GEO Knowledge Hub (GKH) implementation plan

This document is submitted to the Executive Committee for decision.

Dear Members of the Executive Committee,

At its 17th meeting, the Programme Board (PB) considered the draft implementation plan (version v4.1, dated 20 May 2020) for the GEO Knowledge Hub (GKH), as well as the summary report from the GEOSS Infrastructure Development Task Team (GIDTT) who have been working with the GEO Secretariat through successive iterations of the implementation plan. The PB took note of the findings and recommendations contained in the GIDTT report, which are included here.

Upon consideration of the report from GIDTT and discussion of the various aspects and merits of the GKH Implementation Plan, the Program Board endorsed the document and is recommending that the GEO Executive Committee approve the submitted GKH Implementation Plan v4.1.

From GIDTT Report: Findings and Recommendations

1 SCOPE AND SCALE

1.1 Findings:

The GKH is an ambitious effort with a broad scope. The scope and scale of the effort as described in the implementation plan is significant and has broad implications that will have short- and long-term implications on GEO Member and Participating Organizations contributions and planned work activities. The most prevalent implications are identified in additional findings in this document. Additional exploration, community engagements and refinements in coordination with the GIDTT will improve development of the GKH implementation plan to ensure the results support GEO Member needs.

1.2 Recommendations:

1. Proceed with limited and defined prototyping and piloting activities that provides an expanded proof of concept demonstration, report of findings, lessons learned, and proposed scaling that will provide greater understanding and vital information to inform ongoing implementation.
2. The resulting outcomes of recommendation 1.1 shall be used to develop a definition that more clearly defines the GKH, identify capabilities to implement in the GEOSS Platform, and the scenarios that are appropriate for use.
3. The proposed strategy to manage GKH ambitions and scalability is to proceed with a step-wise incremental design and implementation approach, considering also the outcomes of the GIDTT dealing with user requirements, and the future GEOSS infrastructure architecture.
4. Ongoing exploration and development to implement the GKH be limited to one year. After the first year of implementation, additional consideration be given GIDTT, Secretariat, PB and Executive Committee by relevant communities in further scaling and adoption.

2 RESOURCING

2.1 Findings:

Previous findings indicated significant resources will be necessary to extract and document the tacit knowledge of domain experts, making available the process, workflows and models used to frame and address scientific and policy-relevant questions, and to link the relevant data so that it can be processed in a transparent and reproducible manner. General concern exists amongst the community related to resource commitment that will be required by producers, providers, GEOSS Infrastructure components and from the GEO Secretariat who is already committing a number of resources to this cause. A greater understanding of investment and resourcing will garner support and support development of an implementation plan that has adequate resourcing.

2.2 Recommendations:

1. Engage domain experts in understanding the level of effort and investment that will be required by GEO members and contributors to support their contributions to the GKH. Incorporate feedback into the development of the implementation plan.
2. A long-term investment plan that includes costs associated with integration of the GEOSS Platform, staffing and scaling of the GKH, and operations and maintenance of the GKH should be included in the report of progress at the end of the one-year period that is recommended.

3 ENGAGEMENT

3.1 Findings:

Understandably, the current development approach has been with a limited audience to support the exploratory phase of the GKH. Now that the GKH has received an endorsement from the GEO Plenary, greater engagement with the community is needed to understand the approach and to assess levels of readiness in the GEO Community. Engagement with the community will improve assessment of readiness and help develop the implementation plan.

3.2 Recommendations:

1. Additional outreach is recommended to a broader community of users to better understand the current state, gather requirements, capabilities and limitations that may include national and or organizational policy restrictions.
2. Interaction with the Data Sharing and Data Management task is necessary to determine alignment of principles, guidance, workflow approval and other processes that may be necessary.
3. Immediately engage with the GEOSS Platform team to improve understanding and to increase transparency in development and operations.
4. Solicit participation from the GEOSS Platform team, providers and the user community in this development phase of the GKH to provide contributions and recommendations that can inform the development of a long-term implementation plan.
5. Provide regular updates on activities in regard to developers and users.

Based upon these findings and recommendations of the GIDTT, the Programme Board endorsed the GKH Implementation Plan version as presented, comprising development of an extended GKH proof of concept by means of a step-wise approach of an incremental design and implementation process, to be aligned with the future GEOSS development in due course.

Yours sincerely,

Prrogamme Board co-chairs

GEO KNOWLEDGE HUB

Implementation Plan (Draft)

Phase 1 - Proof-of-Concept (July 2020 - June 2021)

VERSION 5 .0 | 17/06/2020

Table of Contents

1. Introduction

The GEO Strategic Plan 2016-2025: Implementing GEOSS reads: *"To realize its Vision and maximise the benefits that GEO can bring to users, through 2025, GEO defines three spheres of activity focusing on advocacy for the value of Earth observations as a fundamental component of timely information; engagement with stakeholder communities to address societal challenges; and delivery of critical data, information and knowledge to inform decision-making."* The GKH is targeting in particular the latter portion, knowledge delivery to inform decision-making. GEO Plenary 2019 acknowledged the value of adding a knowledge hub to improve knowledge delivery and approved ongoing efforts to explore and implement the GKH. This implementation plan outlines the development of the GEO Knowledge Hub (GKH).

Knowledge here is used in the ordinary dictionary sense of *"facts, information and skills acquired through experience or education; the theoretical or practical understanding of a subject." (Oxford).* Therefore, the knowledge contained in, or referred to, by the GKH comprises: observations (facts), including references to the GEOSS Platform content; training materials for skills learning (including reference to capacity building activities); and scientific processes (algorithms, procedures, recipes) and their products (information derived by the theoretical or practical understanding of a subject). In more concrete terms, the knowledge contained in the GKH represents the body of facts, information and skills widely shared in the scientific community. In environmental related areas, which are the ones under GEO's remit, shared knowledge and reusable/replicable science are essential for decision-making at a global scale. For example, when submitting reports and communications to UN conventions such as UNFCCC, UNCBD, UNCCD and UN SDG, the best practices request Member States to provide information that is verifiable and transparent.

The use of the concepts 'knowledge', 'co-design' and 'co-production' in this document has been inspired by the approach taken by important scientific initiatives such as Future Earth and Belmont Forum. In the seminal work of Mauser et al (2013), where the authors consider that knowledge for sustainability needs to combine two approaches:

A. The reductionist view, where knowledge is achieved by focusing only on understanding the parts of the system. Knowledge should — in the reductionists view — ultimately be put in a formal framework and thus be universally recognizable and to a large extent exchangeable across contexts.
B. The user-centered worldview, that considers knowledge as being composed of different configurations and validated practices that emerge as a result of agents' learning within their natural and/or societal contexts. Thus knowledge is mostly what works in a particular context. In this worldview both social and natural

> science knowledge are interdependent and inseparable aspects of the same knowledge.

Mauser et al.(2013) state: "in societally relevant research, the gap between science as the active knowledge producer and society as the passive recipient in the knowledge production process will need to be replaced by a process of co-design and co-production of knowledge". In this context, we consider the following definitions for co-design and co-production:

> **co-design** is the process of joint definition of sustainability challenges and the translation of these challenges into research needs. Stakeholders and academic participants work together to establish the research goals and to agree on their roles (Mauser et al., 2013).

> **co-production** is the process of development of scientific responses that meet the needs of the stakeholders. The research questions are derived from the overall needs of the project and validated by stakeholders to meet concrete needs (Mauser et al., 2013).

In the discussion that follows, the terms reproducibility, replicability and reusability (or generalizability) are used in the sense defined by National Science Foundation's Subcommittee on Replicability and Science (2015):

> **Reproducibility** refers to the ability of a researcher to duplicate the results of a prior study using the same materials and procedures as were used by the original investigator.

> **Replicability** refers to the ability of a researcher to duplicate the results of a prior study if the same procedures are followed but new data are collected.

> **Generalizability** (**reusability or reuse**) refers to whether the results of a study apply in other contexts or populations that differ from the original one.

In short, *reproducibility* involves the original data and code; *replicability* involves new data collection to test for consistency with previous results of a similar study; *reusability* (*reuse* for short) involves the original code (or a slightly modified version) with data for a different region of study, to aim to obtain similar results.

The area of Earth Observations (EO) is undergoing a major change. Public space agencies are making petabytes of data openly available. Cloud services are reducing the need for investment on infrastructure. Machine learning methods are delivering much better results than earlier methods. Data cubes are allowing EO data to be organised for easy access and processing. To benefit from this revolution, GEO has decided to adopt the concept of *results-oriented GEOSS*. The basis for this approach is to promote and facilitate the sharing of data, methods and results in our community.

As we know, sharing does not happen by itself. Using EO for decision making requires a combination of different resources and technologies. The implicit knowledge involved in producing useful results needs to be organized and codified so it can be shared. Organizing the resources that produce EO results for effective sharing and reuse is the primary aim of the GKH.

The GKH, then, is a cloud-based digital library (repository) providing access to knowledge required to ***reuse*** applications of Earth observations (EO). The purpose of the GKH is to reveal all components of a given application using EO data, whether stored within the GKH or housed in other repositories, including: (a) research papers and reports describing methods and results; (b) software algorithms and cloud computing resources used for processing; (c) in situ and satellite imagery data used; and (d) results for verification. The GKH is complementary to the GEOSS Platform and its development is included in the GEOSS Infrastructure Development Foundational Task (GIDFT), which forms part of the 2020-2022 GEO Work Programme (GWP).

The contents of the GKH are linked resources that contain relevant information for EO applications that promote reusability, scalability[1], and co-design/co-production. Examples of resources include an HTML file, a PDF file (report or published paper), a Jupyter Notebook, an R or python markdown file, a GitHub page, a video, a repository entry linking to a dataset store with an assigned Digital Object Identifier (DOI), links to datasets in the cloud, and/or Open Geospatial Consortium (OGC) service or Application Programming Interface (API) links for data. Table 1 provides the main types of resources envisaged to be stored in the GKH, all linked to a given application.

Table 1: Types of resources to be stored in the GKH (list non-exhaustive)

[1] *Scalability* here refers to techniques that ensure that some quality of service is maintained as the number of users goes up or the complexity of the world increases (see Steed et al., 2010). Since they comprise open source components, applications featured in the GKH can be adapted to any required scale.

Type	**Format archived in GKH**	**Purpose**
Text documents	HTML, PDF, and DOC files	Describe methods and results (non-peer reviewed)
Scientific papers	PDF	Describe methods and results (peer reviewed)
Software (self-contained)	Jupyter notebooks, R/python scripts and markdown files	Executable and documented scripts to services that don't require API (e.g, AWS)
Software (API-dependent)	Jupyter notebooks, R/python/Google scripts and markdown files	Executable and documented scripts to services that require API (e.g, Google Earth Engine, Open Data Cube)
Links to software packages	Github links (with metadata and DOI)	Link to executable and documented scripts/algorithms
In situ data	CSV, XLSX, TXT, SHP, NetCDF (and other relevant types)	Description of ground samples
Links to in situ data	Metadata with DOI	Links to in situ data deposited in repositories (e.g., PANGEA)
Links to satellite data	STAC (Spatial Temporal Asset Catalog) files and CEOS opensearch XML scripts.	Provides general description of sets of images stored in cloud services
Links to any other relevant data and/or products	Metadata and other relevant formats	Links to data used in the application which is not strictly from satellites or ground measurements.
Videos	MP4, AVI, MKV	Describe methods, results, capacity development, talks

As has been recognized by the Expert Advisory Group (EAG), the GKH is a complementary application to the GEOSS Platform. The recommendation of the EAG, endorsed by the Executive Committee and by the GEO-XVI Plenary , was to design the GKH as a new module of the GEOSS Infrastructure. The design of the GKH (already endorsed) is that it will be interoperable with the GEOSS Platform. Currently, the GEOSS Infrastructure does not have a component which is dedicated to managing and sharing resources of different types. The extension of the current components to comply with the requirements of the GKH would require substantial resources and software development time. We consider that taking such risks is not required. The underlying technology for the GKH which does 95% of the work required by the GKH (i.e., CERN's InvenioRDM) already exists. The GKH will simply be a customised interface on top of InvenioRDM. Thus, instead of starting from scratch, the GKH will start from a stable and reliable base, which minimizes risks and costs of development.

