SYMPOSIUM ON THE DATA SHARING PLANS AND ON THE SCIENTIFIC BENEFITS OF DATA SHARING IN THE GLOBAL EARTH OBSERVING SYSTEM OF SYSTEMS (GEOSS)

The **Symposium on the Data Sharing Plans and on the Scientific Benefits of Data Sharing in GEOSS,** held on November 16, 2009, brought together more than 120 interested data managers, scientists, and policy makers from around the world to discuss the GEOSS Data Sharing Principles and their potential benefits for research.

Part one of the Symposium opened with introductions by *Roberta Balstad* from Columbia University and chair of the US National Committee for CODATA, and by *Masami Onoda* of the Group on Earth Observations (GEO) Secretariat.

Implementing the GEOSS Data Sharing Principles: Where We Are and Where We're Going

Beth Greenaway, UK Environmental Observation Framework, and GEO Data Sharing Task Force

In 2006, CODATA agreed to lead the GEO Task DA-06-01 on Data Sharing. The CODATA Task Team held several meetings, wrote a white paper, and developed draft *Guidelines for Implementing the Data Sharing Principles.* The white paper was published concurrently in the CODATA *Data Science Journal* and the *Journal of Space Law* in 2009. (This article is available at http://www.spacelaw.olemiss.edu/JSL/articles/35JSL201.pdf and http://www.jstage.jst.go.jp/article/dsj/8/0/GEO1/_pdf).

In 2008, the GEO Plenary in Bucharest agreed to establish a Data Sharing Task Force (DSTF) to augment the CODATA Task Team. The DSTF includes members from all over the world and is led by six co-chairs (including the CODATA Task Team). The DTSF's near-term objectives include finalizing the Implementation Guidelines for the 2009 GEO Plenary and interacting with GEO stakeholders on their data sharing opportunities and needs. The DSTF will prepare an Action Plan by the 2010 GEO Ministerial Summit to help implement the Data Sharing Principles (discussed below) and to enable the development of working procedures for data sharing within GEOSS.

The success of GEOSS is contingent upon compliance with the GEOSS Data Sharing Principles by GEO members. No single set of rules will apply to all types, sources, and uses of data. Having a clear set of guidelines, definitions, and minimum expectations, however, will help to improve the sharing of data within GEOSS.

An Overview of the Key Substantive Provisions of the Implementation Guidelines for the GEOSS Data Sharing Principles

Robert Chen, Co-chair, GEO Data Sharing Task Force; Secretary-General, CODATA; Director, CIESIN, Columbia University; Manager, NASA Socioeconomic Data and Applications Center (SEDAC)

"The societal benefits of Earth observations cannot be achieved without data sharing."

In 2005, the GEO Members adopted the GEOSS Data Sharing Principles by consensus. The Principles provide for "full and open exchange of data, metadata, and products shared within GEOSS, recognizing relevant international instruments and national policies and legislation. All shared data, metadata, and products will be made available with minimum time delay and at minimum cost. All shared data, metadata, and products for research and education are encouraged to be free of charge or no more than the cost of reproduction."

For GEOSS to be successful as a system of systems, a wide range of high quality data needs to flow seamlessly through the component systems to users, and users need to be able to make full use of the data in order to achieve desired benefits. GEOSS is composed of voluntarily contributed systems and data, governed by pre-existing laws, policies, and practices. Therefore, GEO needs to encourage "good-faith voluntary adherence" to the Principles, directly engaging both data providers and data users.

The draft Implementation Guidelines are intended to provide clear definitions and rationales for recommended implementation approaches, recognizing different needs in different Societal Benefit Areas and different practices and policies around the world. GEO has identified nine Societal Benefit Areas that are the principal focus of the organization, which include disasters, health, energy, climate, water, weather, ecosystems, agriculture, and biodiversity. The Guidelines fall into 6 areas: (1) promoting full and open exchange, (2) enabling reuse and re-dissemination of data, (3) ensuring consistency with relevant international instruments and national policies and legislation, (4) consistent pricing policies, (5) reducing time delays, and (6) promoting research and education.

The draft Implementation Guidelines are voluntary, providing a framework for cooperative action to achieve the GEOSS Data Sharing Principles in practice. The Guidelines were developed in direct response to the existing set of Data Sharing Principles and therefore do not (yet) address important issues such as data quality. The Guidelines have been extensively reviewed by GEO Committees and the Members, and modified in response to comments and suggestions. The Guidelines can be changed by GEO in response to an evolving GEOSS and changing needs. (The GEO Plenary subsequently approved the Implementation Guidelines, with some small modifications. See http://www.earthobservations.org/geoss_dsp.shtml.)

