Report on the Development of the GEO Knowledge Hub

This document was submitted by the Secretariat to the Executive Committee and is provided to the Programme Board for information

1   EXECUTIVE SUMMARY

1.1   What does this report contain?

This report contains a description of the implementation of the GEO Knowledge Hub (GKH), following the decision by the 52nd Session of the GEO Executive Committee to approve a one-year implementation plan from July 2020 to June 2021. It also contains a roadmap for further GKH implementation through the end of 2021.

1.2   What is the GEO Knowledge Hub?

The GKH is a digital repository of analytical tools, methods, applications and documents produced by the activities of GEO Work Programme (GWP) and presented for their reuse.

1.3   Why do we need the GEO Knowledge Hub?

Many activities of the GWP focus on developing methods, analytical tools and applications for using Earth observations (EO) to support evidence-based decision making. At present, potential users lack ready access to these tools. The GKH offers a platform for users to discover, learn about and employ these tools; it also provides opportunities for the GEO community to collaborate and provide mutual assistance related to GKH contents.

1.4   Who is it for?

The Knowledge Hub will empower public and private institutions to obtain information so that they are able to produce services relevant to local, regional, and global environmental and societal challenges through the use of EO, building on the results of GWP activities.

1.5   What are the benefits for GEO Members?

GEO Members will have access to the results produced by GWP activities. They will be able to adapt and reuse solutions and applications developed by the GEO community to address their specific needs, which will save them time, effort and other costs associated with discovering, accessing and reusing knowledge.

1.6 **What are the benefits for GEO Work Programme activities?**

The GKH will allow GWP activities to organize, curate and document their knowledge. Consequently, the work of the GEO community will be better organized, accessible and visible. The GKH will ensure long-term preservation of the work of GEO and foster the creation of user forums and knowledge exchange.

1.7 **What does it not do?**

The GKH is not a cloud service for large-scale geospatial data retrieval. The GKH is also not designed for search and discovery of EO data, neither will it provide computational resources for the users to run the applications.

1.8 **What is the infrastructure that supports the GKH?**

The GKH is a user interface built on top of InvenioRDM, a state-of-the-art software package for document management developed by a consortium of institutions led by the European Organization for Nuclear Research (CERN).The GKH is thus customizable to meet the needs of the GEO community.

1.9 **How much will it cost to operate the GKH?**

The operational phase of the GKH will require a combination of resources including programming and engagement skills, and infrastructure related costs. It is anticipated that the additional yearly cost in terms of hiring personnel is approximately CHF 60,000. A more detailed estimate may be found in the Appendix.

1.10 **How much effort is required from GWP activities to be part of the GKH?**

Based on recent experience, a rough guide is that GWP activities will initially require between 1 to 3 person-months to package their knowledge resources (journal articles, tools, applications) for inclusion in the GKH. The effort will be considerably offset by increased efficiency in knowledge sharing and application upscale for broader societal impact.

1.11 **What is the decision the GEO Executive Committee is expected to take?**

The GEO Executive Committee is asked to approve the proposed roadmap for the continued development of the GEO Knowledge Hub through the end of 2021, at which point the GKH is expected to become operational. Prior to the GEO Plenary in November 2021, the GEO Secretariat will prepare a GKH operation plan to be considered at the 56th Session of the GEO Executive Committee for approval.

APPENDIX A

Detailed Report on the Implementation of the GEO Knowledge Hub

(July 2020 – July 2021)

1   PURPOSE

This document presents the status of the GEO Knowledge Hub (GKH) implementation. Additionally, the document focuses on lessons learned and the issues considered most relevant to continued implementation of the GKH. The technical details supporting this document are contained in "GEO Knowledge Hub Implementation Plan", abbreviated as GKH-IP-2020, approved by the 52nd Session of the GEO Executive Committee (ref ExCom 52.5).

2   THE GEO KNOWLEDGE HUB

The development of the GKH is part of the GEOSS Infrastructure Development Foundational Task outlined in the 2020-2022 GEO Work Programme (GWP):

> *"The Knowledge Hub is envisioned as a set of curated and linked documents that contain relevant information for Earth observation applications and which is integrated with the GEO website. It is intended to provide authoritative, validated and reproducible content for evidence-based reporting on policy commitments and decision making."*

The GKH is a central digital repository that provides access to codified knowledge, featuring replicable, open science workflows for applications using big EO data to support sustainable development. The GKH contains documentation linking: (a) research papers describing methods; (b) algorithms/cloud computing resources for processing; (c) EO datasets used (in situ, satellite, airborne, citizen science); and (d) results for verification.