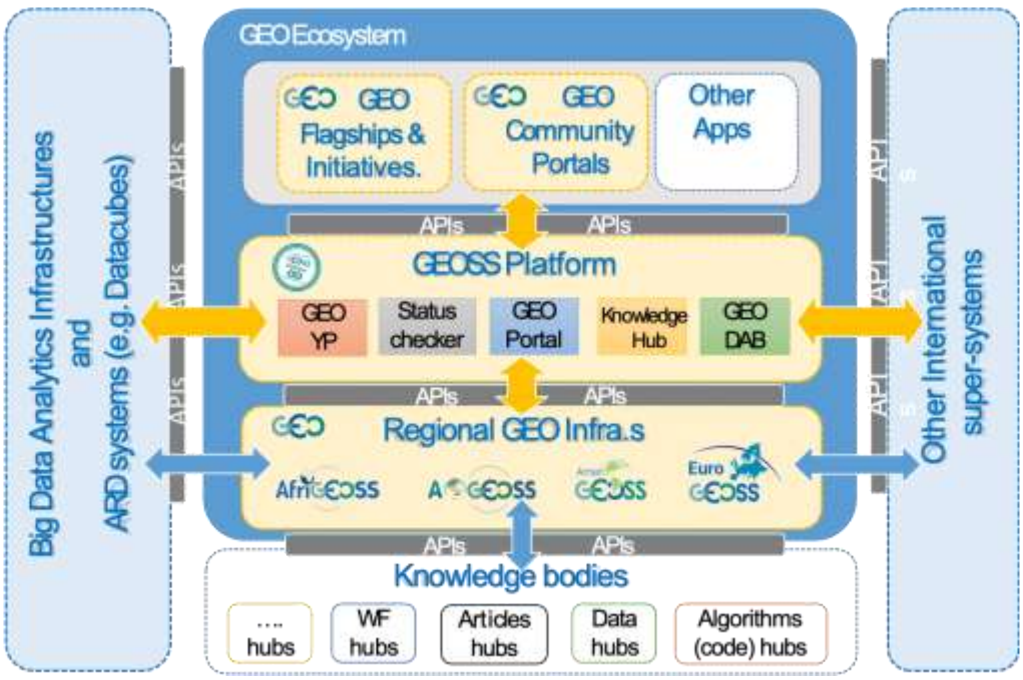

Figure 1: Community interactions within the GEO Ecosystem, including the Knowledge Hub (adapted from GEOSS Evolve study "Challenges to be addressed in evolving the Global Earth Observations System of Systems".)

Figure 1 provides the conceptual GEOSS Platform architecture with the knowledge hub as a component in the Platform. This figure is adapted from the study introducing an overall architecture to advance GEOSS and move from data to knowledge, contained in the GEOSS Evolve discussion paper entitled *Challenges to be addressed in evolving the Global Earth Observations System of Systems* (10 Aug 2018).

2. User Scenarios

In this section, several scenarios are presented in which EO are - or could be - used in a given end-to-end application that addresses a particular issue. These are intended to illustrate the kinds of needs and requirements that exist from a user perspective, both in terms of technicalities and resource content, and the implications these will have for determining characteristics of the GKH. These examples are not intended to form an exhaustive study of all possible users and scenarios, but rather to illustrate how they point to a set of common requirements that can guide implementation of the GKH along with the types of functionalities it should feature, in order to render it the most relevant and responsive to user needs as possible.

Sen2Agri

Sen2-Agri is an operational standalone processing system that generates agricultural products from Sentinel-2 (A&B) and Landsat 8 time series along the growing season. These different products consist of: (a) monthly dynamic cropland masks, delivered from the agricultural mid-season onwards; (b) cultivated crop type maps at 10 m resolution for main crop groups, delivered twice along agricultural seasons; (c) periodic vegetation status maps, NDVI and LAI, describing the vegetative development of crops each time a cloud-free observation is recorded. The Sen2-Agri system is free and open source, allowing any user to generate near real time products tailored to his needs at its own premises or on cloud computing infrastructure. The prospective users of the Sen2-Agri include: (a) Humanitarian actors such as the UN World Food Programme; (b) ministries of Agriculture of developing countries; (c) universities and research groups; (d) private companies.

At the request of the South Africa National Space Agency - (SANSA), the GEO Secretariat in partnership with the Sen2-Agri team has taken the first steps to prepare the Sen2-Agri software in the cloud. The SANSA technical staff is now to run the software using their own collected in-situ data, to generate the various crop related products. The advantages of having Sen2-Agri running in the cloud include:

- Any user in the world is able to use the proposed solution for his/her area of interest, leveraging the Cloud Infrastructure as a Service (IaaS) and Platform as a Service (PaaS) without facing bottlenecks such lack of proper hardware or computer power to run specific algorithms.

- The remote sensing data such Sentinel-2 and Landsat 8 essential to run the output products are already pre-loaded and available in the cloud and thus offering the opportunity to run remote sensing analysis without having to download any EO imagery.

From the perspective of resources to be stored and shared by the GKH, Sen2-Agri includes the following types: scientific papers, software, technical manuals, in-situ data, github link to code, and training material. Each of these resources will be stored separately in the GKH, and they will all be linked to allow those interested in reusing Sen2-Agri to retrieve all of the information available in the GKH.

GEO Cloud Credits Programmes

GEO has invited providers of geospatial cloud services to support our community by making credits available to support the development of EO applications relevant to the GWP. Since the 4th quarter of 2019, three such agreements have been established. These agreements are organized in a similar way: each provider makes cloud credits available to the community to be used in connection with its infrastructure. The credits are allocated based on competitive proposals. We anticipate that about 50 different projects based on these grants will be underway as of July 2020. These and similar future offers are expected to significantly increase the results produced by the community related to the GWP.

Considering GEO's Rules of Procedure regarding interaction with the commercial sector, all of the awards require that the software, data, results and documentation from each awarded proposal be made openly available. The results from each project will be either deposited into the GKH or made available in open repositories with a DOI which will then be linked to the GKH. Each resource deposited into the GKH will receive a DOI, which will be visible to GEO Discovery and Access Broker (DAB) and thus linked to the GEOSS Platform. The GKH will organise the resources provided by the awarded projects into categories, according to their contribution to the themes of the GWP.

For example, INPE/Brazil is using part of its cloud grant to develop software that achieves substantial performance gains for generating data cubes. Those gains are obtained by efficient use of cloud infrastructures. Producing an analysis-ready data cube of Sentinel-2/2A data for the whole of Brazil at 16-day intervals (560 tiles) in a conventional machine would take about 480 days. Using the parallel processing facilities available in cloud providers, this time is reduced to 46 hours, a reduction of 2500\%. The processing costs are also reduced in the same proportion. These results point to a substantial synergy between the cloud computing and the data cube technologies. This is an exemplary case where the availability of such software in the GKH will benefit the whole GEO community, since efficient production of data cubes is a major challenge.

Digital Earth Africa and Open Data Cube

The Digital Earth Africa project (DE-Africa) (https://www.digitalearthafrica.org) is an important new community activity of the GWP. DE-Africa aims to build a common infrastructure for the African continent, providing analysis-ready data from open access satellites such as Landsat-8 and Sentinel-1/1A/2/2A. DE-Africa will support the development of results for EO applications by the local institutions, using the experience gained in the Digital Earth Australia and the Africa Regional Data Cubes.

The basis for DE-Africa is the Open Data Cube (ODC) technology, which has proven capable of delivering successful results. The ODC technology provides a simple and easy to use application programming interface (API), that allows users to develop software to handle different types of problems related to the use of EO for sustainable development.

The GKH is expected to make a significant contribution to the DE-Africa initiative, by providing an infrastructure to share software, data, and documentation related to each intended application. In this way, users in one country can learn from the experience of other African partners. The results by each institution involved in DE-Africa will thus be replicable and generalizable. To this end, the intended thematic organization of the GKH (as described below) will be important to support DE-Africa.

Additionally, DE-Africa initiatives can benefit from information already available in the GKH. For example, the software developed by INPE/Brazil to optimise the generation of data cubes, developed in connection with an AWS grant, is of value to DE-Africa. In a similar way, the work of the Swiss Data Cube to develop a "Data Cube on Demand", which is portable and deployable, will be made available via the GKH and is relevant to the DE-Africa initiative.

Commonalities leading to requirements

All of the above user scenarios have a lot in common. They point to a great diversity of institutions and researchers who are using the new technologies for EO data (cloud computing, machine learning, and data cubes) to generate results. Many of those results are being developed in association with activities of the GWP, such as GEOGLAM, GEO BON and DE-Africa, and need a shared infrastructure to allow reuse and generalization of the results.

Visitors to the GKH should be able to browse through selected topics (e.g., biodiversity) and retrieve linked results containing software, data, and documentation that is relevant to them. To allow efficient retrieval, the GKH will need to support multiple types of documents, provide efficient indexing, and be selective in the content it stores. Considering the facilities already available in the GEOSS Platform, the GKH will not scan the web for EO data, since that functionality is already available in the GEO DAB. It will complement the GEOSS Platform to provide a digital library of selected resources which have proven their usefulness to the GEO community.

3. Objectives, Principles and Assumptions

In the drafting of the GKH implementation plan, several general principles, assumptions and constraints have been made that will have impacts on development of the GKH, and which need to be made explicit in the interest of transparency.

1) **Delivering knowledge**

The GKH is a cloud-based digital library containing an archive providing access to codified knowledge[2] involving the application of EO and Earth science for solving societal challenges. The GKH is concerned with detailing all development phases of applications, elaborated in an open science environment, which support GEO's engagement priorities. As such, *the GKH should not be viewed as delivering products or services and will not itself serve the function of a big data analytics platform.* Rather, the GKH represents a digital library that will provide efficient means for sharing knowledge and scaling-up applications developed within the GWP.

The GKH aims to support the co-design and co-production of knowledge, as defined in the Introduction. More specifically, documents available in the GKH should be organised so that stakeholders can understand their usefulness and can apply them.

2) **Scaling applications**

One of the ultimate aims of the GKH is to amplify the work of Flagships and Initiatives of the GWP such that solutions based on EO being applied at the local/national level may be scaled for implementation in other regions around the world, thus allowing these solutions to "go global." The GKH will focus on assembling the best in "tacit knowledge"[3] gleaned from the experiences of activities of the GWP and translate it into codified knowledge. This would include contributions from sources external to GEO, such as research funding programmes, to the extent that the research supports activities of the GWP.

3) **No computing capabilities provided**

---

[2] Codified knowledge is knowledge that is formalized and explicit; it is articulated, stored and accessed/transmitted, readily available for reuse by others; it can be both procedural and descriptive. (see Semertzaki, 2011)

[3] Tacit knowledge is knowledge that includes skills, ideas and practices. (see Semertzaki, 2011)

The GKH should only store resources, and should not provide computing capabilities for running algorithms. Therefore, there is no need for the GKH to have large storage and processing capacity. Since the GKH will not provide operational processing capabilities, its operation will be based on standard commercial cloud computing services, without additional requirements of high reliability.

4) Maintaining Vendor Neutrality

As defined by Gartner Glossary[4], *"Vendor-neutral describes a state in which no one vendor can control the definition, revision or distribution of a specification."*

Vendor neutrality pertains to two main, yet very different, aspects of the GKH: (1) the digital library itself where the resources will be indexed, and (2) running vendor-neutral computing environments for the actual EO applications. With respect to the first point, neutrality of the digital library will be guaranteed through a containerized architecture. Regarding the second point, the goal is to offer reusable EO application packages that can run in the end-user's defined computing environments.