Part One of the Symposium concluded with a panel discussion moderated by *Roberta Balstad*. Part Two focused on the scientific benefits of data sharing, as summarized below.

Innovations in Information Sharing: Flexible Architectures for a Fast Changing World

Christopher Tucker, Open Geospatial Consortium Board

The GEOSS information sharing vision is enabled by the International Standards Organization (ISO)/Open Geospatial Consortium (OGC) Service Oriented Architecture, which integrates existing processing, exploitation, and dissemination ground segments and sensor gateways. Sharing within this architecture is centered on users and social networks. GEOSS assumes a global data commons, but not all of the best data will exist in the commons. To successfully share beyond the commons, we will need to use other methods such as G-Commerce (i.e., *e*-Commerce for geospatial data). If human activities are the largest source of global environmental impact and if the GEO Societal Benefits Areas are largely tied to long-term human welfare, then it is important to geospatially and temporally manage and share data on socio-cultural dynamics as part of GEOSS.

We do not just share data for the sake of sharing, however. We share them because we have a common mission. With the flood of data and information available to us, we cannot just open the doors to sharing and expect everything to work itself out. The long standing remote sensing paradox of "coverage versus resolution" actually plagues us far beyond the world of remote sensing, and as such, we need some organizing concept to drive our information sharing strategy. Dr. Tucker calls this "Change Intelligence."

The challenges of information sharing require sharing all the way to the sensor; beyond the commons; data on socio-cultural dynamics; and achieving "Change Intelligence." These challenges require innovation if we are to realize the broader goals of GEOSS.

Understanding Ecosystems and Their Services

Anthony Janetos, Director, Joint Global Change Research Institute, Pacific Northwest Laboratory and the University of Maryland

Ecosystems are biological communities and their physical environment. People and infrastructure should be thought of as an integral part of ecosystems, not separate from them. Ecosystem services are the functions of ecosystems from which we benefit and these benefits can be direct or indirect. There are services that ecosystems provide that we depend on, but do not pay for until we have to replace them. This concept has grown to recognize that services can be either outside or inside of existing markets.

Ecosystems face a multitude of challenges. Over the past half century, humans have changed ecosystems more rapidly and extensively than in any comparable period of time in human history. The changes that have been made to ecosystems have contributed to substantial net gains in human well-being and economic development**.** But these gains have been achieved at growing costs that, unless addressed, will significantly diminish the benefits that future generations obtain from ecosystems. There has been a substantial and largely irreversible loss in the diversity of life on Earth, accompanied by a degradation and unsustainable use of ecosystem services. The degradation of these services frequently causes harm to human well-being and represents a loss of natural and capital assets of a country, but the loss of wealth due to such degradation is not reflected in economic accounts.

Perhaps the most consequential human-driven change of Earth's important natural characteristics is in the land-cover vegetation and land-use. About half of original forest area has been converted to agricultural production. As a result, the amount of biologically available nitrogen has roughly doubled, which increases atmospheric concentrations of carbon dioxide. This conversion of ecosystems is also the biggest contribution to loss of biological diversity. The magnitude of the challenge to come involves the fact that most direct drivers of degradation in ecosystem services remain constant or are growing in intensity in most ecosystems.

Data sharing is necessary because stresses on ecosystems and benefits from their services are more than simply local. The strategies for maintaining services from ecosystems require cooperation among many different institutions and we must have a common information base. Because changes are occurring now, we are coping with existing circumstances and have questions about planning for the future to consider. This requires information on current practices and the understanding of factors that control vulnerability, the ability to model the effectiveness of adaptation strategies as part of an integrated response portfolio, and free and open exchange of data for developing a common understanding and appropriate strategies for response.

Earthquakes, Tsunamis, and Nuclear Explosions: Open Data Exchange for Research and Monitoring in Seismology

David Simpson, President, Incorporated Research Institutions for Seismology

Seismology is a global and international science. Earthquakes occur and are observed on a global level and, therefore, international data exchange is essential for monitoring and research. Seismology supports a rich portfolio of research and societal applications, such as deep Earth structure, global seismicity, and earthquake source processes, as well as the mitigation of and response to earthquake hazards and the monitoring of nuclear test ban sites.

