Presenting Solutions and Best Practices on Data Sharing and Data Management in the GEO Community

Chu, W. B.[1]   Uhlir, P. F.[2*]

1. GEO Secretariat, Geneva, Switzerland
2. Consultant on Information Policy and Management, New York 12723, USA

This article summarizes the highlights of a Side Event meeting on Data Sharing and Data Management in the Group on Earth Observations (GEO) Community that was organized by the GEO Data Sharing Working Group (DSWG)(see https://earthobservations.org/index2.php for the GEO website) and by CHU Wenbo of the GEO Secretariat. Over 70 experts participated in the meeting, which was held on 2017 October 23 before the GEO Plenary meeting on October 25-26 in Washington, DC, USA.

The DSWG was established by GEO in 2006 to promote data sharing, which is one of the central goals of the organization. The DSWG also develops data policies for GEO and for the broader international Earth sciences and Earth observations communities.

As the DSWG and GEO Secretariat noted in the publicity for the meeting, "implementation of GEOSS Data Sharing and Data Management Principles is vital to GEO's success in the new decade. Without open and well-managed EO [Earth observation] data, GEO applications or decision making in various areas will not be transparent, scalable and accountable."

The objectives of the meeting focused on four major issue areas, which were reflected in the meetings four sessions. Each session had three to five speakers who gave short presentations and that then left ample time for discussion of the issues with the expert audience. The link to the background information and the slides of all the speakers can be found here: http://www.earthobservations.org/geo14.php?seid=536.

Before the sessions began, Dr. Barbara Ryan, the GEO Secretariat Director, gave the opening remarks. She expressed appreciation of the GEO community for the achievements made in data sharing and management in recent years and thanked the data groups for their dedicated work. She encouraged the community to continue to work together to make EO data more accessible and usable, and supporting international agendas such as the UN 2030 Agenda for Sustainable Development Goals (SDGs), the Paris Agreement on Climate Change, and the Sendai Framework for Disaster Risk Reduction.

The first session, moderated by Robert Chen, of Columbia University in the USA and representing the International Council for Science (ICSU) at this meeting, focused on Why Does Open

Data Sharing and Sound Data Management Matter? This panel invited speakers from several NGOs active in the open data area to look at the value of data sharing and management from broader views, not just limited to Earth observation data. The panelists included Shaida Badiee, Open Data Watch (http://opendatawatch.com/), Anne Hale Miglarese, Radiant Earth (https://www.radiant.earth/), and Paul Zeitz, SDG Compacts 2020.

There were a number of key messages that were identified by the speakers. Shaida Badiee noted that the current challenge about open data is not 'if' but 'how'. Earth observation data is part of the global data ecosystem. GEO should continue working with other data, such as statistical data. National statistical systems will have a greater role as intermediaries for change and closer integration with the GEO data. In countries that have official statistics and EO in the same government department (such as Mexico), both the data and the government department are likely to be more trusted and highly regarded. One step further is to include EO in the statistics law.

Anne Hale Milagrese said that it is critical to share data according to standard interoperable specifications. Innovation in spatial data transfer standards (SDTS) was led for a long time by government investments, but it is now increasingly led by the private sector because there is a large commercial market. The market is expanding significantly because of new applications, such as autonomous vehicles. Interoperability standards are still essential even for closed data, because open and closed data need to work together.

The last speaker, Peter Zeitz, noted that managers and policymakers in the EO arena need to think about the incentives for using and integrating data. Connecting data to businesses and planning processes and making such processes mandatory drives progress more quickly. Those involved in such activities should think about the rewards, particularly monetary ones. Examples that he cited were PEPFAR, a data-driven HIV/AIDS program, and the World Council on City Data, which developed an ISO standard and certifying cities on openness.

During the discussion with the audience, an issue was raised regarding the main barriers to data sharing and whether they are more technical or social and policy in nature. All the panelists agreed that the social and policy barriers were much greater than technical barriers. Barriers arising from a lack of capacity were also identified as being important.