In general, it is essential that the GKH be independent of any particular proprietary components to offer maximum flexibility in choice in order to be viewed as a transparent, neutral, honest information broker. This will be accomplished by promoting community-based, open standards supporting, open source solutions in which all methodologies, algorithms, code/software and results are open source, and can be used in a variety of virtual machines and cloud computing platforms. Further, vendor neutrality is key for supporting interoperability and cross-platform integration, reducing licensing cost and administration, and cultivating agile co-development and exploitation of new emerging technologies[5].

5) Relying on pre-existing, open source, free solutions/software

The development of the GKH will benefit from the substantial ongoing effort by the open source community to develop tools for scientific repositories. A large number of universities and research institutions worldwide are now using open source interoperable software solutions which not only have the benefit of transparency and adaptability, but also greatly expedite development by avoiding

[4] https://www.gartner.com/en/information-technology/glossary/vendor-neutral

[5] The general guideline to be applied in the design of the GKH is to use only technologies that are guaranteed to be vendor neutral. Such is the case of the underlying technology of the GKH (InvenioRDM by CERN).

duplication/reinventing the wheel while accommodating interoperability with proprietary systems and the rapid evolution of technology and societal needs.

However, it is not for the GKH to decide on the transparency degree of a solution. If the relevant community in GEO (e.g., GEOGLAM) considers the application relevant to be shared, and if this application itself is openly available, it will be shared on the GKH.

6) **Interface to proprietary APIs**

Considering that many of the solutions that will be stored in the GKH use cloud-computing services, there is a need to distinguish the contents of open source software available in the GKH from the underlying APIs which might be proprietary. Therefore, the GKH makes a distinction between cloud providers that work as **SaaS** (software as a service) such as Google Earth Engine from those that work as **IaaS** (infrastructure as a service) such as AWS. For solutions that work as **IaaS,** the GKH expects to store the full software as open source. In the case of **SaaS**, where cloud providers have a proprietary API to access their facilities, the GKH will store application code that uses the API, but **not** the API itself. This solution is far from ideal since it limits the generality of the software stored in the GKH, but it is a compromise that allows sharing of reusable software.

7) **Training the trainers**

Although the GKH will be useful to a wide range of stakeholders, from national experts needing to report on policy commitments and national priorities, to individual end users seeking practical solutions to local environmental challenges, the primary user of the GKH will be knowledgeable experts/technicians seeking to scale and customize applications of the GWP[6] for specific purposes related to evidence-based decision-making. Technical experts from research institutions may serve as intermediaries in assisting local end users to make full use of the GKH. GEO intends to leverage the capacity development networks of its partners in a “training the trainers” approach to broadening the knowledge and use of the contents of the GKH.

[6] “Applications of the GWP” is used to describe the results produced by GWP activities that are relevant to the GEO community at large and, more specifically, to GEO Member States to produce information for evidence-based policy making and for responding to global conventions. Possible examples include applications for crop monitoring, food security, deforestation mapping, and water extent and quality.

8) Supporting intercomparison of methodologies

The potential for reuse of the applications stored in GKH will allow experts to compare different methods for producing similar results. Consider, for example, the use of the Sen2-Agri, Google Earth Engine, and DE Africa platforms, outlined in Section 2 (Use Cases). Suppose one member of the GEO community develops an application for crop monitoring as part of the Google Earth Engine call; and that an African partner develops another application as part of the DE Africa project. The GKH will store these solutions alongside Sen2-Agri. Interested users will then be able to compare the three methodologies and to decide which one works best for them.

9) Leveraging existing resources

As a general principle, the resources (both human and financial) required to bring an open source EO application to fruition, along with its inclusion in the GKH, will fit within the existing resources of the GEO Secretariat[7] and the GKH team[8], and the contributing partners (GEO Members and/or Participating Organizations).

There are two main challenges to build the GKH:

1. Software development, which will require a modicum of resources since the work will be mainly a customization of an open source product (InvenioRDM from CERN); nevertheless, initial software development in the GKH will be minimal and can be accomplished by the Secretariat according to the staff duties stated in the *GEO Secretariat Concept of Operations* (CONOPS) document. Thus, instead of starting from scratch, the GKH will start from a stable and reliable base, which minimizes risks and costs of development.

2. Reusable/replicable application selection, organization, and sharing, bundled as a knowledge package[9], which will require an important

[7] The planned effort by the GEO Secretariat is based on the same allocation of hours used to develop the Proof of Concept for the GEO-XVI Plenary in Canberra. The staff allocation to the GKH work had no negative effects with respect to meeting Secretariat responsibilities; consequently, since the same staff time allocation to GKH implementation is foreseen, there are no grounds to expect any negative effect on Secretariat's performance.

[8] In this document, the term 'GKH team' refers to the team that is engaged in the implementation and operationalization of the GEO Knowledge Hub. In year 1 of this implementation, this team will be mostly composed of GEO Secretariat staff and seconded experts. It is envisaged and expected that in follow-up years the 'GKH team' will be broadened to include contributors from the GEO community.

[9] A "knowledge package" is defined here as the ensemble of EO data (both space-based and/or in situ),

engagement between the GKH team and the GEO community, especially the GWP. The resources that will be included in the GKH will be provided by the GEO community, mostly through the activities of the GWP.

There may be cases where synergies can be found with the commercial sector. The GEO Secretariat and/or GKH team may seek additional external sources of funding, including the commercial sector, to develop certain aspects of the GKH, and/or support application development. However the GKH will remain agnostic as to commercial services provided, through open calls and on an as needed basis. In all cases, any arrangements concluded with external funding sources by the Secretariat will be impartial and made transparent to the GEO community, in accordance with *Annex C: Rules of Engagement with the Commercial Sector*, *GEO Rules of Procedure* (March 2019 update).

10) Ensuring quality through peer review

The implementation plans of GWP Flagships and Initiatives, as approved by the Programme Board, contain details as to objectives (including meeting user needs), activities, resources, contributors and deliverables. These are considered to provide sufficient information for identifying candidate knowledge packages for the GKH.

In general, it is expected that leaders and participants in activities of the GWP will be consulted to decide on what results, methods and documents are relevant to be shared with the GEO community as a whole. As part of this process, knowledge packages under consideration for submission to the GKH must provide evidence of having been peer-reviewed for reliability, either through journal publication or formal recognition by experts in the community prior to GKH publication. The GKH team, working with the Programme Board, will be responsible for the due diligence required to ensure candidate GKH knowledge packages have undergone an adequate peer-review process (see Secretariat CONOPS).

11) Providing unrestricted access for discovery and retrieval

Access to the GKH will be unrestricted to any individual who wishes to search, discover and retrieve EO knowledge packages. No account registration will be required.

---

methodologies, algorithms, code/software, computing environment (where relevant) and published report(s) that supports a given EO application (see Introduction).

12) Restricting access for publication

Since reliability and accuracy are critical characteristics of the GKH in terms of applications contained in the repositories, access to the GKH will be limited to trusted individuals (i.e. leaders and participants in activities of the GWP, or others acting on their delegation) only for depositing knowledge packages and making comments. Each GWP activity will be requested to provide one or more points of contact; these persons will act as trustees of the GKH and will oversee the choice of which applications are stored and shared via the GKH. Account registration (single user) will be required for these individuals.

13) Complementing the GEOSS Platform

As has been recognized by the EAG, the GEO Knowledge Hub is a complementary application to the GEOSS Platform. The recommendation of the EAG, endorsed by the Executive Committee and by the GEO-XVI Plenary , was to design the GKH as a new module of the GEOSS Infrastructure. The design of the GKH (already endorsed) is that it will be interoperable with the GEOSS Platform.

The current GEOSS Platform has been designed to fulfill an important and necessary task: search the web for EO data, and organise it to allow searching by interested users. It has two main components: the GEO DAB that searches for data and organises the related metadata; the GEOSS user interface that provides query facilities for retrieval of the metadata which has been discovered by the GEO DAB. Currently, the GEOSS Infrastructure does not have a component which is dedicated to managing and sharing resources of different types. Thus, the GKH will complement the GEOSS Platform to provide a digital library of selected documents which have proven their usefulness to the GEO community.

The interoperability of the GKH to the other components of the GEOSS Platform will be achieved by using an infrastructure to the GKH that is capable of being searched by the GEO DAB as an additional data provider. The GKH should be compliant with one the protocols that are supported by the GEO DAB. In this way, the full contents of the GKH will be searchable and indexable by the GEO DAB.

The GKH will provide documents that are searchable by the GEO DAB in a properly organized metadata format. All documents entered into the GKH will be edited to include missing metadata, so that the DAB can index them correctly.

4. Requirements

In addition to the common characteristics of a shared infrastructure, identified in user scenarios above, a set of key requirements for the GKH has been identified by the Expert Advisory Group (EAG) over the course of several meetings (2018-2019), which set the foundation for the subsection on Principles, Assumptions and Constraints below. These requirements formed part of the document *Results-Oriented GEOSS: A framework for transforming Earth observation data to knowledge for decision making* (EXCOM-48.5) which was approved by the Executive Committee in July 2019, and served as a basis for the development of the proof-of-concept demonstrated to the GEO-XVI Plenary. The requirements include:

1) Support efficient text-based search.
2) Link document sets with different components (web pages, PDFs, links to DOIs, videos, Jupyter Notebooks, R markdown, URLs to data, GitHub pages, instructional videos, etc).
3) Use descriptors (metadata profiles) compatible with current search engine technologies and emerging solutions.
4) Be integrated with the GEO website and GEOSS Platform and based on open source software; use of open standards for interoperability of the GHK repository with other systems.
5) Include revision services for data entry, based on contributions from the GEO community.
6) Describe big EO data catalogues.
7) Promote abstract description of methods used for cloud computing to facilitate interoperability.
8) Support communities of practice to build packages that use open source scripting languages for big EO data analysis in the cloud to promote application portability.
9) Include applications that use models and support the open sharing of modelling software.
10) Work with the GEOSS Platform and other dataset search engines such that they access and promote repositories of continuous in situ data collection.
11) Ensure in situ data sets stored in accredited and recommended repositories are indexed.
12) Build an in situ data repository, managed by the GKH team, to ensure long-term curation/mediation and preservation of data entrusted to GEO by its community.
13) Provide links from the papers stored in the GKH to the accredited long-term data repositories and enable reciprocal links from those repositories in the

GHK.

14) Build a data repository for long-term archival, where needed, of Citizen Science data associated with in situ observations; otherwise link with existing Citizen Science repositories for seamless retrieval of data.
15) Promote and disseminate open source software for building multi-satellite analysis ready data that work on EO cloud platforms.

This implementation plan will ensure that each of these requirements is accounted for in the full deployment of the GKH.

5. Constraints

The proposed design of the GEO Knowledge Hub, while meeting the requirements set up by EAG in document EXCOM-48.5, also has to satisfy some additional practical constraints:

1. The GKH should be developed with the existing GEO Secretariat resources which have been approved by the GEO Executive Committee (EXCOM) in the document *GEO Secretariat Concept of Operations* (CONOPS), and in such a way that its development will not compromise the other duties of the Secretariat as outlined in the CONOPs.
2. The actual software development should be kept to a minimum, reducing the risks of failure and the costs of future maintenance. The GKH should be based on an open source product that has its own development team and that has been adopted by a significant community.
3. The implementation of the GKH should be independent of any particular cloud provider.

6. Addressing the Principles, Requirements and Constraints

Considering the Principles and Assumptions, Requirements, and Constraints outlined in this document, the GKH team evaluated available open source research management data platforms. Four platforms were considered: CKAN, Figshare, Zenodo, DSpace (for a similar approach, see the work of Amorim et al (2017)). From this analysis, Zenodo was chosen initially for the following reasons:

1. Easy deployment and minimal maintenance
2. Allows creating communities to validate submissions
3. Supports Dublin Core and MARCXML for metadata exporting
4. Complies with OAI-PMH for data dissemination
5. Allows easy and fast interface with **github** for making code citable

The compliance with OAI-PHM allows Zenodo to easily interface with the GEO DAB, which understands OAI protocols. The interface to **github** is unique to Zenodo and allows easy code citation. None of the other open-source research data management platforms had an interface to **github**, the world's platform of choice for developing and sharing open source software. Since one of the goals of the GKH is to allow software reuse, we are convinced that, of the current open source research data management platforms, Zenodo is the only one that meets our goals. Table 2 shows how each platform matches the stated requirements.