The Incorporated Research Institutions for Seismology (IRIS) is a facilities program established in the U.S. in 1984. It supports research and education in the earth sciences through the collection and distribution of seismological data for studies of earthquakes, the structure and dynamics of the Earth, and verification of the Comprehensive Test Ban Treaty. The program is developed and managed by the IRIS Consortium. It has three principal tasks that form its mission. It facilitates and conducts geophysical investigations of seismic sources and other properties of the Earth, using seismic and other geophysical methods. It promotes exchange of geophysical data and knowledge, both through use of standards for network operations, data formats and exchange protocols, and through pursuing policies of free and unrestricted data access. And it fosters cooperation among IRIS Members, Affiliates, and other organizations in order to advance geophysical research and convey benefits from geophysical progress to all of humanity.

One of the IRIS core programs is the Global Seismographic Network (GSN), which provides uniform global coverage in cooperation with the U.S. Geological Survey and other partners in the international Federation of Digital Seismograph Networks (FDSN). The GSN has 150 stations in unattended locations worldwide, including remote oceanic islands and polar regions and previously inaccessible regions of China and the former Soviet Union. This provides for integration with national and regional networks in many developing nations. The GSN has IRIS-established standards for observatory instrumentation featuring broadband (3000 sec to 10 hz), high resolution (24 bit), continuous recording, real-time telemetry at most sites, and free and open data access.

Another IRIS core activity is the Program for the Array Seismic Studies of the Continental Lithosphere (PASSCAL). This provides free access to portable instrumentation for use in research programs funded by the National Science Foundation. It is collaborative, international, multi-institutional, and multi-disciplinary. One of its requirements is to deliver data to the IRIS Data Management Center for open release following a two-year proprietary period. PASSCAL has temporary field installations with deployments of tens to hundreds of instruments and experiments, lasting from one week to three years. IRIS management of the USArray component of the large NSF-funded EarthScope project for detailed geophysical observations in North America, is closely linked to PASSCAL.

A third IRIS core program is the Data Management System, DMS, which collects, archives, and distributes data produced by all IRIS programs and contributed by many national and international partners. The archive includes data from over 12,000 stations, from 248 networks that include permanent networks, FDSN global networks, national and regional networks, temporary deployments, research arrays, aftershock studies, and crustal imaging. It currently has 1920 stations that send data in real-time.

The IRIS Data Archive has more than 120 terabytes of data. Its current rate of growth is about 20 terabytes/year and the data distribution is about 50 terabytes/year. Advancing technology, including satellite, cell phones, internet, mass storage devices, and advanced database management systems, are the drivers of open data exchange. The overall system evolves according to data and metadata standards (such as SEED, the “Standard for Exchange of Earthquake Data”), as well as instrumentation standards and quality standards. IRIS is thus seeking to overcome cultural barriers and to advance beyond “scientific colonialism” by encouraging increased collaborations in research and observations with international partners. It is illustrating the values of global data collection as a complement to regional observations. Giving recognition and exposure to data providers gives them international visibility.

Symposium Agenda

15:30	Introduction by the Symposium Chair	Roberta Balstad, Columbia University and U.S. National Committee for CODATA
15:35	Welcoming Remarks	Masami Onoda, Group on Earth Observations (GEO) Secretariat

Part One: Implementing the GEOSS Data Sharing Principles

15:45	Implementing the GEOSS Data Sharing Principles: Where We Are and Where We're Going	Beth Greenaway, UK Environmental Observation Framework
16:00	An Overview of the Key Substantive Provisions of the Implementation Guidelines	Robert Chen, CODATA and Columbia University
16:20	Panel Discussion with the Symposium Participants - Moderated by Roberta Balstad	

Part Two: The Scientific Benefits of Data Sharing

16:30	Innovations in Information Sharing: Flexible Architectures for a Fast Changing World	Christopher Tucker, Open Geospatial Consortium (OGC) Board
16:50	Understanding Ecosystems and Their Services	Anthony Janetos, Director, Joint Global Change Research Institute, Pacific Northwest Laboratory and the University of Maryland
17:10	Earthquakes, Tsunamis and Nuclear Explosions: Open Data Exchange for Research and Monitoring in Seismology	David Simpson, President, Incorporated Research Institutions for Seismology

Symposium Sponsors

The Symposium was organized by the U.S. National Committee for CODATA of the National Research Council Board on Research Data and Information, together with the U.S. National Committee for DIVERSITAS and the U.S. National Committee on Geodesy and Geophysics, and in collaboration with the international Committee on Data for Science and Technology (CODATA), the Group on Earth Observations (GEO) Secretariat, the GEO Data Sharing Task Force, and the Environmental Change and Security Program of the Woodrow Wilson International Center for Scholars (which provided the meeting site).