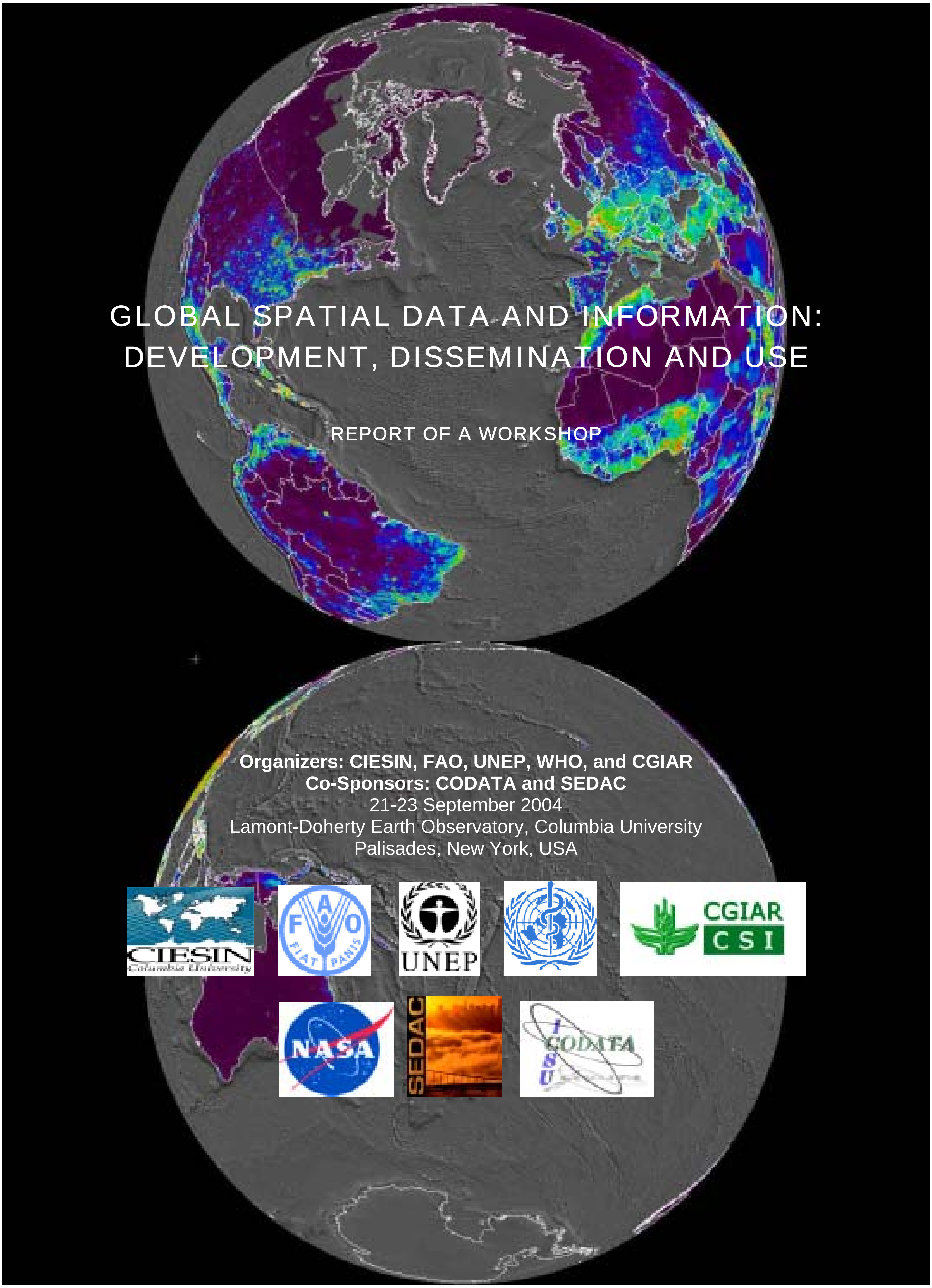


# GLOBAL SPATIAL DATA AND INFORMATION: DEVELOPMENT, DISSEMINATION AND USE

REPORT OF A WORKSHOP

**Organizers: CIESIN, FAO, UNEP, WHO, and CGIAR**
**Co-Sponsors: CODATA and SEDAC**
21-23 September 2004
Lamont-Doherty Earth Observatory, Columbia University
Palisades, New York, USA

*Page left intentionally blank*

# GLOBAL SPATIAL DATA AND INFORMATION: DEVELOPMENT, DISSEMINATION AND USE

REPORT OF A WORKSHOP

**Organizers: CIESIN, FAO, UNEP, WHO, and CGIAR**
**Co-Sponsors: CODATA and SEDAC**
21-23 September 2004
Lamont-Doherty Earth Observatory, Columbia University
Palisades, New York, USA

The Global Spatial Data and Information User Workshop was co-organized by the Center for International Earth Science Information Network (CIESIN) of Columbia University, the Food and Agriculture Organization of the United Nations (FAO), the United Nations Environment Programme (UNEP), the World Health Organization (WHO), and the Consortium for Spatial Information of the Consultative Group for International Agricultural Research (CGIAR). It was co-sponsored by the International Council for Science (ICSU) Committee on Data for Science and Technology (CODATA)and the Socioeconomic Data and Applications Center (SEDAC) of the U.S. National Aeronautics and Space Administration (NASA) Earth Observing System Data and Information System (EOSDIS).

CIESIN, a unit of the Earth Institute at Columbia University, seeks to improve the data and information base for understanding interactions between humans and the environment. This mission places CIESIN at the intersection of the social, natural, health and information sciences. To accomplish its mission, CIESIN creates distributed data and information resources, provides users with new types of interactive analysis and visualization capabilities, and conducts research about human interactions with the environment as well as about the management of data and information.