Table 2 - Research Data Management Platforms

	CKAN	Figshare	Zenodo	DSpace
Easy deployment	❇		❇	❇
Communities	❇	❇	❇	❇
Dublin Core, MARCXML		❇	❇	❇
OAI-PMH		❇	❇	❇
Github interface			❇	

In early 2020, communications with CERN (which created Zenodo) revealed that a new product being developed, Invenio RDM, would provide all functionalities currently available in Zenodo, with a more manageable and maintainable code base. The GKH team evaluated the qualities of the InvenioRDM software by taking part in a community workshop held at the European Organization for Nuclear Research (CERN) in late January 2020. This evaluation concluded that InvenioRDM met the requirements to be the supporting platform for the GKH. InvenioRDM enables organizations to securely house research products and make them discoverable, shareable, and citable -- from publications and presentations to datasets, software, lay summaries, policy documents, and guidelines.

The GKH will therefore be a customization of the [InvenioRDM](https://inveniosoftware.org/products/rdm/) software, developed and maintained by CERN. As defined by CERN[10], "*InvenioRDM (research data management) is a platform that allows researchers to share and preserve scientific results. Researchers can share anything from publications, posters, presentations to datasets and software. Once a researcher has shared a result, they get a DOI (Digital Object Identifier), that allows them to properly cite their result*". The CERN team has considerable experience in developing software for managing research data. To develop InvenioRDM, CERN has partnered with international universities and institutions. The purposes of InvenioRDM are exactly the same as those of the GKH: a) disseminate and archive; b) enable reuse.

Since InvenioRDM is developed and maintained by CERN, there is no need to involve any GEO team in its implementation. The first release of InvenioRDM will be available only in Summer 2020. The system description part of the GKH implementation plan describes the user interface and the steps required for customization of InvenioRDM. The GKH implementation plan is based on a strategy of reducing the actual software development to a minimum and thus reducing risks. Therefore, the GKH team considers to have chosen an approach that meets the constraints outlined above.

7. Information Flow

A simplified GKH data flow diagram is presented in Figure 2, and described in more detail in what follows.

[10] https://inveniosoftware.org/products/rdm/

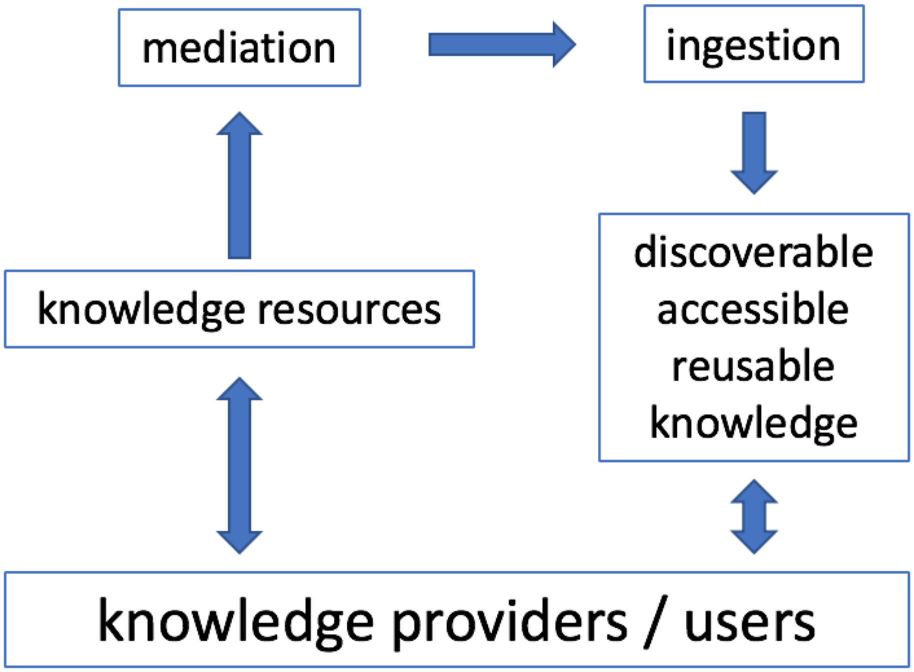

Figure 2 - Information flow in the GEO Knowledge Hub

The *knowledge resources* of the GKH include information relevant to the activities of the GWP. We expect each activity to contribute to the GKH, sharing results and best practices. This practice will be beneficial to all involved, by enhancing the visibility of GEO's work and providing a unique focal point for the results of the GWP. It will help all those interested in EO to have a place to go to learn how best to use big EO datasets. The GKH team will work with the Programme Board to ensure that the GKH contribute to amplify the work of Flagships and Initiatives of the GWP.

Mediation is the process of joint work by the GKH team and the GKH contributors to ensure inclusion of trusted information. The role of the Secretariat in mediation of information to be made part of the GKH and the GEOSS Platform in general has been recognised and accepted by the Executive Committee with its approval of the Secretariat CONOPS, which states that one of the role of the Secretariat Data Officer is: "act as curator of GEO Infrastructure content in close collaboration with Content/Data providers". Since the GEO community needs reusable best practices, information going into the GKH must be verified and organised, while dispersed components of a document set have to be linked. The GKH team will interact with the authors so that the methods, data and software are consistent and follow the GEOSS Data Sharing principles, the GEOSS Data Management Principles and other applicable Open Science principles. As a general rule, the expectation is that leaders and contributors to the GWP activities will provide their expertise to the GKH team when selecting relevant results to be included in the GKH. Additional experts in the field, such as a GEO Community of Practice[11], may be called upon to serve as peer reviewers to help decide the relevance of a particular knowledge contribution to the GKH (see Principles and Assumptions section, item 10, above for more details on peer review procedures).

The *ingestion* process of the GKH will be as automated as possible, but human intervention is required for the final checks. For best query results, it will include the full text of documents; this requires advanced text-based search capabilities. Given the need of GEO to provide open global access, the GKH will store either open access papers or post-prints[12] of journal papers that are not openly accessible.

In summary, the GKH will function as a digital library containing "recipe books" for reusing and scaling up the EO applications for societal benefit globally. These open source solutions will be vendor independent and take advantage of the zero download model of the new digital economy, where big EO data is queried and analysed on the cloud, eliminating the need for costly and quickly-outdated computing infrastructures at the national level. Advances in technology, coupled with the convenience of cloud computing, thus will lower the barrier for governments, particularly those of developing countries, to access and apply EO. In doing so, the GKH will fill the role of a robust, trusted, authoritative information for decision making.

---

[11] As defined by the *GEO Strategic Plan 2016-2025: Implementing GEOSS*: Communities of Practice may form within or engage with GEO in response to specific needs or shared interests. Communities of Practice may, for example, consider aspects of societal challenges, or managing and developing pooled resources. These Communities demonstrate the convening power of GEO by engaging stakeholders along the full continuum of the data and information chain, from providers, to processors, to end users. Existing Communities of Practice may be mobilized to assist with engagement activities aimed at facilitating a comprehensive dialogue between Earth observation providers and end users in order to identify Earth observation products that address users' needs.

[12] Most scholarly publishers allow researchers to share post-prints in public repositories. Post-prints have the same content as the journal paper minus the formatting. A detailed list of publishers' policies is available at http://www.sherpa.ac.uk/romeo.

8. System Design

This section details the technical requirements for the GKH digital library which will be used by providers to register their linked EO application resources and by end-users to retrieve them. Access to cloud computing instances and related information are not part of this section because, in the interest of preserving vendor neutrality, it will be the responsibility of the providers to prepare the environments and up to users to choose where to run applications. Figure 3 provides a general schematic of the components of the GKH.

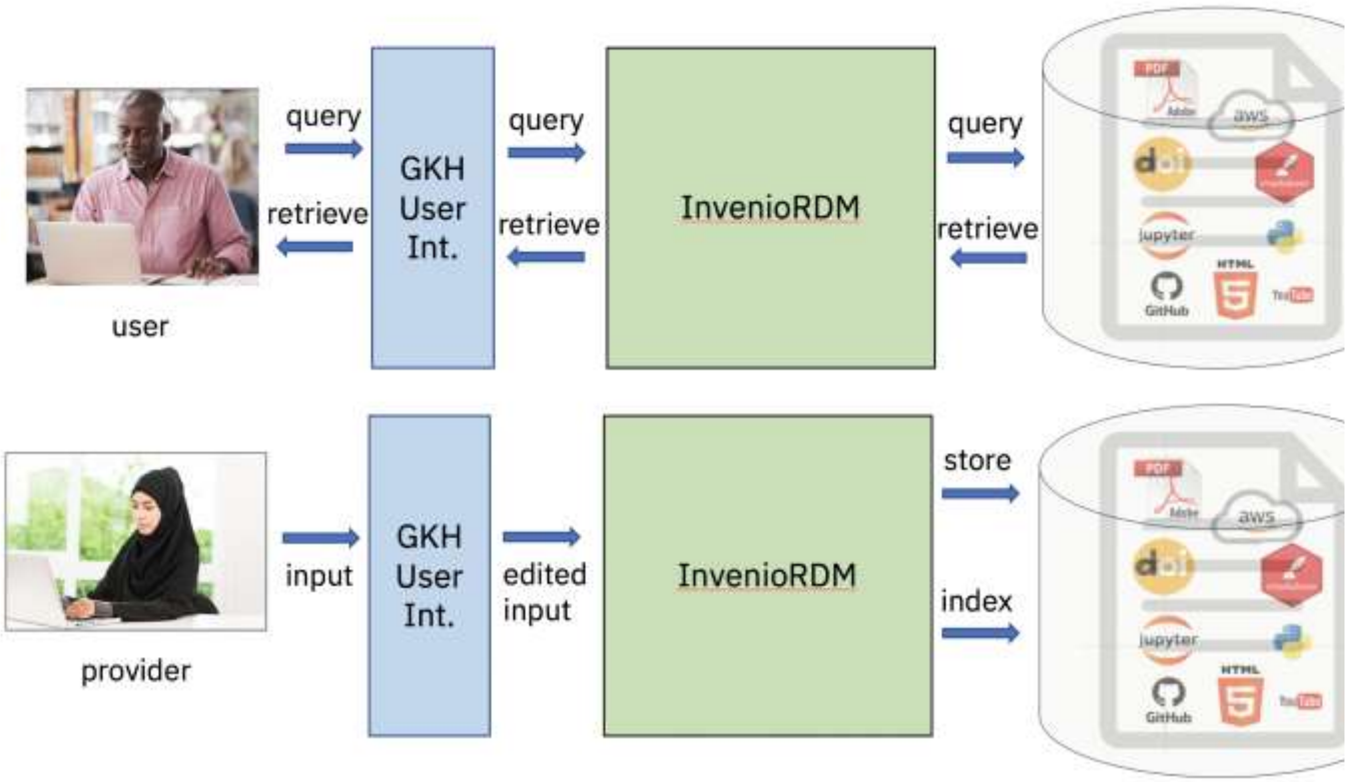

Figure 3 - Overall system design for the GEO Knowledge Hub

As illustrated in Figure 4, the GKH is connected to the GEOSS Platform Infrastructure, linked to the GEOSS Portal through an API to facilitate the integration of existing resources. It will be built on InvenioRDM (https://invenio-software.org/products/rdm/), using only free and open-source technology solutions. It will contain an independant search interface and provide digital library services to GWP activities.

