Summary Report of the First Meeting of the GEO Expert Advisory Group

This document is submitted to the Program Board for information.

GEO Secretariat, 6-7 September 2018

6 September 2018

Following a round-table of introductions, the current state of the Global Earth Observation System of Systems (GEOSS) was reviewed by Stefano Nativi (CNR) and Joost van Bemmelen (ESA). Starting from the description of GEOSS as contained in the *GEO Strategic Plan 2016-2025: Implementing GEOSS*, they noted that GEO has devoted considerable efforts to building the GEOSS infrastructure and capabilities that have made Earth observations (EO) discoverable. The currently resource-oriented GEOSS presently links more than 178 open data catalogs and information systems, comprising over 414 million data and information resources. In order to become a more results-oriented GEOSS, its infrastructure, based on open platforms, will need to evolve towards better serving decision-makers and the general public through improving the accessibility and usability of EO resources and delivery of knowledge-based products and services. Different governance models can be envisaged, ranging from collaborative, to acknowledged, to autonomous, each implicating different styles of interaction among the various constituents. Additionally, different architectural scenarios could be implemented, including star, mesh and silos models having different degrees of GEOSS component centralization. The star or mesh models, would likely best support GEO constraints for future development of GEOSS, and maintaining GEOSS as a System of Systems based on interoperability through Architecture Implementation Pilots (APIs) with existing and new analytics systems and global platforms**,** and the sharing of data and knowledge will continue to be paramount.

Gilberto Camara (GEO Secretariat Director), introduced his rationale for transforming GEOSS from a meta-data to a knowledge hub. In his estimation, the current design of GEOSS is outdated as the research/end user paradigm of searching and downloading EO is obsolete. He challenged the Expert Advisory Group (EAG) is to consider what should GEOSS be and re-design it, regardless of what now exists. Once this has been established, the task will be to determine how much of current GEOSS system is consistent with the re-design. More generally, he explained how GEO is currently functioning as a boundary organization in that knowledge and action are generally treated independently from each other. GEO should rather serve as a catalyst for activities that are co-designed and co-produced, a feature that has been lacking until now in activities of the GEO Work Programme, even though recognized by the *GEO Strategic Plan 2016-2025: Implementing GEOSS*. This would begin to remedy the “knowledge-action” divide and point the way to a results-oriented GEOSS. Further, he argued that there is still a need for public services, as commercial services are often seen as being too expensive. This is particularly true in terms of supporting mandates of the United Nations.

Caroline Richter (GCOS) and André Obregón (GEO Secretariat) spoke on requirements for EO to support the Paris Climate Agreement. Richter noted that a wide range of observations are needed for supporting global policy, and many datasets can serve more than one purpose. The Paris Agreement, for example, has many observational requirements, from monitoring its aims to supporting many of its actions such as adaptation, mitigation, and early warning. Activities such as the Global Stocktake, an essential first step in complying with national reporting under the Agreement, rely on the best scientific assessments available, from a variety of platforms. The Architecture for Climate Monitoring from space encourages collaborative, coordinated activities between the world's major space agencies in the area of climate monitoring with the overarching goal to improve the systematic availability of Climate Data Records. Thus, the global observation community needs to work together to support the Paris Agreement and provide reliable information as a tool. More broadly, the Global Climate Observing System (GCOS) 2016 Implementation Plan recognises the importance of supporting adaptation and mitigation, and monitoring the 3 climate cycles: water, energy and carbon. In particular, GCOS is calling for a Surface Reference Network to improve long-term accuracy, stability and comparability of in-situ observations, in order to achieve simultaneous, coordinated high-quality observations of many Essential Climate Variables (ECVs) as well as provide reference data to constrain and calibrate spatially comprehensive observing systems. Obregón noted that GEO wishes to support the UN Framework Convention on Climate Change (UNFCCC) through GEOSS without necessarily becoming a formal part of its structure. In addition, GEOSS could be used to address several action items from the GCOS 2016 Implementation Plan, including: hosting a knowledge base for guidance and best practice for adaptation observations; providing facilitated data discoverability and access; helping develop a carbon monitoring system; and supplying a mechanism for improved coordination of terrestrial observations.