CIESIN at Columbia University
P.O. Box 1000
69 Route 9W
Palisades, NY 10960 USA
Tel. +1-845-365-8988
Fax +1-845-365-8922
Web: http://www.ciesin.columbia.edu

UN Food and Agriculture Organization
Viale delle Terme di Caracalla
00100 Rome, Italy
Tel. +39-06-57051
Fax +39-06-570-53152
Web: http://www.fao.org/

UN Environment Programme
P.O. Box 30552
Nairobi 00100 Kenya
Tel. +254-20-624299
Fax: +254-20-624269
Web: http://www.unep.org/dewa/

World Health Organization
Avenue Appia 20
1211 Geneva 27, Switzerland
Tel. + 41-22-791-2111
Fax + 41-22-791-3111
Web: http://www.who.int

CGIAR Consortium for Spatial Information
P.O. Box 2075
Colombo, Sri Lanka
Tel. +94-11-2787404
Fax +94-11-2786854
Web: http://www.csi.cgiar.org/

The workshop and this report were partially supported by NASA under contract NAS5-03117 with Goddard Space Flight Center. The views expressed in this report are not necessarily those of CIESIN, Columbia University, NASA, nor the co-organizers and co-sponsors of the workshop.

**Suggested Citation:** de Sherbinin, A., and R.S. Chen, eds. (2005). *Global Spatial Data and Information User Workshop: Report of a Workshop*, 21-23 September 2004. Palisades, NY: Socioeconomic Data and Applications Center, Center for International Earth Science Information Network, Columbia University. Available from http://sedac.ciesin.columbia.edu/GSDworkshop/

# Table of Contents



This report, together with all presentations, is available through the workshop website at:
http://sedac.ciesin.columbia.edu/GSDworkshop/

*The images on the report cover are of CIESIN's Gridded Population of the World v.3 data set visualized using SEDAC's* TerraViva! *software. The software was developed by ISciences LLC and licensed by SEDAC for distribution of its own and other public domain data sets. The software, which comes pre-packaged with the data sets, is available free-of-charge for researchers, policy makers or applied users. To obtain a copy, send your name, title, affiliation and address via email to: ciesin.info@ciesin.columbia.edu.*

# Abbreviations and Acronyms

| | |
|---|---|
| APNHIN | Asia Pacific Natural Hazards Information Network |
| AVHRR | Advanced Very High Resolution Radiometer |
| AWRD | African Water Resource Database |
| CBD | Convention on Biological Diversity |
| CEOS | Committee on Earth Observation Satellites |
| CGDB | Inventory and Comparison of Globally Consistent Geo-Spatial Databases and Data Libraries |
| CGIAR | Consultative Group on International Agricultural Research |
| CIAT | International Center for Tropical Agriculture |
| CIESIN | Center for International Earth Science Information Network |
| CIMMYT | International Maize and Wheat Improvement Center |
| CODATA | Committee on Data for Science and Technology |
| COTS | Commercial off the shelf |
| CSI | Consortium for Spatial Information of the CGIAR |
| DCW | Digital Chart of the World |
| DE | Digital Earth |
| ECAT | Electronic Catalogue of Scientific Names |
| EIP | Enterprise Information Portal |
| ESSP | Earth System Science Partnership |
| FAO | Food and Agriculture Organization of the United Nations |
| FEWS | Famine Early Warning System |
| FGDC | Federal Geographic Data Committee |
| FIVIMS | Food Insecurity and Vulnerability Information and Mapping Systems |
| GBIF | Global Biodiversity Information Facility |
| GCMD | Global Change Master Directory |
| GEO | Group on Earth Observations |
| GIEWS | Global Information and Early Warning System |
| GIS | Geographic Information System |
| GISD | Geographic Information for Sustainable Development |
| GLiPHA | Global Livestock Production and Health Atlas |
| GPW | Gridded Population of the World |
| GRDC | Global Runoff Data Centre |
| GRUMP | Global Rural-Urban Mapping Project |
| GSDI | Global Spatial Data Infrastructure |

| | |
|---|---|
| GTOPO30 | Global 30 Arc Second Elevation Dataset |
| HYDE | History Database of the global Environment |
| ICARDA | International Center for Agricultural Research in the Dry Areas |
| ICSU | International Council for Science |
| IFPRI | International Food Policy Research Institute |
| IGBP | International Geosphere-Biosphere Programme |
| IGO | Inter-governmental organization |
| IGOS | International Global Observing System |
| IP | Intellectual Property |
| IPRs | Intellectual Property Rights |
| ISCGM | International Steering Committee for Global Mapping |
| ISO | International Standards Organization |
| ISPRS | International Society for Photogrammetry and Remote Sensing |
| IWMI | International Water Management Institute |
| JRC | Joint Research Centre |
| LCCS | Land Cover Classification System |
| MA | Millennium Ecosystem Assessment |
| MDGs | Millennium Development Goals |
| NASA | National Aeronautics and Space Administration |
| NGO | Nongovernmental organization |
| NMO | National mapping organization |
| NOAA | National Oceanic and Atmospheric Administration |
| NSIDC | National Snow and Ice Data Center |
| OGC | Open Geospatial Consortium |
| PMPG | Poverty Mapping Project Group |
| PPF | Peace Parks Foundation |
| RIVM | Dutch National Institute of Public Health and the Environment |
| SALB | Second Administrative Level Boundaries |
| SDI | Spatial Data Infrastructure |
| SEDAC | Socioeconomic Data and Applications Center |
| SIMAG | Spatial Information Management Advisory Group |
| UDDI | Universal Description, Discovery, and Integration |
| UNEP-GRID | United Nations Environment Programme – Global Resource Information Database |
| UNICEF | United Nations Children's Fund |
| UNGIWG | United Nations Geographic Information Working Group |
| VMAP | Vector Smart Map |

| | |
|---|---|
| WCMC | World Conservation Monitoring Center |
| WHO | World Health Organization |
| WMO | World Meteorological Organization |
| WSIS | World Summit on the Information Society |
| WWF | World Wildlife Fund |

# Workshop Statement

The Global Spatial Data and Information User Workshop was held on 21-23 September 2004 at Columbia University in Palisades, New York, USA. We, the participants in the workshop, are representative of a wide range of global-scale data and information developers, managers, distributors, and users from both governmental and nongovernmental organizations around the world.