In response to another question from the audience regarding how to promote the private sector to share their data for public-interest uses, the panelists stated that Uber has shared some of its data with the government of New York city and Facebook has also shared their data with the World Bank and the OECD. However, some privacy protection issues need to be resolved in order to move forward.

The second session focused on the Best Practices in GEO Initiatives and Flagship. It was moderated by Mariel Borowitz, an associate professor at Georgia Tech University and currently on detail to NASA in Washington, DC. GEO Initiatives are activities coordinated by GEO Members and Participating Organizations to develop and implement prototype services with Earth observation data and information according to GEO priorities. GEO Flagships have similar goals, but are expected to provide operational or near-operational services. Currently there are 22 Initiatives and 4 Flagships in the GEO Work Programme (http://earthobservations.org/geoss_wp.php).

The panel for this session included representatives from GEO Initiatives that have undertaken or planned data sharing and data management activities. The panelists included Stefano Salvi, of the GEO Geohazard Supersites and Natural Laboratories (GSNL); Jonas Eberle, of the GEO Wetlands Initiative; and QIU Yubao (replacing Peter Pulsifer, who could not attend the meeting), representing the GEO Cold Regions Initiative (GEOCRI).

The speakers raised several key messages. One is that researchers have begun to understand

that they need to make data open. The challenge is to move from open data to open science, which can really support decision-making. Stefano Salvi recommended using research objects that capture the workflow from data to results into a single object with a citable DOI (Digital Object Identifier). QIU Yubao noted that it is equally important to build a human network that can work well together in addition to a technical network that is interoperable.

In response to a question regarding how GEO data groups can improve the products for GEO Initiatives, QIU Yubao pointed out that he considered the GEO data principles to be too general for implementation, although they provide a useful framework. He suggested that the data groups showcase some initiatives with a prepared slide deck for the GO Initiative to use. Jonas Eberle added that it is important to demonstrate with use cases how the GEO data sharing and management principles can be established in practice; that is, how to write a data management plan. Stefano Salvi expressed the opinion that more direct interactions between the GEO Initiatives/Flagships and the data groups would be very helpful.

In response to a question regarding Initiatives/Flagships interaction with the GCI (the GEOSS Common Infrastructure), the panelists observed that there is a significant challenge in the long-term funding for adequate data infrastructure, especially to support the developing nations in joining the Initiatives/Flagships data activities. They hope to see possible solutions and success stories shared among the Initiatives/Flagships.

There was a Welcome to the afternoon session provided by LI Pengde, GEO Co-Chair Representative of China and Co-Chair of the United Nations Committee of Experts on Global Geospatial Information Management (UN-GGIM). Dr. LI shared some of his experiences in the UN and China about open data stimulating various positive impacts on society. He encouraged GEO and UN-GGIM to explore further collaboration.

The afternoon sessions were chaired by Paul Uhlir, who was representing the international Committee on Data for Science and Technology (CODATA). The first afternoon session addressed Experience and Best Practices in Data Management Policy and Implementation among GEO Members and Participating Organizations. It was moderated by Carrie Seltzer, who was at the meeting representing the Belmont Forum.

GEO Members are governmental ministries and come from 105 countries, plus the European Commission. The Participating Organizations in GEO are non-governmental organizations (NGOs) and intergovernmental and international organizations, rather than nations. There are now 118 of them.

The panelists included Jeff de la Beaujardiere, of the National Oceanic and Atmospheric Administration (NOAA), USA; Chris Jarvis, of the Environmental Agency, UK; SHI Ruixiang (replacing LIU Chuang, who was unable to attend), of the Chinese Academy of Sciences (CAS), China; and Gilberto Camara, of the National Institute for Space Research, Brazil.