After reviewing several recent natural disasters and their consequences, such as the 2017 North Atlantic Hurricane season being the costliest on record ($330 billion in losses, MunichRe), James Norris (GEO Secretariat), noted that the Sendai Framework for Disaster Risk Reduction 2015-2030 (adopted at the 3rd UN World Conference in Sendai, Japan, March 18, 2015) placed the focus on disaster *risk* management, as opposed to disaster management. This means adapting a strategy that seeks to reduce existing risks while preventing new risks and strengthening resilience. Objective and measurable Outcome Targets under the Sendai Framework provide the means for international benchmarking of progress relative to a quantitative baseline 2005-2015. EO and geospatial information can significantly reduce the cost burden to countries as they seek to achieve baseline information and comply with the Sendai Targets. The UN Economic and Social Council (ECOSOC) has adopted a Resolution to ensure that geospatial information can be available to those who need it, at the time they need it, and GEO is working to support implementation of Sendai Framework targets E,F and G through engagement with the UN International Strategy for Disaster Reduction (UNISDR). More generally, EO can support all stages of disaster risk management cycles (prevention, preparedness, response, recovery). GEO is already contributing to disaster risk reduction efforts through activities of its Work Programme, including the Data Access for Risk Management (GEO-DARMA) Initiative, the Global Wildfire Information System (GWIS), the Global Agricultural Monitoring (GEOGLAM) Flagship, and the Geohazard Supersites and Natural Laboratories (GSNL) which provides near real-time data information in response to seismic activity globally. Where GEOSS can play an augmented role as knowledge hub would be in assisting national governments as they report to the UN on their progress against measuring and reducing disaster risks. GEOSS can serve as a an open and collaborative platform which utilises reliable trusted data to support national progress reporting and allows scaling from local to global levels. Integration and analysis of complex risk models is crucial to the success of national reporting, and GEOSS could further provide the mechanism to integrated both space-based and in-situ data for incorporation into different risk models.

Jessica Espey (SDSN) highlighted four key elements of the 2030 Agenda which have profound implications for how data, specifically EO data, can be used for the fulfilment of the Sustainable Development Goals (SDGs). These included: using more data, of higher resolution, so as to better identify the poorest and most vulnerable (Leave No One Behind); ensuring data are systematically informing policy and decision-making (Evidence (Data) Based Policy); employing a placed-based approach to implementation, as encapsulated by Goal 11 (Sustainable Cities and Communities); using geospatially-explicit information to minimize trade-offs between goals e.g. agriculture/water provision and ecosystem preservation; and pursuing the "…transparent and accountable scaling-up of appropriate public-private cooperation to exploit the contribution to be made by a wide range of data, including EO and geospatial information, while ensuring national ownership in supporting and tracking progress" (Harness the data revolution). She noted that, with the current situation, the mandate for monitoring the SDGs falls principally with National Statistics Offices (NSOs), which are governed by Chief Statisticians, and dominated by statisticians with expertise in demography, household surveys and national accounting, but very little in the way of EO. Indeed, the use of EO and geospatial data is generally limited as NSOs are often distinct from national mapping agencies, and GIS capacity is usually scattered across government with minimal coordination. The amount of data required for monitoring the SDGs, coupled with a lack of EO and geospatial information expertise and no budgetary increases, places a huge burden on the NSOs. Espey concluded with five recommendations for GEO, having implications for the re-design of GEOSS:

1. Focus on strengthening 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 SDG data production;
2. Work with local academic institutions to co-produce resources which explain the value and contribution of EO to decision-making and SDG policy in specific contexts;
3. Support local academics to be the technical advisers to government and help them identify entry points within the government policy-process / data production cycle;
4. Build international coalitions of supply side data producers and demand-side data users, to identify the data needs for different SDG pathways, and model these pathways in different countries using EO data; and
5. Advocate continued investment in EO capacity and products, including global public goods such as Landsat, which are fundamental for low capacity countries.

