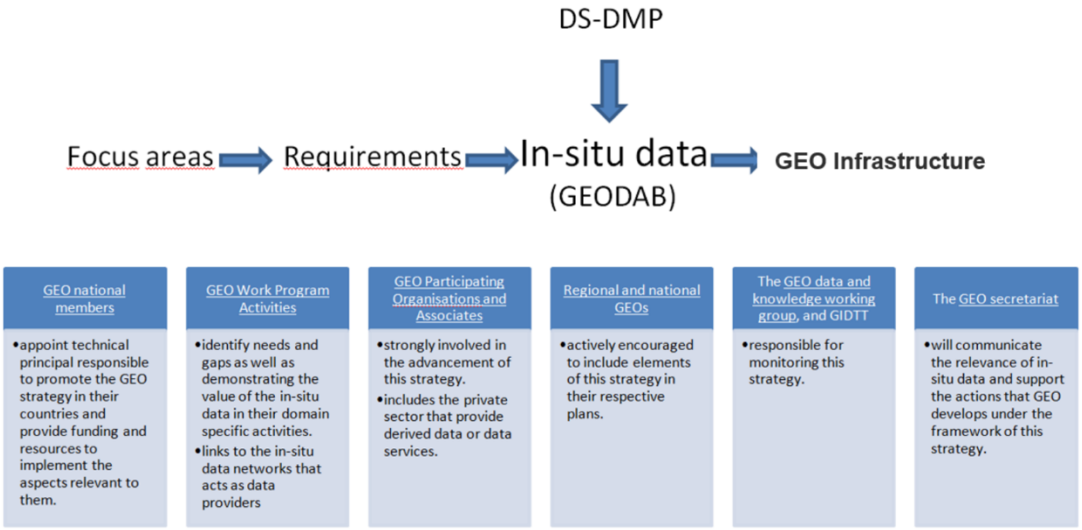

Figure 2: Role of different GEO actors in implementing the GEO In-situ Data Strategy

While the implementation mechanisms described above are equality important, actioning the in-situ data strategy should start by:

Determining user needs (starting with GEO activities) to demonstrate the application of in-situ data;

- Collect in-situ data requirements through existing tools: e.g. G-Reqs, that are not covered by current in-situ data and observations;
- Host discussion forums at GEO events, e.g. ODOK, where in-situ data consumers and producers regularly meet and address known challenges associated with in-situ data e.g. identified gaps;

- Secure the involvement of GEO Activities and existing regional and national networks;
- Produce in-situ data management guidelines for in-situ data networks.

The implementation of the GEO In-situ Strategy will be monitored by the In-Situ Data subgroup. The operationalisation mechanisms will be reviewed annually and prioritised accordingly as part of the annual work plan for the subgroup.

ANNEX A: EXISTING TOOLS

GEO supports the community with a set of tools that are regularly reviewed and updated based on the evolving needs of users. The tools currently available are:

- **GEO Infrastructure**
    - **GEOSS Platform**

GEOSS acts as a bridge between the contributing data providers and the users. It provides the necessary functionalities to link together existing and planned observing systems around the world. The associated GEO-DAB (Discovery and Access Broker) acts as the enabler for connecting and establishing interoperability among the various in-situ data providers.

GIDTT is in the process of defining the revised GEO Infrastructure, which will be presented to GEO EXCOM in March 2025.

    - **GEO Knowledge Hub: https://gkhub.earthobservations.org/**

A key element of the GEO tools and services is the Geo Knowledge Hub (GKN) that acts as a digital library providing access to Earth Observation (EO) Applications developed by the GEO Work Programme Activities as well as other resources supporting the GEO mission towards open data and knowledge.

- **G-reqs: https://g-reqs.grumets.cat/**

The Geospatial Requirements tool (G-reqs) captures identified user needs for in-situ data in a standardized form that builds into a comprehensive requirements database (Maso et al., 2023). The resulting database can be used to match requirements with existing datasets, allows gaps in-situ data coverage/ availability to be dicntied, and provide the basis to define new in-situ observing systems where necessary.

- **libinsitu: http://libinsitu.org/**

The libinsitu tool is an open-source python library (BSD licence) to transform heterogeneous time series of in-situ measurements into well-formed binary NetCDF files, compliant with Climate and Forecast (CF) convention, ready for sharing (see Blanc et al., 2022). The aim is to push for FAIR principles and the usage of interoperable and self-describing data, easing the scientific community to leverage open data.