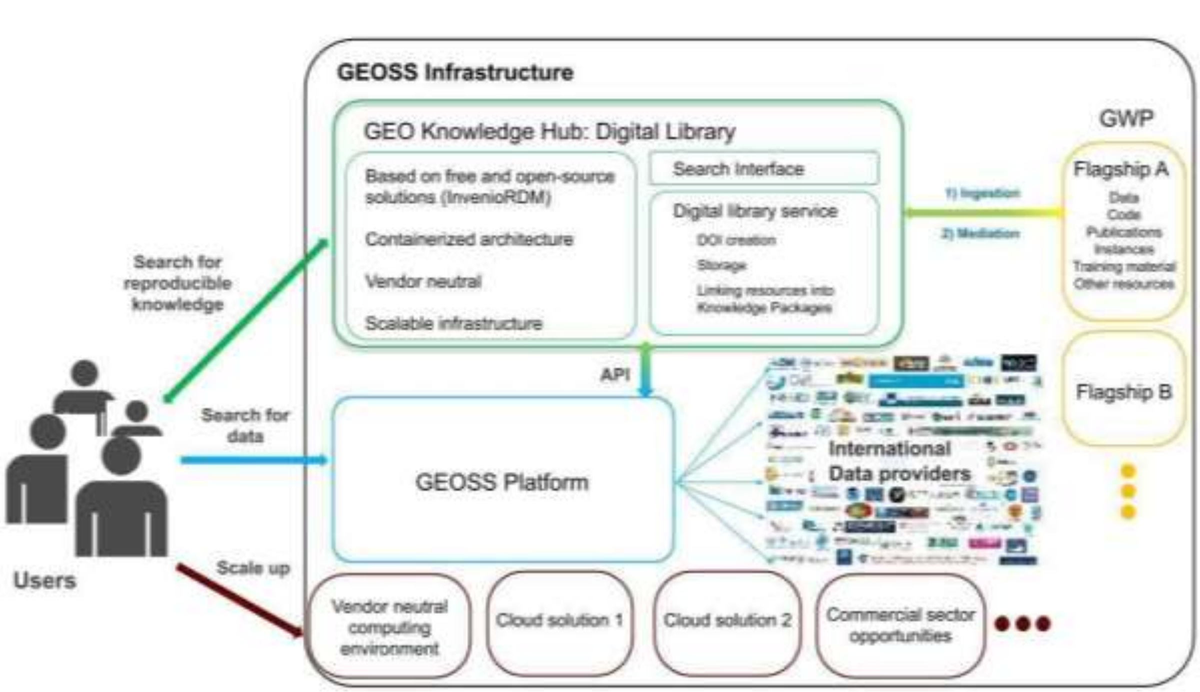

Figure 4 - Role of the GKH in the GEOSS Infrastructure

Hardware

The GKH will run in a cloud-based environment with an infrastructure designed to be portable and vendor neutral, as explained in the Principles and Assumptions above. It will be accessible through a public, fixed IP. The computational infrastructure sizing assessment is part of Task 1. The processing power, internal memory as well as the storage capacity for the GKH can be easily adapted to include the initially selected knowledge packages (Tasks 3 & 4).

Software

The GKH is being developed on top of InvenioRDM, a free and open-source framework for building large-scale digital repositories. The Invenio project is actively developed and maintained by CERN and a large user community. Invenio relies on the following stable, widely-used, open-source technologies (Technical details here:https://invenio.readthedocs.io/en/latest/general/requirements.html). The source code will be made available via Github and will use Python, Javascript frameworks, HTML and CSS. The main development tasks will be managed via the Github issues.

User Interface

The design of the user interface will be an ongoing process where the input of the actual knowledge providers (mediation/ingestion) and the actual users (search/discover/reuse) is used in a co-design effort to meet their needs. Note that the initial user interfaces (both on ingestion and search side) are loosely based on the proof-of-concept example presented in Canberra (using Zenodo). These efforts are more detailed in Task 1.

9. Outline of Major Tasks

In this section, descriptions of the major components required to implement the GKH are provided. They have been identified by the GKH team, the GEOSS Infrastructure Development Task Team and the GEO Programme Board as the optimal number and design to implement the initial instance of the GKH. These tasks may be modified or supplemented as the GKH becomes established, subject to experience and feedback from the GEO community. It is anticipated that any modifications to these tasks would happen on an annual basis, in consultation with the above-mentioned bodies and subject to approval by the GEO Executive Committee. Thus, the tasks cited herein constitute the body of work to be accomplished for the period July 2020 - July 2021, with an update containing any potential course-corrections for the period July 2021 - July 2022 to be submitted to the 2nd quarter Executive Committee meeting in 2021 for consideration, and so on.

Task 1: Implement GEO Knowledge Hub (GKH)

Objectives

Ensure that the selected GWP activities have access to the first version of the GKH (digital library) to share and discover EO applications (see tasks 3 and 4 below).

Activities

This task will establish a first version of a customized digital library based on InvenioRDM where knowledge packages from selected GWP activities will be featured (see tasks 3 and 4). Customization includes changes to the web interface, deployment of the platform in a cloud-based environment (with possibility for portability and adaptation of the computational infrastructure when needed).

It is not expected that the GKH will be a large digital library. Based on the current state of the art with respect to Earth observation data analytics, the number of documents to be stored and managed by the GKH will remain at a reasonable level. It is expected that there will be at most a few hundred relevant documents in the next three years. Github will be used as a sharing platform for people to access the code.

To ensure vendor neutrality, the GKH will be deployed in both local and cloud environments. After this initial test phase, the containerized platform can be easily deployed/transferred to a permanent hosting environment, for which the GKH team is actively seeking solutions.

During this stage, feedback from the knowledge providers and end users will be crucial to identify improvements to the user interface and the functionalities of the initial version of the digital library. All code related to the initial version of the digital library will be hosted on Github. In parallel, user documentation will be developed to describe the knowledge ingestion process, as well as the search features and a users' forum will be established.

Human/financial Resources

This task will be conducted by the GKH team. Key persons involved are:

- Rik Baeyens (IT Officer, GEO Secretariat)
- Florian Franziskakis (In situ Data Specialist, GEO Secretariat)
- Gilberto Ribeiro de Queiroz (Seconded expert, Brazil)

Criteria for completion

Task 1 will be considered as completed when the digital library is in a first and stable version used to support tasks 3 & 4. This means that the GKH website is operational, allowing providers to register knowledge packages, store resources and assign DOIs to them if needed, link to external resources such as code repositories. Completion of this task is also subject to users being able to search and retrieve these knowledge packages and resources.

Duration

Timeframe for completion of the task is expected to be 2-3 months after approval of the Implementation Plan by the 52nd meeting of the GEO Executive Committee (July 2020).

Task 2: Enhance GEOSS Platform Interface to the GKH

Objectives

The aim of this task is to integrate the information available in the GKH with the GEOSS Platform[13]. Its motivation is to create synergies by implementing interoperability between the GEOSS Platform and the GKH. It fits with the aims expressed by the GEOSS Platform team to expand the kinds of information retrievable by the GEOSS Portal interface. Besides providing access to descriptions and location of EO data, the long term aim of the

[13] As has been recognized by the EAG, the GEO Knowledge Hub is a complementary application to the GEOSS Platform, not an addition to it. The recommendation of the EAG, endorsed by the Executive Committee and by the GEO-XVI Plenary (Canberra), was to design the GKH as a new module of the GEOSS Infrastructure.

GEOSS Platform is to enable its users to have access to methods, algorithms, software and reports available in the GKH. This capability was demonstrated at the GEO Week 2019.

The general idea for this path of evolution of the GEOSS Platform is the concept of developing and delivering *knowledge packages*. A knowledge package is a consistent set of data, software, products and reports that supports an EO application. Organising a knowledge package requires combining the access to data with the access to the resources (models, algorithms, specifications, fit for use, results, etc.) required to make use of the data. It also requires that the resulting knowledge package be organised inside the GEOSS Platform so as to make it easily discoverable, understandable, accessible and usable.

On the basis of the current discussions between the GKH team and the team responsible for developing the GEOSS Platform, this task will be done under the responsibility of the GEOSS Platform team. Its development will use the API of the InvenioRDM platform to extract information from the GKH. To the extent possible, InvenioRDM supported open standards, such as OAI-PMH, will be used in preference to platform-specific API calls.

The GEOSS Platform currently harvests resources from Zenodo through its API. Zenodo will migrate to InvenioRDM during the second half of 2020. As the GKH will also be built on top of InvenioRDM, resources from the GKH can be harvested by the GEOSS Platform via similar API protocols as Zenodo given minimal adjustments. InvenioRDM is architected from the ground up to be resilient to changing business requirements, including stable APIs built on top of Zenodo APIs.

Activities

The activities in this task will be further developed by the GEOSS Platform team and will be included in future versions of this document.

Human/financial Resources

The development of the GEOSS Platform has been supported by the European Commission and the European Space Agency. The team includes specialists from ESA, CNR (Italy) and JRC. The financing of the activities under this Task will be subject to discussions between the European Commission, ESA, and the GKH team.

Key Persons

- Joost van Bemmelen, Guido Colangeli (ESA)

- Paolo Mazzetti, Mattia Santoro (CNR)
- Stefano Nativi (JRC)
- Gregory Giuliani (University of Geneva)
- Eldrich Frazier (USGS)
- Paola De Salvo (Data Officer, GEO Secretariat)

Criteria for completion

- The criteria for completion of this task will be proposed by the GEOSS Platform team.

Task 3: Identifying Reusable Applications from GEO Work Programme (GWP)

Objectives

As noted in the Introduction, reuse of EO applications is an essential requirement of GKH knowledge packages. Thus, a key feature of the GKH is that the EO application knowledge packages featured therein will be conducted in an open-science context, adhering specifically to the principle of reusability, as defined in Section 1 above.

Working closely with Task 4 (see below), this Task will promote and identify open science solutions with the "ingredients" and "recipes" needed to provide clear, step-by-step instructions for reusing a given application of EO, across the full information chain (from data, methodologies and algorithms, to computing environments and results). Part of this effort will include support for the GEO community to learn how to apply the applications in a "train the trainers" approach, for activities of the GWP that wish to progress from a local/regional scale up to a global one.

Activities

1. Open-science reusable EO applications developed by activities of the GWP, with competencies and resources contributed by GEO Members and POs as outlined in implementation plans approved by the Programme Board, will serve as the basis for populating the GKH. The GWP Flagships and Initiatives in particular are the main vehicles by which new knowledge is being developed in the GEO community. The primary work of the GKH team will be to highlight the GKH as an efficient means for sharing knowledge and scaling-up applications developed within the GWP, and thus providing motivation for the different GWP communities to be engaged with, and contribute to, the GKH with the resources that form the components of a given application. To this end, the GKH team will work with the

leadership of the GWP Flagships and Initiatives to establish a GKH point of contact who can serve as a Liaison Officer and keep the GKH team informed of each phase of application development within that activity. The GKH team will also ensure that the methodologies used in application development forming part of the knowledge packages for inclusion in the GKH have been properly vetted through a peer-review process.

2. The GKH team will also work with the commercial sector through periodic open calls to provide opportunities to the GEO community to leverage new technologies for data analytics and cloud computing that can contribute reusable, open-science applications of EO. Recent examples include, but are not limited, to the GEO-Amazon Web Services (AWS) EO Cloud Computing Programme, and the GEO-Google Earth Engine (GEE) Licensing Programme. The GKH team has also been in discussions with Microsoft, Mundi (Copernicus Data and Information Access Services), Sobloo/Airbus, Earth Observation Data Centre for Water Resources Monitoring (EODC), and SMEs such as Sinergise and Meteorological Environmental Earth Observation (MEEO) with regards to potential data analytics and cloud computing facility contributions.

   As outlined in Principle 6, Section 3 (Interface to proprietary APIs) the GKH distinguishes between cloud providers that work as **SaaS** (software as a service) such as Google Earth Engine from those that work as **IaaS** (infrastructure as a service) such as AWS. For solutions that work as **IaaS,** the GKH expects to store the full software as open source. In the case of **SaaS**, where cloud providers have a proprietary API to access their facilities, the GKH will store application code that uses the API, but the API itself. This is not an ideal solution, since it limits the generality of the software stored in the GKH, but it is a compromise that allows sharing of reusable software.