The primary goal of the GKH is to advance the transformation of EO data into knowledge-based services for evidence-based decision making. Several Flagships and Initiatives of the GEO Work Programme (GWP) have called for the establishment of a centralized, efficient means for transferring knowledge and scaling-up applications developed as a result of their efforts. In this way, the GKH will respond to the need for greater efficiency in reproduction and dissemination of applications from the GWP, while lowering the barriers for developing countries to access to free cloud services and help with the hosting, processing and analysis of big Earth observation data in support of sustainable development.

3   IMPLEMENTATION STATUS

Initial GKH implementation has focused primarily on development of two modules: (a) data search; and (b) document submission. Based on keywords provided by users, the **data search** module retrieves the relevant documents. Related documents pertaining to a given application of EO are presented together as knowledge packages. **The document**

submission module allows users to submit documents to the GKH, either individually or as part of a knowledge package. Before insertion into the GKH database, these submissions should be endorsed by members of the GEO community or by designated persons in the GEO Secretariat.

As of June 2021, the **data search** module will be completed, with the capabilities described in GKH-IP-2020. The **data search** module has been demonstrated at the Programme Board meeting.

The development of the **document submission** module is behind schedule, due to delays in the implementation of the InvenioRDM spatial data infrastructure, which forms the foundation of the GKH. This development is led by CERN and suffered delays due to the Covid-19 pandemic. InvenioRDM version 3.0, a pre-operational version, was originally planned for mid-2020. It was not released until April 2021. Based on the availability of InvenioRDM version 3.0, the GKH team expects to achieve the full implementation of the GKH by the end of 2021, in time for presentation at the GEO Plenary.

Simultaneously, the GKH team has been working closely with the leadership of Flagships and Initiatives of the GWP to produce knowledge packages from EO applications developed as part of their activities. Engagement thus far has included:

1. GEO Global Agricultural Monitoring (GEOGLAM)- Sen2Agri;
2. Global Observation System for Mercury (GOS4M);
3. GEO Human Planet;
4. Global Wildfire Information System (GWIS);
5. Global Network for Observations and Information in Mountain Environments (GEO Mountains);
6. GEO Global Water Sustainability (GEOGloWS);
7. GEO Land Degradation Neutrality (GEO LDN);
8. Brazilian Data Cube – Land Use/Land Cover (LULC).

Five knowledge packages have been completed as a result of this engagement, and others are in advanced stages of preparation.

Finally, the GKH Webinar Series introduced the GKH to the broader GEO community, showcased the initial deployment of knowledge packages, and provided the opportunity for interactive questions and answers with both package providers and the GKH team.

4   LESSONS LEARNED

During the first year of GKH implementation, we learned three important lessons: (a) the need for technical expertise to adapt *InvenioRDM* to the needs of the GKH; (b) the importance of strong interaction between the GEO Secretariat team and the GWP activities to include documents in the GKH; (c) the need for expertise in the GEO Secretariat to be able to guide Members, POs and GWP activities on how to profit from the recent technological advances in cloud computing and big EO data.

Regarding *InvenioRDM*, while full development of the GKH has been impacted by delays in the deployment by CERN of version 3.0 of the product, the overall experience has been positive. The initial releases of *InvenioRDM* have met the basic requirements with

respect to support software for the GKH. It provides a powerful application programming interface (API) that has reduced the workload of the GKH implementation team. Nevertheless, we learned that the GKH requires a permanent effort of a half-time expert to maintain and improve the GKH user interface.

In terms of the interaction with the GWP, we found that the Secretariat needs to allocate dedicated staff time to engage with the GWP activities so that can properly contribute to the GKH. This topic is further addressed in the next sections.

Experience in developing the GKH has shown that the Secretariat needs to have the technical capacity to adequately engage with members of the GEO community on topics such as data and knowledge infrastructures, cloud computing, analysis-ready data, data standards, etc. While the Secretariat is not expected to play a leading role in these areas, it is important that the Secretariat is able to bring together relevant stakeholders to pursue opportunities, track developments, advise GEO governance bodies on their implications, and support GWP activities to benefit from latest technological advances.