Michael Obersteiner (IIASA) and Fernando Ramos (INPE) observed that policy making in a complex world needs more than ever to be based on science. Obersteiner stated that both structural (geography, time, sector) and dynamic (volatility, transformation, deep uncertainty) complexity are contributing factors. He explained the concept of data-rich algorithmic policy making to address complexity, which entails an assessment of the economic, political and social instruments used as inputs into a matrix analysis of transformations needed across various sub-system domains (energy, food and biosphere, cities, sustainable consumption and production, digital revolution, human capacities and demography) when ascertaining the best ways to inform decision making in a given target arena. An example would be the information requirements for SDG monitoring and application frameworks. Taking an algorithmic approach to policy making allows optimal combinations of EO and big data, artificial intelligence, crowd sourcing, response prediction, and political analytics as opposed to the manual approach which relies on literature surveys, public statistics, and simple expert systems. He further argued that policy planning for global targets was best accomplished when nationally co-produced and co-designed to be reproducible. EO are an essential component of the solution, and GEOSS must help simplify processes required to comply with policy. Ramos added that interdisciplinary models, such as GLOBIOM, which make use of a wide variety of EO data and information, can be very persuasive and thus useful to inform and change government policy. Access to, guidance on usage of, and information from these types of interdisciplinary models should form an essential holistic functionality of the GEOSS re-design.

Baudouin Raoult (ECMWF) reviewed the use of cloud computing services at the ECMWF. Noting that, according to Wikipedia, "cloud computing relies on sharing of resources to achieve coherence and economies of scale, similar to a public utility", he explained the various layers cloud computing can provide, including Saas (Software as a Service), PaaS (Platform as a Service), and IaaS (Infrastructure as a Service). In the case of Copernicus, the ECMWF provides the Copernicus Climate Data Store (CDS), an on-premises private cloud serve providing improved access to observations, global and regional climate reanalyses, global and regional climate projections and seasonal forecasts through a unified web interface. The CDS includes an authoritative set of software ("CDS Toolbox") that allows users to develop applications based on content of the CDS. Working with partners such as EUMETSAT and Mercator, the CDS forms part of the Copernicus Data and Information Access Service (DIAS). Additionally, the European Weather Cloud (ecCloud) provides virtual machines that allow users to run data mining algorithms, perform downscaling and offer services to their own end-users directly from the virtual machines. He concluded by noting the power of ECMWF cloud computing and analysis tools are available to all with a reasonable internet connection, thus confirming that cloud computing could serve as a viable option for GEOSS in solving societal challenges.

Noel Gorelick (Google) demonstrated the Google Earth Engine (GEE) which combines a multi-petabyte catalogue of satellite imagery and geospatial datasets with planetary-scale analysis capabilities and makes it available for scientists, researchers, and developers to detect changes, map trends, and quantify differences on the Earth's surface. He noted that while GEE affords open access, it is itself a managed environment and thus is not open source. The GEE follows the "freemium" model based on top of the Google cloud which allows a basic degree of utilization, but extra functionalities will entail a fee. Automating image classification through Convolutional Neural Networks (CNN) is another feature that can make use of the Google cloud. Instead of pre-processing the data to derive features such as textures and shapes, a CNN takes just the image's raw pixel data as input and "learns" how to extract these features, and ultimately infer what object they constitute. GEOSS as a knowledge hub could provide added value to users in providing data aggregator functions along with guidance on the use of technology such as CNN.

7 September 2018

Edzer Pebesma (University of Münster) presented his views on reproducible science, cloud-based observation data processing, and open EO. He defined reproducibility as the capacity to reach identical outcomes based on shared data, software and computational capabilities. Informing decision-making with EO involves Information that is derived from data through analysis, which involves computation. Reproducibility is an essential feature of the knowledge process because it allows us to place confidence in information that comes from a shared understanding of what the data are, how analyses are performed, and which computations are involved. Of course, computations involve software, which gains trust through repeated use, verification and benchmarking, and access to the code itself. He cited the open data cube as file-agnostic means to access to EO imagery, which can boost usability of EO data since spatial dimensions are complemented with other dimensions such as temporal or spectral, and researchers can directly alter, aggregate, or map functions over dimensions of a user-defined cube without being concerned about how the data in the processing platform is organised. His suggestions for a results-oriented GEOSS included:

1. providing (develop, share access to) compute nodes where big EO data can be sensibly analysed (i.e. not by tile, but by image collection);
2. providing a user-defined data cube view;
3. implementing simple methods (open source, or otherwise open API) to perform computations;
4. developing benchmarks showing that these computational nodes produce the same results given the same queries (openEO);
5. developing best practice executable documents (e.g. Jupyter notebooks) showing how one could use given software/service infrastructure to answer questions (openEO); and
6. creating a sustainable ecosystem of users to share knowledge.