We **recognize** that global-scale datasets on the environment, agriculture and food security, health, population, and poverty represent:

a) The cumulative and collective knowledge of humanity about critical aspects of the environment and sustainable development;

b) Essential information resources needed by scientists, decision makers, applied users, educators, and many others to advance science, support education, ensure sustainable development, and meet the United Nations Millennium Development Goals; and

c) The long-term foundation for shared understanding and effective action to improve the quality of human life and the environment.

We therefore **acknowledge** the collective responsibility to:

1) Make global-scale data and derived information as widely accessible and usable to all types of users as possible, while recognizing the intellectual property rights of the underlying data sources;

2) Promote the appropriate use of these data and information resources among all types of users, through provision of suitable metadata and documentation, expert guidance, outreach to key user communities, and other means;

3) Improve the quality, comprehensiveness, and usability of global-scale datasets and derived information through collaboration with the relevant data sources and managers, the scientific community, diverse data users, and key sponsors;

4) Improve the capacity of data sources, data managers, and data and information users in developing countries to contribute to and benefit from global-scale data and information resources;

5) Promote efficient and seamless integration of global-scale data development, management, and access with corresponding local, national, and regional data programs, initiatives, and networks;

6) Establish effective coordination with other related data and information efforts including the development of national and global spatial data infrastructure, ongoing intergovernmental data programs, relevant international efforts to develop and implement open standards, and present and future international scientific initiatives; and

7) Ensure the long-term stewardship of these data including their long-term preservation and access.

We therefore **support**, in our personal capacities, collaborative efforts to:

1) Develop and implement bilateral and multilateral international collaboration in global-scale data and information development, management, and dissemination;
2) Further refine and articulate general principles for global-scale data and information development, sharing, access, and stewardship based on the responsibilities outlined above;
3) Expand interactions with diverse global data and information user communities to ensure clear identification and continued responsiveness to user needs;
4) Address key gaps and weaknesses in current global datasets as identified in the workshop discussions and subsequent analyses;
5) Address key needs for capacity building, training, development of interoperability, and portal development as identified in the workshop and subsequent analyses;
6) Work on the development and implementation of an efficient international infrastructure to support global data development, access, and use drawing on open standards, appropriate technologies, sound science, and professional data management;
7) Establish one or more international coordination mechanisms for these global-scale data activities, such as the proposed Spatial Information Management Advisory Group (SIMAG);
8) Develop appropriate linkages with relevant international organizations and networks;
9) Promote awareness of the value and utility of global-scale spatial data and information in key venues such as the World Summit on the Information Society (WSIS), the Integrated Global Observing Strategy (IGOS), the Group on Earth Observations (GEO) initiative, and the 2005 UN Summit to review progress since the 2000 Millennium Declaration; and
10) Work with the appropriate governmental authorities and agencies, funding sources, scientific bodies, and other organizations to secure the resources needed to implement these activities.

# Introduction

Many different global- and regional-scale datasets on the environment and human development are being developed and disseminated by a range of institutions around the world. With the increasing use of more open, interactive mapping servers and greater capabilities by users to access and utilize large global datasets comes the potential for increased problems related to inconsistent data integration and visualization, variable data quality and documentation, uncoordinated proliferation of different versions of data sets, unnecessary duplication of effort, excessively complex restrictions on data re-dissemination and use, and incomplete or incorrect citation and attribution of data.

The Global Spatial Data and Information User Workshop brought a core set of institutions actively involved in global spatial data and information development and dissemination together with representatives of key user communities to address both short- and long-term needs for coordination and collaboration. The overall purpose of the workshop was to coordinate spatial data and information dissemination among key actors in the global data community in order to better meet user needs at global as well as regional and national levels.

The three-day workshop was co-organized by the Center for International Earth Science Information Network (CIESIN) of the Earth Institute at Columbia University, the Food and Agriculture Organization of the United Nations (FAO), the United Nations Environment Programme (UNEP), the World Health Organization (WHO), and the Consortium for Spatial Information of the Consultative Group on International Agricultural Research (CGIAR). CIESIN's NASA-funded Socioeconomic Data and Applications Center (SEDAC) served as the local host. The workshop was also co-sponsored by the Committee on Data for Science and Technology (CODATA) of the International Council for Science (ICSU). This workshop followed a successful technical workshop at FAO headquarters in Rome, Italy on 18-19 March 2004 on coordination of and cooperation in spatial information management capacity development, which resulted in the FAO-led Spatial Information Management Advisory Group (SIMAG) initiative.

The workshop addressed five principal areas with regards to the production, dissemination, and use of global data sets:

1. Stock taking of global data sets and identification of significant gaps and overlaps in the following thematic areas: environment; food and agriculture; population, poverty and health.
2. Standardization/harmonization of spatial data and information, including:
    a. Standards and conventions for mapping and Geographic Information Systems (GIS), e.g., consistent projections, scales, and boundary files.
    b. Additional information on data sets (beyond discovery metadata) that would facilitate their appropriate use (e.g., data set guides).
3. Identification of user needs for online services, and for education and capacity building in how to use those services.
4. Flexible user access through interoperability and open standards of both catalog searching and Internet mapping.
5. Integration of biophysical and socioeconomic data, including substantive and methodological issues.