5   RESOURCE REQUIREMENTS FOR THE GEO WORK PROGRAMME ACTIVITIES

In addition to resources managed directly by the GEO Secretariat, each activity of the Work Programme will need to allocate resources to organize its knowledge and make it available in the GHK. Based on recent experience, a rough guide is that GWP activities will initially require between 1 to 3 person-months to include their knowledge resources (journal articles, tools, applications) in the GKH. The effort will be more than offset by increased efficiency in knowledge sharing and application upscale for broader societal impact. Over time, resources from the GWP are expected to surpass the GEO Secretariat operational contribution, if the GKH is to succeed.

6   RESOURCE REQUIREMENTS FOR THE SECRETARIAT

The current resource allocation for the initial phase of implementation of the GKH has relied on staff from the GEO Secretariat and on voluntary contributions from Brazilian experts. We divide the cost estimates in two parts: (a) Conclusion of the development phase (July-December 2021); (b) Operational phase (January 2022 – onwards).

For the conclusion of the implementation, it is required to hire a consultant at 100% time that will take the role of lead developer and will work together with the seconded expert from Brazil (Gilberto Queiroz).

The operational phase of the GKH will require a combination of skills, including:

a) Programming skills: keep the software up to date with the evolving development of InvenioRDM and to make improvements requested by users. We estimate this effort as a half-time software specialist hired as a consultant.

b) Engagement skills: maintain close contact with the GEO community and specifically with the GWP to increase and improve the documents available in the GKH. This position would likely be best filled with resources from the Secretariat, with existing contacts and the long-term mission objectives necessary to perform this task well.

c) Software/Infrastructure related cost: The *InvenioRDM* core is developed and maintained by a Consortium at CERN which covers 97% of the GEO Knowledge Hub. An additional 3% is needed to customize it to meet the needs of GEO Knowledge Providers and users. The cost below represents the efforts towards that 3%.

TABLE 1

Cost estimates for operation of the Geo Knowledge Hub

Item	July-Dec 2021	2022
Lead developer at 100%	US$ 60,000	
Development consultant at 50%	INPE Brazil (seconded)	US$ 60,000
Engagement with GWP	GEOSEC staff	GEOSEC Staff
Cloud infrastructure/Year	US$ 2,000	US$ 10,000

7   ENSURING THE SUCCESS OF THE GKH

Based on the results and lessons learned in the first year of implementation, the document now outlines several indications of the actions required to ensure that the GKH becomes a useful and relevant resource for the community. The recommended actions are:

- Ensure that the planned implementation of the GKH is completed by the end of 2021;
- Allocate adequate GEO Secretariat staff to enable effective interaction with the GEO community and thus ensure GKH relevance;
- Identify a mix of internal and external resources that will be responsible for supporting the operational mode of the GKH;
- Engage strongly with the GWP to enhance and improve the documents available in the GKH, as well as reinforce and scale-up activities of the GWP through the sharing of tools and knowledge resources.

The success of the GKH will depend on the continued close collaboration between the GEO community and the Secretariat. The GKH is not an end in itself, but a means of promoting sharing and reuse of EO applications and results by the GEO community. Ideally, it will become a key tool for capacity development and empowerment of GEO Members, especially institutions from least developed countries.

8   CONNECTION WITH THE GEOSS PLATFORM

There are two key pieces of the GEOSS infrastructure: a) the GEOSS Platform, composed of the GEO Discovery and Access Broker (DAB) and its user interface, the GEOSS Portal; and b) the GEO Knowledge Hub. These systems address very different needs and user scenarios and are based on different technologies.

The primary function of the DAB is to extract information from data providers that allows discovery and access to Earth observation datasets. The DAB can decode a large variety of metadata formats and to harmonise such information for user access.

The GKH is intended to provide access to the methods, tools, training data and other components needed to reproduce the results developed by GWP activities. The resources themselves are generally stored externally and are linked by digital object identifiers (DOIs). The resources to be included in the GKH are curated by the GWP activities themselves, that is, they choose and maintain those resources that are considered most relevant. The GKH also provides a full-text search function for users who wish to access the resources.

During the next implementation phase (July-December 2021), the GKH development team will work with the GEOSS Platform team to develop a working connection between the two systems. Ongoing conversations between the development teams have identified different options for this connection. The connection between the GEOSS Platform and the GKH will enable users to find additional data in the GEOSS Platform that could be used with the tools in the GKH.

9   PROPOSED ROADMAP

The proposed roadmap sets the following milestones for the second half of 2021:

1. Completion of the development of the GKH document submission module in time for presentation at the GEO 2021 Plenary;
2. Inclusion of knowledge packages in the GKH, even before the document submission module has been completed;
3. Availability of the GKH data search module to the community by August 2021;
4. Development of a strategy for GKH operations, including defining the roles and responsibilities of the various stakeholders, and associated workflows;
5. Presentation of the strategy for GKH operations at the GEO Plenary 2021.