3. In addition to collecting open, reusable EO applications from the GWP, the GKH team will actively promote the use of open science solutions based on EO, and by extension, good use of the GKH, in a “train the trainers” approach to capacity development. Activities will include electronic fora to exchange best practices, webinars, presentations, and side-events in the context of regular GEO meetings, such as the GEO Symposium and Plenary.

Human/Financial Resources

The GKH will be primarily populated from applications developed by the GWP, in the course of activities already defined in implementation plans (including resources) and as approved by the Programme Board. Thus, the contributions of GEO Members and Participating Organizations to the GWP will form the major component of the human and financial resourcing for this task. Additional resources allocated to this task will include staff from the GEO Secretariat.

Other sources of financial resources could include commercial sector contributions, such as the GEO-AWS EO Cloud Credits Programme and the GEO-GEE License Programme. Additional external funding arrangements are under discussion at the time of this writing.

Key Persons

- Douglas Cripe (Senior Scientist, GEO Secretariat)
- Paola De Salvo (Data Officer, GEO Secretariat)

Duration

- This task constitutes part of the core work of the implementation of the GKH. As such, it is an ongoing task, since it is expected that applications and related documents will be included in the GKH on a regular basis. During the first phase of the GKH Implementation (through June 2021), we expect to demonstrate that the Task objectives can be fulfilled with the basis of the results from Task 1.

Criteria for completion

- A routine consultative procedure with the Flagships and Initiatives of the GWP has been established with the GKH team.
- A number of robust “cookbooks” using a variety of EO and computing platforms have been identified and used by the GEO community.
- A community for best practice exchange has been established (including commercial sector projects).

Task 4: Showcasing GEO Community Resources (pre-operational) in the GKH

Objectives

The GKH’s primary purpose is to amplify the GWP by sharing their results to the end-users. In this task, the GKH team will work in very close collaboration with the GWP activities identified in Task 3 (see above) to make sure that their knowledge and tools developed are discoverable and accessible to the global community in the GKH.

Activities

The activities on Task 4 will consist of assessing different user scenarios associated with the GWP, some of which have been outlined in Section 2 above. The idea is to start from concrete user needs (Sen2-Agri, GEO BON EBVs, Digital Earth Africa) and invest time together with the respective community, so that the process of storing information in the GKH proceeds along the following steps:

1) All the various knowledge components will go under a quality control process of the knowledge Provider and the GKH team to secure its effective discoverability, accessibility and usability.
2) The GKH team will work with the developers to make such tools available globally, either: (a) as a virtual machine that could run on any cloud service; (b) as reusable scripts that could run on services such as GEE.
3) The knowledge and tools made available will be complemented by ancillary essential documentation such training material, publication, case studies. The scenario above described will be applicable to knowledge elements and packages developed by the various GEO WP Activities.
4) A process of capacity development towards the knowledge providers in using and populating the GKH will also be part of this task, using dedicated virtual webinars as well face to face meeting.

Human/Financial Resources

The GKH will be primarily populated from applications developed by the GWP, in the course of activities already defined in implementation plans (including resources) and as approved by the Programme Board. Thus, the contributions of GEO Members and Participating Organizations to the GWP will form the major component of the human and financial resourcing for this task. Additional resources allocated to this task will include staff from the GEO Secretariat.

Key Persons

- Paola De Salvo (Data Officer, GEO Secretariat)
-

Duration

- This task constitutes part of the core work of the implementation of the GKH. As such, it is an ongoing task, since it is expected that applications and related documents will be included in the GKH on a regular basis. During the first phase of the GKH Implementation (through June 2021), we expect to demonstrate that the Task objectives can be fulfilled with the basis of the results from Task 1.

Criteria for completion

- Successful deployment of select EO applications from the GKH available via cloud services, verified by at least one external party.

Task 5: Improvements to the GKH

Objectives

Provide an open-source digital library designed to serve the community and to ensure that all components of reusable EO applications are visible, accessible, and where GWP Activities can share their knowledge and experience in the long term.

Activities

Based on feedback from the community (through tasks 1, 3 and 4), this task will be focused on creating a fully customized digital library that fits the GEO community needs to go towards Open Science. This extensive customization will be a continuation of our collaborative effort with the CERN InvenioRDM development team. This collaboration has been established and the GKH team is recognised by the CERN team as a relevant collaborator in the development of InvenioRDM.

Considering the development plan outlined by the InvenioRDM consortium, the functionalities of the GKH will be progressively expanded to expose new InvenioRDM capabilities to the GEO community. Among the expected capabilities to be included in future versions of InvenioRDM include quality control mechanism of the resources, full-text indexing and search of stored text documents.

As feedback from the community is of paramount importance, an integrated users community forum will be established (based on the StackExchange model). Contribution guidelines for developers and open-source sharing of the code will be maintained, as well as documentation.

Human/Financial Resources

In continuation of task 1, resources allocated to this task will include staff from the GEO Secretariat. Initial deployment of the GKH will take place in a cloud-based environment and be designed to have maximum portability. Currently, public cloud solutions are being sought for hosting the GKH as this would be the optimal environment in terms of transparency, neutrality and sustainability.

At the end of this task, a cost analysis will give a better understanding of the appropriate funding mechanisms to put in place for the long-term sustainability of the GKH.

Key Persons

- Rik Baeyens (IT Officer, GEO Secretariat)
- Florian Franziskakis (In situ Data Specialist, GEO Secretariat)
- Gilberto Ribeiro de Queiroz (Seconded expert, Brazil)

Criteria for completion

- This task will be considered as completed when the GKH is ready to be made publicly available, giving the possibility to all providers to make contributions, upon registration and through quality control mechanisms.

- Completion of this task also includes allowing public users and EO practitioners to search through the fully customized, operational web interface of the GKH and retrieve relevant resources[14].

Duration

- This task is expected to last until the end of the present implementation plan with regular updates (each 3 months).

[14] Upon recommendation by the EAG, the GKH search function should be independent from the DAB function of the GEOSS Platform, as extension of the DAB to form part of a digital library would be a more resource-demanding option than the current proposal.

10. Management Overview

The development of the GKH is part of the the GEOSS Infrastructure Development FT, outlined in the 2020-2022 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*."

As explained above, the GKH will be primarily populated with the constituent resources of EO applications developed by the GWP, in the course of activities already defined in implementation plans (including resources) and as approved by the Programme Board. The process for including applications in the GKH involves the participation of the leaders and participants of GWP activities, and the notion of "authoritativeness" will come from the experts/contributors involved in the activities of the GWP themselves, as well as proof of some form of peer-review by external experts in the field. Each GWP activity will be requested to provide one or more points of contact; these persons will act as trustees of the GKH; they will oversee the choice of which applications are stored and shared via the GKH.

At the same time, the GEOSS Platform team, as members of the GEOSS Infrastructure Development Task Team (GIDTT), is invited to work with the GKH team in implementing the GKH as a component of the GEOSS Platform. As envisaged in the GWP, the development of the GKH will be advised and monitored by the GIDTT, whose duties are to *"ensure coordination, and integration as appropriate, among the various components of the GEOSS infrastructure for management and sharing of data, information and knowledge resources*". Among the duties of this Task Team are: "*establish performance expectations for each component and means to measure their achievement, and monitor progress of each component at regular intervals against the performance expectations'*. As befits all FTs of the GWP, the development of the GKH will be reported periodically to the GEO Programme Board.

Therefore, the governance of the GKH implementation has been well established as part of the GEO structures, including development oversight by the GEO Executive Committee[15]. In particular, the current document provides input to the GEOSS Infrastructure Development Task Team and to the GEO Programme Board concerning the relationship of the GKH and the other components of the GEOSS infrastructure. More specifically, the performance expectations of the GKH and the progress monitoring milestones are

[15] Draft Report of GEO XVI (6.2)

discussed in Section 6 of this document.

Considering the actual implementation of GKH, the initial phase, covered by the current document (July 2020 - June 2021) assumes a minimal team that does not require additional resources than those listed in the Secretariat CONOPS, whose tasks and estimated percentage of time allocated are outlined in Table 3 below (please refer to the CONOPS document for more details). The expected dedication of the GKH team is about two persons full-time, which is equivalent to 15% of the total personnel allocation at the GEO Secretariat.

Table 3: Human Resources dedicated to implementation of GKH

ROLE	NAME	Roles
GEO Secretariat Director	Gilberto Camara	1. Continuously assess the status and progress of the GKH 2. Connect with Member States and Participating Organizations to identify ways that GKH can support their needs 3. Evaluate how well GKH is meeting its goals and objectives. Estimated time allocation: 10%
Senior Scientist	Douglas Cripe	1. Has primary responsibility for the Secretariat contribution to GEOSS Implementation FT. 2. Lead in planning and coordination of the GKH implementation. 3. Identify projects of the GWP to most likely produce results relevant to the GKH. Estimated time allocation: 15%
GKH Manager	Gilberto Queiroz (seconded from Brazil)	1. Lead the Secretariat team responsible for the implementation of the GKH. 2. Support the GWP to develop the relevant contributions of each to the GKH, including reusable research. Estimated time allocation: 25%

Data Officer	Paola De Salvo	1. Engage with activity of GWP to ensure the EO data, algorithms, models as outputs of their respective work plans are available discoverable and accessible in the GKH; 2. Contribute to the GKH by providing technical advice and guidance on content: essential EO data – algorithms and models for the various GEO Activities and GEO engagement priorities: Sendai, Climate SDGs; 3. Act as curator of GKH content in close collaboration with Content/Data providers. 4. Engage with the GEOSS Platform Team to ensure co-ordinations of actions and maximise collaboration. Estimated time allocation: 60%
In situ Data Specialist	Florian Franziskakis	1. Responsible for the development of the in situ data component of the GKH. 2. Engage GWP Activities, Member States and Participating Organizations to submit relevant in situ data and associated documentation to GKH. 3. Develop software modules for ingestion, indexing, querying and presentation of in situ data and related documents in the GKH. 4. Develop a database for in situ data in the GKH, based on open standards. 5. Manage and curate in situ data in GKH in accordance with GEOSS Data Management Principles. Estimated time allocation: 60%
Space Data Specialist	Vacant (duties to be fulfilled by other GEOSEC staff)	1. Responsible for the development of the space data component of the GK (*currently, these duties are filled by the in situ Data Specialist*). 2. Engage GWP Activities, Member States and Participating Organizations to submit relevant information on the use of space data and associated documentation to GKH (*currently, these duties are filled by the Data Officer*).
IT Officer	Hendrik Baeyens	1. Take part in the design and implementation of the GKH. 2. Responsible for management and maintenance of information technology solutions used in the GKH.

		3. Coordinate the use and maintenance of the GKH for relevant GEO documents. Estimated time allocation: 20%
Capacity Development Officer	Vacant (expect secondment from Australia)	1. Support the GEO community to use the new technologies enabled by the GKH. 2. Work with GEO Open Data Cube community to support institutions in Member States to better extract information available in the GKH. 3. Organize training sessions and webinars to improve the capacity of Member States and Participant Organizations to use the new technologies of big data, cloud computing, and machine learning; 4. Support capacity development at both individual and institutional levels for long-term initiatives of the GWP (e.g., Digital Earth Africa).
GEOSS Infrastructure Development Task Team	Team Members	5. Ensure coordination, and integration as appropriate, among the various components of the GEOSS Platform for management and sharing of data, information and knowledge resources.
GEO Programme Board	Representatives	6. Support the Secretariat with identification of GWP activities and knowledge packages suitable for inclusion in the GKH.

To ensure sustainability on the effort of the help desk role and user support, the help desk will be in format of virtual forum engaging the respective relevant domain experts This effort will be sustained and coordinated by the GKH team, and be supported by the GEO community.