The workshop was attended by seventy participants, many of whom represented large organizations or networks of data developers, disseminators and users. The core of the workshop was five breakout group discussions that addressed the following topics in depth (guidelines, discussion topics and group members for each of the breakout groups are in Annex 2):

- Principles for data sharing and access
- Gaps in global data and collaboration in data development
- Data search, discovery, and documentation; role of portals
- Technical data interoperability
- Science data integration

This report provides a summary of the plenary presentations and more detailed reports of the breakout group discussions. Presentation files for the plenary presentations may be downloaded from the workshop website: http://sedac.ciesin.columbia.edu/GSDworkshop/.

# Summary of Plenary Sessions

The plenary sessions covered a number of substantive issues relevant to the development, dissemination and use of global spatial data. The following sub-sections cover each plenary presentation in turn.

### *Introductory Session*

**Dr. Robert Chen** of CIESIN provided an overview on the motivation for the workshop. The following were some of the issues that prompted the workshop:

- There has been a proliferation of global datasets on a wide range of topics, and a greater diversity of data sources in both developed and developing countries, including emerging regional and topical data networks.
- The increasing complexity of global databases and more frequent overlap and cross-disciplinary applications.
- The emergence of new tools for providing online visualization of and access to global-scale data – both "centrally" and from distributed data servers.
- Increasing concerns about intellectual property rights, data attribution, metrics on data use, data quality and uncertainty, liability.
- The diversity of users, from scientists to applied users to students and educators.
- The need to improve quality and efficiency of data development, maintenance, archiving, dissemination, and user support.
- The need for better justification, assessment of benefits, and more resources for global data development.

Dr. Chen then went on to describe the emergence of a multiplicity of Internet mapping and data download portals. Problems that he hoped would be resolved, at least in part, by the discussions at the workshop include:

- Unnecessary confusion among users about which dataset to use for their application.
- Inability of users to remember what version of a dataset they used to create a figure or how to cite it.
- Poorly matched coastlines, boundaries, point locations, labels, and projections in Internet map services.
- Misinterpretation or misuse of data due to use of inconsistent reference data or parameters, confusing terminology, or poor documentation.
- Unnecessary duplication of datasets across many data distributors and extended use of obsolete datasets due to slow updating.
- Widespread use of relatively poor quality but unrestricted datasets due to unnecessary restrictions, poor documentation, limited awareness, and/or lack of access to higher quality datasets.

- Time wasted creating and implementing specific data sharing agreements and gathering user metrics from secondary distributors.
- Persistent data quality problems on certain key topics, in specific regions, or with respect to spatial and temporal scale and resolution.

He also expressed the hope that several positive outcomes would emerge from the workshop, such as general principles for global data sharing, dissemination, and attribution; identification of ideas and opportunities for collaboration in prototyping, developing, and implementing new global datasets; common architecture, expanded search tools, and improved documentation resources for portals; consensus on standards, conventions, protocols, and approaches to promote harmonization and interoperability in global data; and identification of key institutions, networks, user groups, and funding sources not represented at the workshop.

**Mr. Jeffrey Tschirley** of the FAO spoke next on the subject of the Spatial Information Management Advisory Group (SIMAG) initiative. He began by outlining some FAO priorities. FAO wants to put information within reach of decision makers and managers. This includes creating knowledge networks and disseminating statistics, data and analyses. FAO also wishes to bring technical knowledge to the field, which means developing capacity to collect and use information on biodiversity, land quality, climate variability, agronomy, rural livelihoods and food security. Finally, FAO conducts policy analyses on agriculture, forestry, and fisheries to achieve rural development, alleviate hunger, reduce poverty.

Mr. Tschirley then proceeded to list a number of challenges faced by everyone in the data community, among them:

- The variable quality of most of the key terrestrial data and information;
- the generation of information that is genuinely demand-driven;
- National policies that restrict data access;
- Poor data and information management, update, assembly, and assimilation;
- Weak mechanisms for data and information sharing, including collaborative analysis;
- Uncoordinated investment in end-to-end institutional capacity;
- Unharmonized data and information collection and dissemination methods and standards; and
- Unused data and information, including point data, field analyses, archived satellite data.

In response to these challenges, FAO has proposed the launch of SIMAG, a coalition of data providers who would learn from complementary experiences and expertise, address common challenges to their respective institutions, and build toward a multi-scale membership of users, providers, and managers. The idea is to build habits of collaboration, whereby organizations network through inter-linked projects to promote technical quality and vocational education and training. SIMAG would also serve as a catalyst and a voice for national, regional, and global cooperation. The vision is for a consultative and inclusive process to encourage cross-sectoral communication that would be based on inter-disciplinary management approaches. Members would also be asked to make an institutional commitment to information sharing, and capacity-building.

To achieve this, SIMAG needs a shared and clear vision and strategic framework, alignment with other information management initiatives at national and global levels, coordination of funding efforts, institutional commitment, and a spirit of partnership characterized by information and expertise sharing and in-kind support. He proposed that the workshop could contribute to this vision by developing terms of reference (objectives, activities and outputs), definition of members' roles, a strategy for implementation, and pilot projects.

### *Plenary Panel 1. Access to Global Spatial Data and Information from the User's Perspective*

During the last decade, technological advancements in remote sensing, GIS, and other geospatial data management tools in tandem with development of the Internet have contributed to the creation of many global and regional spatial databases, online mapping services and geospatial data clearinghouses. However, the dramatic expansion of this wealth of geographic data and information has not necessarily been matched by increases in the ability of users to take full advantage of the available resources, or to make sense of similar products. This plenary panel set the stage for the workshop by addressing user needs for data access and documentation.

**Dr. Dennis Ojima** of Colorado State University and the Land Science Project of the Earth System Science Partnership (ESSP) spoke on the user needs for data in the area of global environmental change science. He noted that among the challenges of serving this community is that global change science involves the study of multiple stresses, interactive sectors, and increasing human pressures. In addition to this, there are multiple users – scientists, managers, policy makers and the public.