Given the current resources available for implementation and the stability of the base software, it is *highly likely* that the GKH will become operational at the end of 2021. Once full implementation is complete, the focus of the GKH will shift from *implementation to operations*.

The concept of the GKH and its added value in terms of scaling up GWP activities is gaining traction with the GEO community. The Secretariat needs to ensure that this momentum continues such that the ultimate goal of the GKH – that of sharing applications of EO to address environmental issues – is fully realized and aligned with community expectations.

APPENDIX B

Community Endorsement of the GEO Knowledge Hub

Initial reaction to the GKH received from the leadership of a select number of GWP Flagships and Initiatives, as well as the broader GEO Community, has been positive, including the following endorsements:

- Ian Jarvis, GEOGLAM:

  "Over the last decade GEOGLAM has been focused on turning research into operational monitoring systems to enhance global food security. In the process, the community has produced a tremendous amount of knowledge, usually driven by project funding. GEOGLAM has been concerned with retaining and organizing these knowledge resources once project funding ends, but until recently sustained support for knowledge management has been beyond our means. The GEO Knowledge Hub (GKH) was seen as a possible solution to this challenge, and as a result GEOGLAM has been a major early supporter, helping develop the GKH proof of concept.

  Based on the excellent progress by the GEO Secretariat so far, the GEOGLAM community believes the GKH concept has been proven. It is now time to move boldly beyond concept and ensure the GKH system is developed to a point where we can openly access, enter, organize, discover, and maintain our community knowledge resources. We believe this will be a transformative initiative for GEO, a great step forward for the global EO community, and a significant contribution to open science. Beyond GEOGLAM, we believe the GKH will provide a multi-sectoral knowledge integration platform to address the wicked challenges embedded in our shared GEO priorities. The time to move beyond proof of concept is now, and GEOGLAM looks forward to helping the GEO community move forward.

  We are looking forward to the GEO KH collected knowledge packages for establishing cooperation and exchanging datasets with GEO projects in order to address societal challenges, and science and policy data demands. Most of societal challenges need to be addressed by pooling together different expertise and datasets, and the knowledge packages that you are collecting provide a valuable insight in what resources are available. We feel the GEO-KH should be able to maintain the knowledge packages and hopefully receive updates from the knowledge providers."

- Carolina Adler, GEO Mountains Initiative:

  "We foresee a benefit in that it would allow us to combine, package, and present relevant data and information in formats that are accessible and that respond to explicit knowledge needs expressed by users, especially those in policy and management. It also provides us a means to connect and collaborate with other partners (data providers) whose data resources show potential and relevance for our thematic area of focus around mountains".

- Daniele Ehrlich, GEO Human Planet:

  "The GEO KH could facilitate the dissemination of the knowledge – beyond what the knowledge providers do directly. In fact, the consistent knowledge structured within GEO KH will facilitate the access of users. The GEO Hub could also keep track of the downloads from the user community and feed that information back to the knowledge providers. I would also encourage GEO projects to come together to identify thematic areas that remain uncovered, and to rise the visibility of those knowledge packages and datasets used by other GEO projects within the GEO community. I strongly believe that the value of the combined knowledge packages you are collecting under the GEO KH will be more than the sum value of all knowledge packages".

- Tomas Artes Vivancos, Global Wildfire Information System (GWIS):

  "In the case of wildfires, the European Forest Fire Information System (EFFIS and GWIS at global scale with GEO) produces and shares data, works in multidisciplinary fields and tries to fill the gap between research and civil protection mechanisms.

  A common nexus or a hub is an essential piece to break the silos between research fields and organizations. A common place to work and share data could lead to improve data accuracy, multidisciplinary analysis, exchange of data and knowledge between actors. This last fact would speed up the production and disseminate the knowledge.sharing the data and knowledge at global scale would allow us to analyze and produce potential knowledge that would be impossible only focusing on a single country and using only one knowledge field."

- Stefan Nilsson, Manager International Relations (Swedish Meteorological and Hydrological Institute, Swedish GEO Alternate):

  "The first GEO Knowledge Hub webinar was a real success, and it clearly showed the potential of the GKH. The presented examples from GEOGLAM, the Human Planet Initiative and the Land use via the Brazilian Data Cube were all excellent showcases."