- Self-assessment for FAIR compliance (Findable, Accessible, Interoperable, and Reusable (FAIR) Data): https://fair-data.netlify.app/

- Table designed to facilitate the assessment of data compliance with the FAIR principles. It has been adapted by Lionel Menard and Kalamkas Yessimkhanova as an outcome of the deliberations and collaborative efforts within the GIDDT group meetings. The original source is from the GO FAIR Initiative:

- **GEO Data Management Principles Implementation Guidelines: https://gkhub.earthobservations.org/records/gg85h-x8466**

Set of recognized GEO principles covering discovery, accessibility and usability of data, among other topics. They can be applied to the entire data management lifecycle, and are complimentary to the FAIR principles, which focus primarily on aspects of metadata, and the TRUST principles, which primarily focus on the curation and preservation of data and related resources.

- NiMMbus: (https://www.nimmbus.cat/)

NiiMMbus is a solution for storing geospatial resources on the MiraMon GIS cloud. The system implements the Geospatial User Feedback standard (http://www.opengeospatial.org/standards/guf). It allows users to provide comments, ratings, questions, creating a citation of an external geospatial resource (pointing to an external catalogue or repository) and associate feedback items to it.

- **Copernicus In-situ component activities: https://insitu.copernicus.eu/.**

The Cross-Cutting Coordination of the Copernicus In-Situ Component, led by the European Environment Agency (EEA), maps the availability of data and identifies gaps or bottlenecks in access. It also supports the provision of cross-cutting data and information and manages partnerships with data providers to improve access and use.

REFERENCES

Materials used in the preparation of this document. Some, but not all, are directly referenced and/or hyperlinked in the text where appropriate.

In-Situ Concept v0, GEO-XIII-2-Inf-01(Rev1) APPENDIX 4. In-situ Observations: Coordination Needs and Benefits) pp. 113-149 https://old.earthobservations.org/documents/geo_xiii/GEO-XIII-2-Inf-01(Rev1)_2016_Work_Programme_Progress-Report.pdf

First Steps Towards a GEO In-situ Data Strategy
https://old.earthobservations.org/documents/pb/me_202109/PB-21-11_Towards%20a%20GEO%20In%20Situ%20Data%20Strategy.pdf

Steps towards an In-situ Data Coordination within the GEO Community https://earthobservations.org/storage/app/media/documents/Working-Groups/Data/In_Situ_Concept_NoteV3.pdf

Canberra Declaration 2019

https://earthobservations.org/storage/app/media/documents/Events/GEO-Week-2019-Canberra-Ministerial-Summit/MS%204.2_Canberra_Declaration.pdf

GEO Data Management Principles Implementation Guidelines

https://gkhub.earthobservations.org/records/mq2sr-9jp64

GEO Data Working Group - Law and Policy Subgroup. (2023). Data Licensing Guidance. Geneva, Switzerland: Group on Earth Observations Secretariat. https://doi.org/10.60566/c3yd5-2s987

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18 https://doi.org/10.1038/sdata.2016.18

Maso, J.; Brobia, A.; Voidrot, M.-F.; Zabala, A.; Serral, I. G-reqs, a New Model Proposal for Capturing and Managing In-situ Data Requirements: First Results in the Context of the Group on Earth Observations. Remote Sens. ***2023****, 15, 1589.* https://doi.org/10.3390/rs15061589

Philippe Blanc, Raphaël Jolivet, Lionel Ménard, Yves-Marie Saint-Drenan. Data sharing of in-situ measurements following GEO and FAIR principles in the solar energy sector. 2022. ⟨hal-03811628⟩

Strobl, P.A., Woolliams, E.R. & Molch, K. Lost in Translation: The Need for Common Vocabularies and an Interoperable Thesaurus in Earth Observation Sciences. Surv Geophys (2024). https://doi.org/10.1007/s10712-024-09854-8*)*

Vaduva, Corina & Iapaolo, Michele & Datcu, Mihai. (2020). A Scientific Perspective on Big Data in Earth Observation. https://doi.org/10.1007/978-3-030-43981-1_8*.*

This document has been developed through a collaborative effort by current and previous members of the In-situ Data subgroup and the Data & Knowledge Working Group in GEO. The main authors of the document are listed in Table 1 below.

Helen	Glaves	British Geological Survey, UK
Joan	Maso	CREAF, Spain
Henrik	Steen Andersen	European Environment Agency (formerly)
Paola	De Salvo	GEO Secretariat
Florian	Franziskakis	GEO Secretariat (formerly)
Marie-Francoise	Voidrot	OGC
Felipe	Carlos	GEO Secretariat
Jose Miguel	Rubio	European Environment Agency
Robert	Downs	

Table 1: Main authors of the GEO In-situ Strategy document (January 2025)