Developing and testing theory and models requires integration of complex *in situ* process data with large gridded data sets. The required data are multi-scale, come in many formats, and originate from multiple disciplines. To maximize user control of information systems requires a rapid prototyping and development cycle, which implies incorporating existing state-of-the-art components rather than building things from scratch. Data systems must allow user-driven, knowledge-based querying of multiple data types.

In summary, Dr. Ojima mentioned that the data are available in many cases, and the understanding is well formulated. However, there is a mis-match in end-user needs (e.g., researchers vs. managers), and in the analytical tools used by end users (e.g., integration of decision making tools with research models). He added that scaling of information – to get from the individual observations to the level required by users – is inadequate

**Dr. Robert Ford** of Loma Linda University presented on the Geographic Information for Sustainable Development (GISD) initiative, an activity involving many partners that he coordinated from his previous position at the U.S. Agency for International Development. Provocatively titled "Who needs spatial data?," much of Dr. Ford's presentation focused on the digital divide. He characterized the data world as being split between the elites and those who are at risk of being left behind. Among the latter are the local-level users in developing countries, such as land managers, nongovernmental organizations (NGOs), community groups, farmers, small businesses, and churches. He provided a case study of coastal communities in Honduras that were impacted by Hurricane Mitch in 1998, and today are coping with unfettered coastal development for tourism (Figure 1). He argued that local communities and decision makers need geographic information in order to cope with natural disasters and to plan development effectively. But, he asked, "how do we get the health and sustainable development community more involved in the SDI-building process so that both the poor and vulnerable truly benefit?"

Finally, he observed that many academic programs in U.S. universities are unable to fully make use of geographic data and information. He argued that GIS needs to be integrated into public health programs and many other areas.

**Ms. Jillian Thonell** presented on the challenges of global data from the perspective of the Millennium Ecosystem Assessment (MA). The MA is an ambitious effort to understand the changes to global ecosystems, to develop likely scenarios for future changes based on what is known of the drivers of such changes, and then to develop responses. The primary audience for the assessment is the major environmental treaties, followed by policy makers, the private sector and civil society. Ms. Thonell described the sources of the data utilized in the MA, including models (hydrologic,

**Figure 1. Contrasting land uses in Honduras.** *Local communities and land managers need spatial data in order to better manage tourism development.*


population, climate and land-atmosphere, ecosystem process models, and global terrestrial ecosystem models), remote sensing, inventories (natural resource, biodiversity, socio-cultural), indicators of ecosystem condition, and indigenous knowledge. On the positive side there are many useful global data sets that are freely available and easily downloadable through the Internet. Among the challenges the MA has experienced, there is a lack of trend data on biodiversity and land-cover changes, and there are gaps in global spatial maps of invasive species and cultural diversity.

On the positive side, she noted that there are multiple data sources covering the same thematic area. However, there are difficulties in assessing data quality – for instance, what is the "expert opinion" versus actual "ground truth." At times, regional data sets can be useful in helping to interpret the global data, since many global data sets are at too coarse a scale to be useful for local or even national level analyses.

Finally, she praised the existence of geospatial data clearinghouses, but said that they have the potential to leave users confused (with too much information or technology/science jargon). Furthermore, although metadata are good, there is a need to provide tools to choose from multiple sources, such as by providing information on strengths and weaknesses or an indication of the quality of the data set. For future data development, it is important to assess the data that were available for the MA and carefully plan mechanisms for filling the gaps.

**Mr. Craig Beech** of the Peace Parks Foundation (PPF), based in southern Africa, presented a regional perspective on user needs. The PPF has identified areas for tranboundary parks and conservation projects based on a number of different data layers, including land cover maps and roads based on data from the Advanced Very High Resolution Radiometer (AVHRR). The areas with minimal human influences, and where the benefits for peace and security are greatest, are those selected for so-called "Peace Parks" (Figure 2).

**Figure 2. Conservation areas (in brown) in southern Africa.** *Priority transboundary areas identified through an overlay of several global data sets.*

### *Discussion*

The Chair of the session, **Dr. Stanley Wood** of the International Food Policy Research Institute, initiated the discussion. He said that decisions are made at the country and community levels that affect global change. In some senses there are two meanings to the term "global data sets" – one is the traditional meaning of a global-scale data set covering the world, and the second addresses data sets that are required everywhere (in the sense of being widely applicable to many problems), and that may require a different, presumably higher, spatial resolution.

He also observed that it is very helpful to the user to know the quality and reliability of the spatial data they are using. In an ideal world one might have one map with the data and another with the underlying reliability. Finally, he noted that getting longitudinal data is difficult; historic information is often available, but underutilized.

Jeffrey Tschirley of FAO noted that, as a user himself, policy makers often have difficulty defining their own needs. Returning to the question of the digital divide, the question was posed: How do we get this information to communities, those who are not part of the "digital" world? Can these data be extracted from digital format? Glenn Hyman of the International Center for Tropical Agriculture (CIAT) suggested that we need to develop "thin" clients or tools to be able to access the data. Someone suggested that there should be a shift from providing raw data to providing knowledge-level services in accordance with user needs. For the latter, data and tools need to be integrated to achieve a useful purpose.

## *Plenary Panel 2. Global Spatial Data and Information Dissemination Initiatives*

This plenary session addressed data development and dissemination efforts in four major thematic areas: agriculture and food security, the environment, health, and population and poverty.