11 Intellectual Property Rights, Security and Privacy

General Principles

As a design principle, all of the content of the GKH will adhere to the GEOSS Data Sharing and Management Principles[16], be available under an open source and open access license when possible. The responsibility for defining the specific license to be used is of the knowledge provider. If required, the GKH implementation team will provide support to the knowledge providers in understanding the differences between the licenses.

Software licenses

For software, providers will be recommended to use one of the licenses listed by the Open Source Initiative (https://opensource.org/licenses). There are three main flavors of open source licenses: "copyleft" ones such as Gnu Public License (GPL) (which restrict the use of open source code within a non-open environment), "permissive" such as the Massachusetts Institute of Technology (MIT) license (which has no restrictions for its use) and dedicated ones such as the NASA agreement (which protects the rights of the US government). The GKH team will do its best efforts to provide guidance to providers if required.

Other documents

For documents, the recommendation to providers is, whenever possible, adopt one of the Creative Commons licences: CC-BY (permissive, just asks for citation), CC-BY-SA (mandates that all derivative works be openly shared) , CC-BY-NC (forbids commercialisation) and CC-BY-SA-NC (combines the previous two restrictions). As in the case of software, the GKH will support users to choose which licence meets their needs. In any case (documents or software), there will be no transfer of intellectual property rights from their owners to GEO. Knowledge producers retain their rights.

When storing papers that have been published in scientific journals in the GKH, there are two possibilities. Either the papers are made available as open access or the journal has the copyright of the final version. Open access refers to papers where: (a) the journal policy is to allow free access to all papers published; or (b) the author has paid to the journal an additional fee to make the paper openly available. In the former case, there is no restriction for depositing the paper in the GKH. In the latter situation, GKH will abide by the journal's intellectual property rights (IPR) policy. For the majority of scientific journals that allow it, GKH will store the "post-print" version (the final version of the paper without the journal's letterhead and formatting).

Data security: individual and collective rights

In terms of data security, there are two considerations. The first one concerns the integrity

[16] As noted during the GEO Data Technology Workshop, 23-25 April 2019, the GEOSS Data Sharing Principles are more expansive than the FAIR (finadable, accessessable, interoperable, reusable) Principles.

of the information deposited in the GKH. For this purpose, the GKH Implementation Team will use the security system enabled by the cloud provider to make sure the information is replicated in different geographic locations and backups are made on a regular basis. The second concerns the rights of access to modify the contents of the GKH. Such access will be restricted. The user interface that will be made available for depositing data in the GKH will not provide direct access to the GKH database. The documents will be deposited by the user in a temporary storage. They will be revised and checked by the GKH management before being included in the main database.

As regards privacy, one should distinguish between individual and collective privacy. As regards personal privacy, the GKH is not expected to hold personal information that goes beyond what is already shared in documents and research papers: author's names and institutional addresses. For in situ data collection that involves personal data, as in the case of some citizen science initiatives, the GKH team will consult with legal experts and the WMO legal office to fully understand its responsibilities for hosting data produced by third parties, in the light of legislation such as the European General Data Protection Regulation.

One should also consider the case of collective privacy rights. For example, some Member States may be concerned that the GKH allows diffusion of methods and data that enable extraction of what such countries deem as sensitive. In this and similar issues, GEO has to refer to the appropriate international legislation, and especially to the 1967 Outer Space Treaty. In force since 1967, it has been ratified by 103 states. The treaty stipulates that all activities must, among other restrictions: be governed by International Law; provide free access to space with no interference from other states; only peaceful activities may be carried out; national governments are responsible for the activities of their enterprises. Furthermore, the UN Committee on the Peaceful Uses of Outer Space (COPUOS) has repeatedly recognized the value of EO as a means of providing essential information for sustainable development. Recent reports by COPUOS (e.g, AC.105/1014 from 2012 and AC.105/1115 from 2016) recognise the work of GEO and the importance of open provision of data and results from EO. Therefore, the GKH will only host data and results that are consistent with the accepted international practices and recommendations. To meet such requirements, data, documents, software, and methods stored and managed by the GKH have to be fully open and accessible to all interested parties.

In-situ data

Dealing with in-situ data is one of the most difficult challenges facing the GEO community. In the context of this discussion, in-situ data refers to the ground samples and measurements used to calibrate algorithms that extract information from satellite data, or initialize/validate model simulations. In most cases, the combination of in-situ and satellite data is a requirement for achieving good results with EO data.

Experience with machine learning methods has shown that the limiting factor in obtaining good results is the number and quality of training samples. Large and accurate data sets provide good results, no matter the algorithm used. Increasing the training sample size results in increased classification accuracy. This points to an key issue on the sharing of EO application: not all Member States are willing to share their in-situ data.

The choice of data policy for in-situ data does not affect the ability of Member States institutions to use cloud services. Using local in situ data in cloud computing facilities can be done in full compliance with the Member States chosen data policy. Member States will continue to own their data and retain full intellectual property rights, while at the same benefiting from the latest technological advances.

From the point of view of the GKH, a distinction needs to be made between two cases:

1. Applications that allow reproducibility: this is the case where users have deposited in GKH the full set of resources that allow others to duplicate the result;
2. Applications that allow replicability and/or reuse: this is the case where users have deposited in GKH the full set of resources minus the in-situ data. In this case, others can reuse the code to solve their problems or can try to replicate the result with a different set of in-situ samples for the same area.

As stated earlier, the main aim of the GKH is to facilitate the reuse of applications developed by members of the community. Our goal is to simplify the capacity development of Member States by improving the sharing of recipes that allow others to learn from one's experience. Therefore, we consider that the GKH can be developed to be compatible with different in-situ data policies.

The GKH will in no way interfere with the IP rights of data owners. Data deposited by providers in the GKH will be associated with a license chosen by the data owner. The GKH will not aim to enforce such rights. It is the responsibility of the data owner to legally pursue any legal action in case of violation of data policies.

12. Performance Monitoring

Unlike traditional information systems, whose performance is measured in terms of metrics related to system throughput (e.g., MFlops, transactions per second, downtime), the assessment of the GKH depends essentially on its content, and also, more broadly, on its impacts. In that sense, performance of the GKH will depend not only on the quantity of its contents, but also on the quality of its implementation. Therefore, we propose the following indicators be used to assess the performance of the GKH:

- Complete knowledge packages available in GKH (including all associated resources as mentioned above, such as papers deposited, shared scripts and software modules, datasets - both in situ and remotely sensed).

- Degree of reusability of knowledge packages contained in the GKH (based on GWP activity leadership, and also user feedback).

- Facility of use of GKH by national experts seeking to apply EO for informed reporting on policy commitments and/or solving environmental challenges related to sustainable development (based on user feedback).

- Evidence of uptake of GKH information as demonstrated in decision-making by GEO Member governments and other stakeholders.

As stated above, a knowledge package is a combination of data, software, documents, and methods which enables application reuse. It represents the "gold" standard for the contents of the GKH. However, not all contributions to the GKH will consist of knowledge packages. The inclusion of individual papers and of software modules is also relevant. For many contributors, it will be simpler to share resources comprising an application than it is to prepare a fully organized and documented knowledge package. Therefore, the GKH will be open to all relevant contributions.

Further, GKH usage will be monitored and statistics will be made available on a fully and open basis. Enabling open monitoring policies will provide a key reward element for communities regarding their contribution efforts to the GKH, as well as the benefits of adhering to the GKH as a principal channel for promoting and disseminating EO applications they have developed.

13. Future Perspectives: Dissemination and Exploitation

This section outlines the longer-term strategy for GKH uptake and exploitation by the GEO community, and provides the context for potential future activities of GKH implementation beyond the initial phase covered by this document.

The GEO community produces many useful and relevant results, which it aims to share globally. All successful EO applications have different components: policy mandates, national priorities, satellite and in situ observations, collaborators, methods and algorithms, and deployment strategies. Such resources will be of different types to reflect the entire information flow of the research and knowledge pertaining to a given domain.

The primary goal of the GKH is to advance transformation of EO data into knowledge-based services for evidence-based decision-making. The GKH is envisioned to provide linkages to policy mandates, the GWP with related project and task workflow alignment, task monitoring, and to implement dissemination of knowledge and technical expertise capabilities as integral part of a results-oriented GEOSS. Linked documents containing relevant information for Earth observation applications that provide access to the underlying data, models, methods and algorithms are necessary to promote dissemination of knowledge and technical expertise for reusability, scalability, and co-design/co-production. Since there is a substantial amount of new knowledge developed in the GWP, the effectiveness of the GKH depends on building strong links with the activities of the GWP. The GKH team will work with the Programme Board to identify opportunities for Flagships and Initiatives of the GWP to leverage the GKH in order to increase uptake of end-to-end services and applications they have developed, and amplify their reach globally. This will best be accomplished if a GKH focal point can be designated for each GWP activity that wishes to contribute to the GKH.

Additionally, the GKH team will liaise with GEO Members and Participating Organizations contributing to the GWP to identify experts distributed globally to form a virtual Knowledge Hub help desk. The help desk experts can serve as intermediaries with end-users, providing them guidance and support in making maximum use of the GEO Knowledge Hub and interpret results. In turn, users of the GKH will be invited to supplement the trainer community for EO topics, in an ever-widening feedback loop.

Further, delivery of fit-for-purpose country-relevant knowledge-based products and services will be a key measure of the success of the GKH implementation. This may be accomplished through the leverage of existing, or establishment of new, Communities of Practice for dissemination of knowledge and technical expertise.

In parallel, the GKH team will take active action to promote Open Science in EO and, by extension, good use of the GKH. This is a multi-front approach that requires, among others, the following actions:

a) *Working with funders:* Much of EO results are produced using research grants from public agencies. GEO will engage with these agencies to highlight the importance of getting the most impact out of public funds, and to show that full reusability of results benefits everyone.

b) *Incentive-building*: GEO will highlight those groups that are actively sharing their data and methods and promote their work in the community. This can be done using GEO events, newsletter, webinars and other communication activities.

c) *Building software repositories*: GEO will organize the information provided in the Knowledge Hub to maximize reuse. Following the example of Bioconductor[17], it is feasible to organise the software available to make it easier for users worldwide to find out packages that are useful to them. Other relevant documents (e.g., training material) will be stored directly in the GKH.

d) *Capacity-development*: As a general principle, GEO intends to leverage the capacity development networks of its partners (e.g. GPSDD, ITC, RCMRD, SDSN and others) in a “train the trainers” approach to disseminating this knowledge. GEO will work with its Members and Participating Organizations to organize capacity development activities and postgraduate diploma courses that are required to bring about a new generation of users that are familiar with big EO data analysis. Work with local academic institutions to co-produce resources which explain the value and contribution of EO to decision-making and policy in specific contexts. The GKH team could even explore options for bringing select individuals to institutions with expertise in capacity development for intensive training.

Finally, GEO, through its a large network of Members and Participating Organizations, can strengthen partnerships with official statisticians, Government Ministries and other national and local data users to raise their awareness of the potential of EO for SDGs, and encourage a collaborative ecosystem approach to national data production for policy compliance. GEO Members and Participating Organizations would be asked to help identify agencies within governments or universities and other affiliated agencies that can provide assistance within the country. In some cases, the entry points may be within agencies themselves, for example, most statistical agencies have geospatial divisions that may be good engagement points since they understand both statistics, sampling, etc. for national statistics as well as the potential for EO.

[17] https://www.bioconductor.org/

14. Risk Assessment

In this section, we consider risk factors associated with the development of the GKH, assess their impact and consider mitigation measures.