Matt Hansen (GLAD/University of Maryland) emphasized that there was a great need for concurrent views and harmonized approaches nationally when it came to tackling issues such as forest inventory and land cover. Often these definitions become overshadowed by political concerns (e.g., some countries consider replacing a diverse forest with monoculture does not constitute deforestation). He also pointed out that perfect maps do not exist, rather the issue is the degree of accuracy to be in order to be useful for reporting purposes. Although satellites may provide a first approximation for quick location of items of interest, on-the-ground reference data is still critical for verification purposes. Especially when maps are used as indicator products, efficient collection of reference data must be highest priority. He called on GEO to rethink the balance of data coming from its Members and Participating Organizations and provide better access to in-situ data (GEO is not just about remote sensing).

Douglas Cripe (GEO Secretariat) traced the mandate of GEO with respect to in-situ EO networks across key guidance documents, including the initial *GEOSS 10-Year Implementation Plan*, the *GEO Strategic Plan 2016-2025: Implementing GEOSS*, and Ministerial Declarations such as the one from Mexico City (2015). For GEO, in-situ observations are seen as fundamental to support process studies, satellite data validation, and algorithm and model development, as well as the detection, documentation and attribution of change. Further, the Mexico City Declaration highlighted the fact that GEO should respond to the needs of global policy implementation, including the 2030 Global Goals for Sustainable Development, the Sendai Framework for Disaster Risk Reduction 2015-2030, the United Nations System of Environmental and Economic Accounts, and the UNFCCC. He then gave several examples of where GEOSS could support the in-situ observational needs of GEO Members and Participating Organizations (e.g.: GCOS, UNGGIM, EC-Copernicus, WMO, WHO) and other activities of the GEO Work Programme, as well as serve as a repository for legacy datasets generated by projects with limited lifespans (e.g. Limnades dataset for in-situ water quality). Finally, he remarked that GEO could play a role with respect to collecting, curating and providing access to citizen science, helping boost use of these types of data which have the potential to be much more greatly exploited than is currently the case. He concluded by noting that GEO (through GEOSS) has an essential global role to play by becoming the trusted, curated world-wide repository for authoritative in-situ data in the service of its Members and Participating Organizations.

Break-out group summaries:

(a) Knowledge production and organization

- Good term to capture the intent of this discussion: “pathways to co-design and coproduction”.
- At the broad conceptual level, the idea of co-design/co-production has been long embedded in GEO principles, for example, advocate/engage/deliver.
- Co-design and co-production assume we know who are intended users are – but do we? We can think about this in terms of the knowledge cycle (collect, organize, share, use) or the policy cycle (formulation, adoption, implementation, evaluation). On which points does GEO want to intervene?
- We must recognize that knowledge sharing occurs at multiple levels, it is multi-layered.
- Co-design also occurs at different levels. There is a kind of co-design at the global level: e.g. on SDGs, UN agencies are talking with GEO about what is possible for using EO data for indicators. Also, that co-design also doesn’t imply large scale adoption, but that some representative stakeholders work directly with the developers from early in the process.
- While GEO is working with a few countries to test methods of incorporating EO information in SDG reporting, GEO has not yet undertaken broad engagement at the national level. How can GEO reach national agencies that need to produce indicators?

- Several approaches were identified. GEO has a large network of Members and POs and can 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.
- GEO may not be able to engage every country one by one; it should perhaps work with other partners, e.g. Sustainable Development Solutions Network (SDSN) and the Global Partnership for Sustainable Development Data (GPSDD). Another approach is to pick up on good existing examples, such as Egypt's national SDG monitoring program.
- Another important aspect relates to the policy planning and operational side. This includes using EO data to model and predict, not just report. For example, with SDGs, how can a country model their current trajectory, develop policy scenarios and then determine the best policy course. This is not happening in many areas now and may be a new area where GEO could do more.