**Dr. Robert Zomer** of the CGIAR Geospatial Data Initiative began with an overview of *agriculture and food security* data initiatives, prepared jointly with **Dr. Tim Robinson** of the FAO. The following organizational websites were among the data sources he referred to in his presentation:

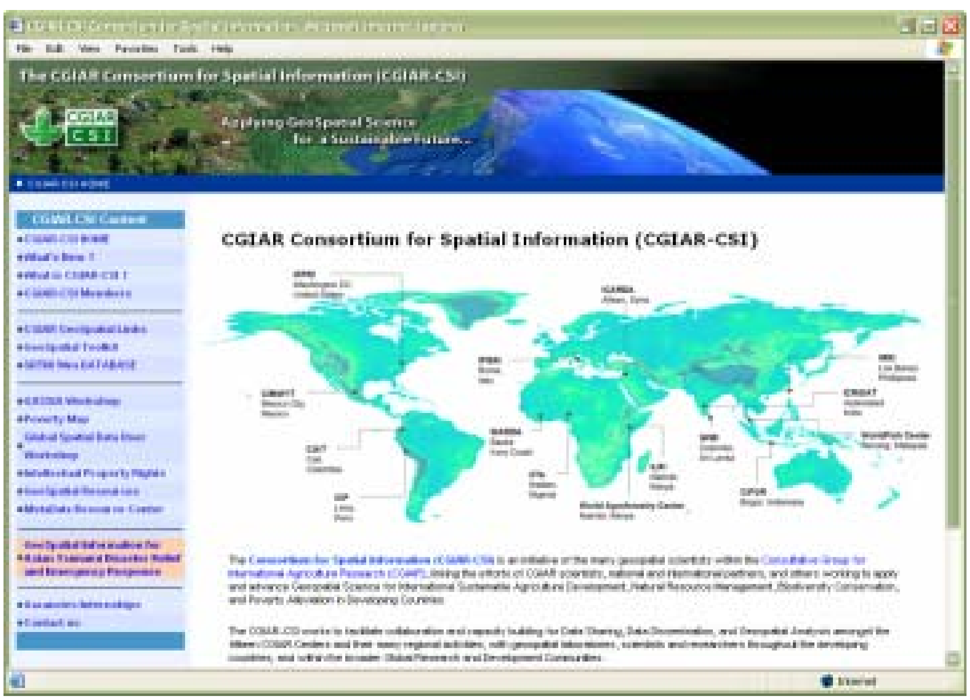
- The Global Spatial Data Infrastructure Association (http://www.gsdi.org).
- FAOSTAT (http://apps.fao.org).
- FAO's Statistics Division, which has a number of mapping services (http://www.fao.org/es/ess/).
- FAO's GeoNetwork data repository (http://www.fao.org/geonetwork/).
- The Famine Early Warning System (FEWS) for data on food security (http://www.fews.net).

- Food Insecurity and Vulnerability Information and Mapping Systems (FIVIMS), networks of national information systems that assemble, analyze and disseminate data on food insecurity (http://www.fivims.org).
- FAO's Global Information and Early Warning System (GIEWS), which includes an online mapping package called the "GIEWS Workstation" (http://www.fao.org/giews/). GeoWeb also offers capabilities to view thematic maps and some satellite imagery (http://geoweb.fao.org).
- FAO's Global Livestock Production and Health Atlas (GLiPHA) (http://www.fao.org/ag/aga/glipha/).

The remainder of his presentation focused on CGIAR's Consortium for Spatial Information (CSI) (http://www.csi.cgiar.org/). The CSI works to facilitate collaboration and capacity building for geospatial data sharing, dissemination, and analysis among centers of the CGIAR. Centers are responsible for various aspects of the work: The International Water Management Institute (IWMI) provides coordinated data management and tools, the International Maize and Wheat Improvement Center (CIMMYT) focuses on the geographic dimensions of crop varieties, the International Food Policy Research Institute (IFPRI) on impact assessment and policy research, the International Center for Agricultural Research in the Dry Areas (ICARDA) on natural resource degradation, and CIAT on poverty mapping. The CSI as a whole focuses on integration, training, and capacity building. Collectively, partners in the CSI have a wide range of regional and global data sets, including a world water and climate atlas, forestry data, and spatial data on coral reefs.

Among the data gaps in this area, Dr. Zomer pointed out the following. For farming and production systems, he noted that there is a need to develop a hierarchical classification similar to the Land Cover Classification System (LCCS) and greater disaggregation of the "mixed farming" category. For global irrigated areas mapping, he said there is a need for spatially disaggregated country statistics, maps using remote sensing, and accounting for

**Figure 3. CGIAR Consortium for Spatial Information website.** *This consortium is a model for data sharing among organizations working within a specific thematic area.*



small irrigation systems. Beyond data, the most significant gap in his assessment is the poor capacity at the national and local levels to utilize spatial data, and low perceived demand at both ends of the user spectrum. He echoed the thoughts of Dr. Ford about the need to bridge the geospatial digital divide by creating sustainable two-way interactive data exchanges and information flows across scales.

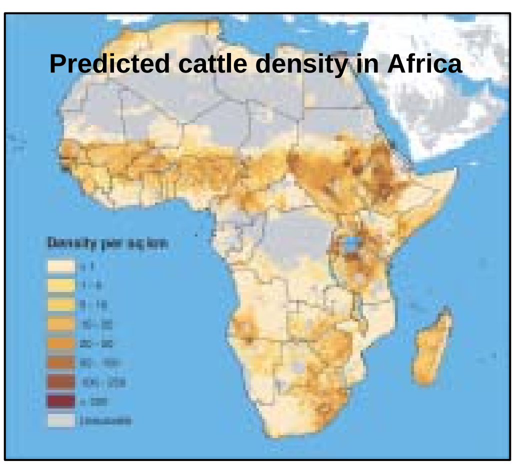
**Dr. Ashbindu Singh** of UNEP North America then presented on *environmental* data. He pointed out that environmental data fall into three broad categories: textual data (e.g., treaty data bases), statistical data (e.g., measurements of air and water quality), and spatial data (e.g., forest cover, protected areas). Among the problems he cited in environmental databases is that many of them are dated, they lack comprehensive coverage, their quality is uncertain, and they cannot tell us what is happening where in a scientifically credible fashion. The availability and quality of environmental data are a major concern. There is a lack of time series GIS data layers that could be used to influence decision makers. Fortunately, remote sensing and GIS technologies are increasingly providing significant input to data generation and analysis. But generally, he observed that there is more of a focus on access mechanisms (web sites) than on content development. Finally, there are constraints in data development insofar as geospatial data handling technologies are still quite expensive and rather difficult to use.