Carrie Seltzer initiated the panel by sharing the Belmont Forum's mechanism for implementing data policy: requiring submission of data management information from Belmont Forum funded projects, encouraging training in data management skills, and collecting data policies from participating funders.

Dr. de la Beaujardiere introduced NOAA data policies (https://nosc.noaa.gov/EDMC/), Data Portal (https://data.noaa.gov), data access support (https://coastwatch.pfeg.noaa.gov/erddap/index.html, including the Unified Access Framework at https://geo-ide.noaa.gov/ and the Dataset Identifier Project). He pointed out that data management in and of itself is not the goal; the goal is to use and reuse data, and to extract maximum value from the data. He recommended full use of cloud services.

Chris Jarvis shared some of the UK's experiences in the implementation of an 'Open Data by default' data policy. Here was clear evidence, including an increase in the use of environmental

data by 25 times, that was demonstrated regarding the benefits by adopting the Open Data policy.

SHI Ruixiang shared China's new practice with the journal on Global Change Data Publishing and Repository (GCdataPR) (http://www.geodoi.ac.cn). It promotes an open data policy for end users, DOI for citation, peer review for quality control, and partnership with traditional journals for data management. For developing countries that still struggle with open data policy at the government levels, data publishing provides a vehicle for scientists to help each other on data acquisition from the bottom-up, although it is not an alternative way to use and reuse government data.

Finally, Gilberto Camara introduced the Brazilian experience concerning open data. Greater transparency builds better governance (for example, Real-time Deforestation Monitoring was enabled by open satellite data), and array databases (data cube) advances data integration for science and decision-making.

The last panel of the Side Event, moderated by Miles Gabriel of the UK delegation, focused on the Best Practices in the GEOSS Common Infrastructure. The panelists included Siri Jodha Khalsa, Institute of Electrical and Electronics Engineers (IEEE); Massimo Craglia, Joint Research Centre (JRC), European Commission; Robert Downs, representing the World Data System (WDS); and Steve Browdy, IEEE.

Siri Jodha Khalsa gave an overview to the GEOSS Common Infrastructure (GCI), the GEOSS Data Sharing Principles (DSPs) and the Data Management Principles (DMPs). The current approach of implementing DSPs and DMPs in GCI is the conformance of the DSPs and DMPs to the GEOSS Yellow Pages (http://www.geoportal.org/yellow-pages). Key elements at the data repository level are data repository certification and status checker, and at the dataset level, the assessment of data fitness for use and user feedback.

Massimo Craglia introduced the findings from an in-depth analysis of 1.8 M metadata records, covering 89% data within the GCI. 1.4 M out of 1.8 M data are made available through the GEOSS Data-CORE, though there are 520 ways to describe "open data", which causes unnecessary confusion. There is also a lack of persistent identifiers (PIDs) for datasets and difficulty in understanding their curation levels. Paths for improvement include the use of shared thesauruses or lists of common concepts, cross-domain mappings, improved data management practices (including PIDs), and consolidation at the level of major facilities or data collections.

Robert Downs presented the view of implementing DSPs and DMPs based on synergies between them and identifying ways to prioritize aspects of implementation to amplify the benefits of data sharing and data management.

The last speaker, Steven Browdy, addressed the interoperability issues raised by the DMP survey. Notably, there were use conditions to promote the legal interoperability of data. Now only Data-CORE compatible licenses and waivers are considered, so the survey raised the need for expansion of such mechanisms. Sematic interoperability is another area that needs better mediation or mapping between vocabularies. Finally, the automation of PID implementation would be highly desirable.

The meeting participants recognized that it is necessary for the GEO Community to continue paying more attention to the implementation of the Data Sharing and Data Management Principles. The participants provided examples from around the world, and described best practices and solutions for major and common challenges, which are possible paths to reach the GEO objectives.

Readers of this summary are encouraged to access all the speakers' slide presentations for further details. The link to that again is: http://www.earthobservations.org/geo14.php?seid=536.