1. **InvenioRDM project continuity**: CERN and its partners could decide to suspend or terminate the InvenioRDM project, which has been chosen as the basis for the GKH.
2. **GKH User Interface development**: The GKH development team fails to deliver a user interface that is considered to be efficient and easy to use by the GEO community.
3. **Lack of interest by the knowledge providers:** The GEO community fails to respond to the opportunity to collaborate with results that would be shared using the GKH.
4. **Lack of interest by the prospective users of the GKH:** The prospective users of the GKH consider that the hub does not provide useful information to their projects.
5. **Hosting of GKH in the cloud:** The implementation of the GKH may become vendor dependent.
6. **Long-term evolution of GKH:** As the GKH grows, so will the work of maintenance, improvement and promotion of the system.

The first case (**InvenioRDM project continuity)** refers to a situation where CERN and their partners decide to halt the investment on the development of the InvenioRDM. The primary user of the InvenioRDM is CERN itself, an institution that has a substantial demand for management of research data. Given the large amount of data generated by the experiments at CERN, and the global reach of the researchers that use such data for their work, CERN needs a system such as InvenioRDM. Given the size and variety of data handled by CERN, it is unlikely that InvenioRDM could be replaced by the current generation of commercial data management platforms. These commercial platforms are mostly targeted for administrative purposes such as handling emails and company documents, and doing product and client management. Arguably, only the big Internet companies (such as Google or Amazon) would have the capability of developing a platform that meets the needs considered by InvenioRDM. Taking into account the international nature of CERN and the need to ensure data protection and IP rights enforcement, it is unlikely that CERN would be willing to replace an open source solution for a commercial product from one such big company. Therefore, we consider the likelihood of the interruption of InvenioRDM project to be **low.**

The mitigation action proposed in the case of interruption of the InvenioRDM project is to revert to the use of the Zenodo platform. Zenodo is an existing and stable platform that was tested and validated in the Proof of Concept of the GKH presented at the GEO Plenary in Canberra. Although Zenodo does not have all facilities that will be implemented in the InvenioRDM platform, it supports the GKH requirements 2-15 outlined in Section 4 of this document. The only missing requirement is full text-based search. Therefore, we consider using Zenodo to be a satisfactory solution as a mitigation option.

The second case (**GKH User Interface development**) refers to the possibility that the GKH team who is tasked with developing the user interface of the GKH as a front-end to the InvenioRDM software fails to achieve a satisfactory result. As it happens, the GKH team (including the seconded expert from Brazil) has already engaged in workshops and teleconferences with the InvenioRDM team and has already built draft versions of the GKH user interface. Based on the current experience, there are no major impediments for a swift implementation of the user interface. Therefore, we consider the likelihood of a faulty development of the GKH use interface project to be **low.**
The mitigation strategy for an unlikely failure to deliver a working user interface to the GKH will be to hire a short-term consultant to do the job. It is expected that a good quality user interface to InvenioRDM can be developed by a qualified person in three months.

The third case (**Lack of interest by the result providers**) considers the situation where the GEO community fails to contribute to the GKH by providing reusable results. Since the stages of maturity of the GWP activities are quite uneven, there are currently few initiatives that are actively producing results that can be shared. There are already initiatives that are expected to contribute (some of them listed in Section 2, User Scenarios); these are considered to be sufficient to provide enough results to be shared by the GKH in the first few years of its operation. From that point onwards, the challenge will be for the GEO community to full recognize the benefits of sharing and contributing to the advancement of the global EO data usage. However, until the present moment, it has been challenging to implement a strategy for sharing results from the GWP more broadly as an element of a results-oriented GEOSS.   Therefore, we consider the risk of lack of interest by the result providers in developing and sharing results to be **moderate**.

The mitigation strategy for the lack of interest by results providers, a serious issue, would involve demonstrating the benefits/impacts generated by uptake of GKH resources at GEO events, in order to promote the added value the GKH brings. Community feedback would also be essential, and solicited at such events. Further engagement and development of the GKH could possibly be based on assessments such as a cost/benefits analysis. In addition, a re-examination of relevant GEO governance structures may be warranted. In parallel, the GKH team could engage with the global EO community for shareable results that can be used to populate the GKH.

The fourth case (**Lack of interest by the prospective users of the GKH**) would reflect a situation where those that could benefit from the information available in the GKH would not be in a position to use or understand such information. This would be typically the case when the prospective users lack the appropriate expertise to use the solutions shared in the GKH. Possessing adequate know-how for using the new technologies of cloud computing, machine learning and data cubes is a requirement in order to make maximum use of the capacities that big EO data affords. Since the understanding of how to use these technologies is not widespread, we consider the risks associated with lack of interest by prospective users to be **moderate**.

To develop mitigation strategies to deal with the fourth situation, GEO needs to put in place a significant effort in capacity development. A key element of this effort is the production of material that complements the reusable results available in the GKH. For example, it will not suffice simply to share in the GKH a useful product such as the Sen2-Agri; there needs to be detailed documentation on how to use it in a step-by-step approach. Additional documentation on the technical basis for Sen2-Agri (also in a didactic fashion) would be an important aid for prospective users. Similar approaches would be required for most of the applications made available in the GKH. This type of risk can also be viewed as presenting an opportunity. By exercising the process of sharing EO results, the GEO community will grow and become stronger.

The fifth case (**Hosting of GKH in the cloud**) refers to the situation in which the cloud environment that will be used to implement the initial version of the GKH will create a vendor dependency. Considering that the GKH will be developed based on portable software and open formats and standards, we consider that the risk of such dependency is **low**.

The mitigation actions to avoid vendor dependency are taken at the design and implementation stage. The GKH team is working carefully to ensure that there will be no vendor dependency at any stage during the implementation of the GKH.

The sixth case (**Long-term evolution of GKH**) points to the risk that the the GKH team will need to have a continued involvement in the actions of data mediation, ingestion and quality control of the GKH; it also considers the time to be used by the GKH team to help prospective users benefit from the resources available in the GKH. The assessment of this risk depends on whether the concept of *results-oriented GEOSS* is fully adopted by GEO. If that is the case, it becomes natural to consider that one of the functions of the GKH team is promoting reusable results. In fact, there is no other body in the current GEO governance structure that could perform this task better and more efficiently than the GEO Secretariat. Therefore, considering that GEO embraces the concept of *results-oriented GEOSS,* maintaining, improving and promoting the use of the GKH becomes an essential part of the work of the Secretariat. In this context, the risk associated with the commitment of the GEO Secretariat to the long term maintenance of the GKH is **low**.

The mitigation strategy for the risk of involvement of the GEO Secretariat in the long-term evolution of the GKH has to be based on a serious assessment of the role and mission of GEO. If GEO fails to support its Member States (especially developing nations) in the use of EO for decision-making, arguably it needs to be reformed and reorganized. The GKH is part of a bigger strategy of reaching out to the full membership of GEO to promote the use of EO. Therefore, it is to be expected that there will be a dedicated team whose full-time job is to improve and promote the GKH. Eventually, the GEO Executive Committee, considering the opinion of the GEO Programme Board, could decide that a dedicated team outside of the GEO Secretariat and GKH team has to be given the responsibility for the GKH. Whatever the case, the GKH team will have to be fully committed to work actively with the GEO Work Programme and the GEO community at large. It needs to be a highly dedicated team, whose primary responsibility is not software development but capacity development.

The risk assessment is summarized in Table 4 below.

Table 4 - Risk Assessment and Mitigation Strategy

Risk	**Likelihood**	**Mitigation Strategy**
InvenioRDM project continuity	Low	Use Zenodo (existing solution)
GKH User Interface development	Low	Hire a consultant
Lack of interest by the result providers	Moderate	Focus on demonstrating added-value of GKH to GEO community.
Lack of interest by the prospective users of the GKH	Moderate	Strong investment in capacity development
Hosting of GKH in the cloud	Low	Avoid vendor dependency from day one
Long-term evolution of GKH	Low	Embrace the concept of results-oriented GEOSS as one of the missions of the GKH team

Glossary

API	Application Programming Interface
AWS	Amazon Web Services
CC	Creative Commons
CCS	Cascade Style Sheets
CERN	European Organization for Nuclear Research
CNR	National Research Council (Italy)
CONOPS	GEO Secretariat Concept of Operations
COPUOS	UN Committee on Peaceful Uses of Outer Space
DAB	Discovery and Access Broker
DOI	Digital Object Identifier
EO	Earth Observations
FT	Foundational Task
GEE	Google Earth Engine
GEO	Group on Earth Observations
GEOSS	Global Earth Observation System of Systems
GIDTT	GEOSS Infrastructure Development Task Team
GKH	GEO Knowledge Hub
GPL	Gnu Public License
GPSDD	Global Partnership for Sustainable Development Data
GWP	GEO Work Programme
HTML	Hypertext Markup Language
IaaS	Infrastructure as a Service
IP	Internet Protocol
IPR	Intellectual Property Rights
ITC	International Institute for Geo-Information Science and Earth Observation
JRC	Joint Research Council of the European Commission
MIT	Massachusetts Institute of Technology
ODC	Open Data Cube
OGC	Open Geospatial Consortium
RCMRD	Regional Centre for Mapping of Resources for Development
RDM	Research Data Management
SaaS	Software as a Service
SDG	Sustainable Development Goals
SDSN	Sustainable Development Solutions Network
UN	United Nations
UNCBD	UN Convention on Biological Diversity
UNCCD	UN Convention to Combat Desertification

UNFCCC	UN Framework Convention on Climate Change
USGS	United State Geological Survey
WMO	World Meteorological Organization

References

2020-2022 GEO Work Programme http://earthobservations.org/documents/geo16/GEO-XVI-7.2_2020-2022%20GEO%20Work%20Programme.pdf

Amorim, R., Castro, J. Silva, J., Ribeiro, B. "A comparison of research data management platforms". Universal Access in the Information Society, June 2016, 16(4) DOI: 10.1007/s10209-016-0475-y.

Draft Report of GEO XVI.

GEOSS Evolve Team, 2018. *Challenges to be addressed in evolving the Global Earth Observations System of Systems.*

GEO Rules of Procedure (March 2019 update) http://earthobservations.org/documents/GEO_Rules_of_Procedure.pdf

GEO Secretariat Concept of Operations (CONOPS) http://www.earthobservations.org/documents/excom/ec48/ExCom48-06_rev1%20GEO%20Secretariat%20Concept%20of%20Operations.pdf

GEO Strategic Plan 2016-2025: GEOSS Implementation http://earthobservations.org/documents/GEO_Strategic_Plan_2016_2025_Implementing_GEOSS.pdf

Kitzes, J., D. Turek, and F. Deniz, 2018: *The Practice of Reproducible Research: case-studies and lessons from the data-intensive sciences.* University of California Press.

Mauser, W., G. Klepper, M. Rice, B. . Schmalzbauer, H. Hackmann, R. Leemans, and

H. Moore. “Transdisciplinary Global Change Research: The Co-Creation of Knowledge for Sustainability.” Current Opinion in Environmental Sustainability, Open issue, 5, no. 3 (September 1, 2013): 420–31. https://doi.org/10.1016/j.cosust.2013.07.001.

Proposed Design and Proof of Concept of the GEO Knowledge Hub http://earthobservations.org/documents/geo16/GEO-XVI-6.2(Rev1)_Proposed%20Design%20and%20Proof%20of%20Concept%20of%20the%20GEO%20Knowledge%20Hub.pdf

Results-Oriented GEOSS: A framework for transforming Earth observation data to knowledge for decision making http://earthobservations.org/documents/geo16/GEO-XVI-6.1(Rev1)_Results-Oriented%20GEOSS_A%20framework%20for%20transforming%20Earth%20observation%20data%20to%20knowledge%20for%20decision%20making.pdf

Semertzaki, E., 2011: *Special Libraries as Knowledge Management Centres*. Elsevier B.V.

Social, Behavioral, and Economic Sciences Perspectives on Robust and Reliable Science, 2015: *Report of the Subcommittee on Replicability in Science Advisory Committee to the National Science Foundation Directorate for Social, Behavioral, and Economic Sciences.* National Science Foundation.

Steed, A., and M. F. Oliveira, 2010: *Networked Graphics: Building Networked Games and Virtual Environments.* Elsevier B.V.