He suggested that some of these constraints could be met by creation of more geocoded, orthorectified satellite images, and development of more user friendly technologies for data analysis. He also said that the global data community needs to showcase data applications that have solved real world problems in order to create greater support for its efforts.

Dr. Singh advocated for an operational program to monitor the terrestrial environment on an annual basis and generate policy relevant information. Furthermore, he would like to see a comprehensive system to pull together and analyze the rich data collections available from multiple sources (i.e., economic, social, cultural and environmental) and present results which both specialists and non-specialists can comprehend.

**Box 1. FAO's Global Livestock Information System**

One of the major limitations in livestock planning, policy development and analysis is the generally poor quality and accessibility of information on the distribution and abundance of livestock. With the objective of redressing this shortfall, FAO's Livestock Production and Health Division is developing a global livestock information system in which available data are collated and standardized, gaps are filled by predicting livestock numbers based on empirical relationships between livestock densities and environmental variables in similar agro-ecological zones, and the information is made publicly available through an interactive web application, the Global Livestock Production and Health Atlas (GLiPHA) (http://www.fao.org/ag/aga/glipha/), and through FAO's GeoNetwork data repository.




The spatial nature of these livestock data facilitates analyses that include: estimating livestock production; mapping disease risk and estimating the impact of diseases of livestock; estimating environmental risks associated with livestock due, for example, to land degradation or nutrient loading; and exploring the complex relationships between the many components of human wellbeing and livestock.

Predicted livestock distributions have now been produced, at a spatial resolution of 3 minutes of arc, for the entire globe for cattle/buffalo, sheep, goats, pigs and poultry/chickens. Examples of the observed and predicted cattle distributions for the African continent are given below.

*Based on a presentation by Tim Robinson (FAO).*

**Dr. Steeve Ebener** of the World Health Organization presented on global *health* data sets. He mentioned that there is a paucity of data on health facilities – As an example he mentioned that out of 600,000 facilities known worldwide from country censuses, only 50,000 have spatial reference information. There is also a great need for health district boundary data. This complements work of the UN Geographic Information Working Group (UNGIWG) to develop a global Second Administrative Level Boundaries data set (SALB). He also mentioned a number of gridded data sets in the health arena that are useful, such as the MARA model of climate suitability for malaria.

One data service that Dr. Ebener mentioned was the WHO Global Atlas of Infectious Diseases (http://www.who.int/globalatlas/). WHO has recently launched an online GIS Metadata Explorer based FAO's geonetwork data repository tool (http://www.who.int/geonetwork/). The Southeast Asian Regional Office Integrated Data Analysis System is a common integrated tool for surveillance, data collection activities, analysis and presentation using indicators and to prevent duplication of work across various technical programs in the region (http://193.220.119.50/sidas/). However, these sites remains password protected for the moment and therefore inaccessible to the general public. One of the main obstacles to data dissemination in the public health field is confidentially. As a result, the health sector may have the fewest global spatial data sets of the sectors covered by the workshop.

**Dr. Deborah Balk** of CIESIN then spoke on the topic of *population and poverty* data. She began her talk with a brief review of the evolution over time of population data availability. She noted that, since 1900, when only population counts were available for developed countries, the population data community has paid more attention to global scope, to data comparability, to problem-oriented science, and to spatial frameworks. Next, she addressed the development of globally consistent population data sets such as the Gridded Population of the World (GPW) dataset distributed by SEDAC (http://sedac.ciesin.columbia.edu/gpw). The first GPW was released in 1994 and had only 19,000 input units (an input unit is a national- or sub-national administrative unit); in 2000 the second version was released with 127,000 input units and in 2004 the third version was released with 375,000 input units. Since GPW, other products such as LandScan (http://www.ornl.gov/gist/landscan/) have been developed, as well as projected and backcast spatial population databases (e.g., RIVM's HYDE, http://arch.rivm.nl/env/int/hyde/).

**Figure 4. Extract from the Global Urban-Rural Mapping Project urban extents database.** *Roads connect the cities, but global road databases are of poor quality.*

She next spoke about urban population data and CIESIN's new Global Rural-Urban Mapping Project (GRUMP). The project has produced three data sets: settlement points (more than 70,000 settlements covered), urban extents (more than 23,500 urban areas with population of greater than 5,000), and a population grid reallocated to urban areas at 1 km resolution. One of the gaps she mentioned in passing is the need for better roads data for a variety of population and poverty mapping applications.

Turning to poverty data, there has been a dramatic increase in the past three years in efforts to produce poverty maps. CIESIN has developed poverty maps for the UN Millennium Development Project (Sanchez *et al*. 2005). Poverty Map, a consortium of UNEP, CGIAR, CIAT, and FAO (see http://www.povertymap-.net), has also been generating maps. However, the data sets themselves are often difficult to access, and at the current time are only exchanged amongst researchers.

In terms of the future of data development in this area, Dr. Balk painted two contrasting pictures. In one scenario, there is a continuation of the status quo, with "stove piping" by disciplines and institutional bureaucracies, inadequate integration, failed opportunities for cross-fertilization, redundancy, and unproductive competition. In the other scenario, there is a quantum increase in cooperation and institutionalization, coordination between groups, pooling of input data sources, a division of tasks, guidelines on common products, endorsement from national statistical offices and UN agencies, and productive interaction on methodological hurdles. If the latter scenario prevails, it will result in more relevant, efficient and timely outputs.