(b) in-situ data

- What is the definition of in-situ observations?
    - o Everything non-satellite (e.g. geo-referenced socio-economic data, statistical datasets)?
    - o A feature of in-situ data is that it should be measured on site, by a sensor, for verification of remote sensing.
    - o Reference data for benchmarking and calibration/validation which are essential parts of the system and serve particular purposes for derived datasets (e.g. poverty mapping).
    - o Must be careful of the slippery slope – do household surveys become in-situ as well? What can GEO put into a global system – this is the essential question.
    - o GEO documentation has traditionally referred to EO from "satellite, airborne, and in-situ" platforms.
- Does GEO want to get involved in policy? GEO, through GEOSS, serve to collect and curate authoritative data needed to inform various policy frameworks.
    - o In-situ data can become a political issue as well – there are sensitivities around sharing various data such as stream gauge. In some cases, data could be stored on an 'as-needed' basis.
    - o In order to be of value, GEOSS must strictly adhere to data sharing policies of Member states who will share data. Need to have capability of serving its community while not going beyond their mandate to GEO.
    - o Questions of comparability and reliability. Is there some minimum set of data that all GEO Member states should agree to, to be of use? If just 10 countries share their census data, this will be a great validation for work being done deriving population masks.
- The issue of in-situ data can presents an opportunity for GEO to infuse some excitement and help unlock datasets. Resources could be mobilized through targeted citizen science campaigns.
- To what extent does the ambition of creating a global repository go too far? What is given back if countries share their data? Global Biodiversity Information Facility (GBIF) was able to make trade-offs with data providers, ownership remained with providers, GBIF paid for digitzation of their data.
- What kind of services does GEO want to offer addition to serving as a data repository?
- IPR issues need to be addressed – GEO needs to be seen as part of solving problems

(c) reproducible science

- We need to expand the concept of a knowledge hub for :
    - allowing the sharing of experience; and
    - being a means for collaborative work.
- We don't have to re-invent anything, the technology already exists.
    - Having data location and software available via URLs .
    - Github, executable notebooks, docker files, docker images.
- What do we consider reproducible science?
    - Paper, documentations AND Runnable code AND Reference data accessible via APIs.
    - In the GEO Context we should not only consider science but also operations and helping the transition from science to operations.

(d) cloud computing for EO

- Metadata → data & software & computing.
- Knowledge hub / platform for sharing of knowledge and software.
- Build partnerships where both sides have clearly identified benefits.
- Make data and software available via APIs.

In summary, discussions at various points in the agenda recommended that GEO should re-design GEOSS to:

- be the world's repository of authoritative in-situ observation data (with GEO issuing a call for commitment from GEO Members to provide more in-situ data).
- help fill in gaps with respect to the climate observing system. In particular, areas where GEO can add value include:
    - adaptation (requires high resolution datasets for which capacity is now lacking, also help connect global to local scale);
    - data access;
    - coordination of in-situ measurements (especially in the absence of GTOS); and
    - coordination of carbon monitoring system.
- consult the GCOS implementation plan which can serve as a guidance document for GEO community.
- be seen as place to maturate scientific knowledge for disaster risk reduction (such as the IPCC in compiling science reports for UNFCCC - there is no equivalent for the Sendai Framework).
- be seen as the place for sharing knowledge, expertise, and technology support.
- make efforts to strengthen ties with UNSPIDER and UNOOSA. There is an opportunity for GEO to provide inputs for the next Global Assessment Report (GAR), particularly in the area of drought.
- play a role in providing near real-time data to feed early warning systems (perhaps through citizen science).

- support the needs of the re-insurance industry.
- focus its efforts on the handful of SDGs where EO can make a demonstrable difference.
- identify end-to-end pathways that go all the way to reporting, and provide the necessary technical and EO support to do so.
- provide clarity around privacy issues with respect to cloud computing and private work spaces.
- help users make sense of the complex world of EO and find complementarities among the multiple providers with various holdings and different missions.
- fill a niche by providing reproducible knowledge and algorithm validation.