### *Plenary Panel 3. Inventories of Global Spatial Data, Standardization, Search, Portal Development*

The presentations of Plenary Panel 2 highlighted many activities related to the development and dissemination of global data sets. In this plenary, the emphasis was on core data sets for sustainable development research and operational programs, efforts to develop standards, and the development of portals and online mapping tools for data access.

**Mr. Joseph Dooley** of Spatial Data Services & Mapping presented work related to two projects, the African Water Resource Database (AWRD) and the FAO Inventory and Comparison of Globally Consistent Geo-Spatial Databases and Data Libraries (CGDB). According to Dooley, the production of AWRD, which is a seamless continent-wide geospatial database of water resources down to the catchment level, provided important insights into the availability and problems associated with existing baseline framework data sources such as the Digital Chart of the World (DCW or VMap0). The structural organization and storage methods with which each of the global source databases are encoded have limited their adoption within the international development community.

The FAO CGDB Inventory attempts to itemize data sources that can be used to support general base mapping, emergency preparedness and response, and potential analytical sources of geospatial data for the FAO Poverty Mapping Project Group (PMPG), the FAO as a whole, and the UN. The inventory builds and expands on the 16 core data layers identified by the UNGIWG CGDB Task Force and is designed to provide a baseline for UNGIWG and the FAO PMPG to build on. The inventory is also potentially unique in that it:

- Categorizes each CGDB layer into one of ten topical indexes;
- Identifies potential global baseline framework data libraries and other sources in both the public domain and from commercial sources; and
- Evaluates the sources identified by accessing, processing, and determining the estimates of the level of effort required to make the data comparable and directly useable in relation to each CGDB layer.

A large number of "framework" data sets were evaluated, including Vector Smart Map Level 1 (VMAP-1), DCW, GTOPO30, and Hydro1k. In addition, the CGDB inventory looked at data in the following topical areas:

- Boundaries: coastal, administrative, and areas of special interest
- Human Health: boundaries and facilities
- Human Population: population centers and distribution
- Transportation: roads, railways, airports, harbours, and navigation routes
- Bathymetry and Elevation Databases
- Geophysical: geology, geo-morphology, seismic, hydro–geology, and soils
- Surface Hydrology: waterbodies & water points, drainage, and watersheds
- Satellite imagery, orthorectified mosaics, land cover & vegetation data

- Climatic Data: temperature, rainfall, and atmospheric emissions

The inventory is available in electronic format on request from Mr. Dooley (see Annex 3).

**Mr. Allan Doyle** of EOGEO (Earth Observation & GEOspatial) spoke next on closing the interoperability gap. Mr. Doyle suggested that there have been great advances in geospatial data standards in the last several years, but for any given application, bringing together geodata, the technology, and beneficiaries is still difficult. He argued that we need more than just standards. He also noted that interoperability can foster solutions to data inconsistency issues, insofar as the inconsistencies become more obvious, users demand solutions, and data and service providers start to address them.

In closing, he argued that there needs to be more testbeds, prototypes and pilots. Also, in light of the fact that many organizations are setting up web services (NGOs, government agencies, and the private sector), the community needs to weave operational threads throughout those services to show what can be done with the data. They need to be reproducible, usable in developing countries, and usable by non-technical people.

**Mr. Richard Cicone** of ISciences, LLC spoke on data fusion. In the lexicon used by Mr. Cicone, geospatial data include geographic data (raster and vector formats), geostatistical data, and geotextual data. Geostatistics are data that are related to a national or subnational unit, and therefore can be georeferenced. Geotextual data are text databases, such as treaty databases, that are linked to some geographic entity, be it national or regional. He argued that data fusion is not just overlaying maps. Integration of geospatial information elements is central to understanding a particular problem. Yet integration is challenging, particularly when boundaries do not match. Maps are created using different vectors for features such as coastlines. For example, a land cover map may identify an area such as a bay or inlet, whereas a population map may show people living there. Figure 5 shows how political boundaries for a small island state do not match the elevation and bathymetry data from another data set. Shorelines are often defined by an independent water mask. The resulting position and shape of an island varies from one dataset to another.

**Figure 5. Mismatch between elevation data and political boundaries data for Wallis and Fortuna Island, South Pacific.**

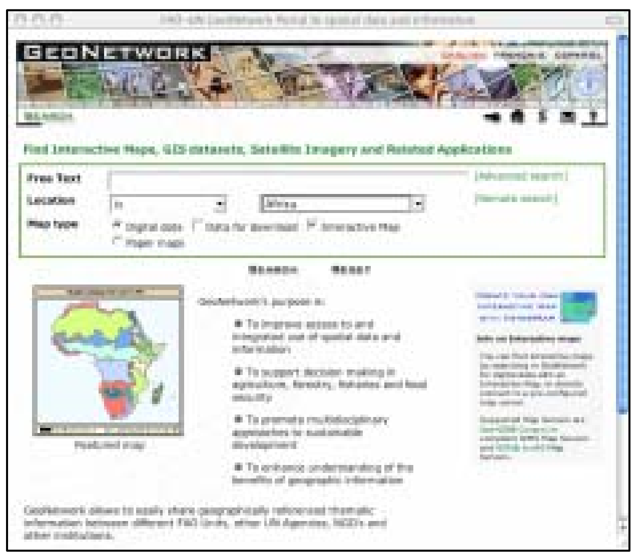

Turning to geostatistical data, he pointed out that difficulties arise with matching data to map entities because there are no standard names or codes that are used universally. Thus, the process can rarely be fully automated, since someone is needed during the import phase to resolve issues that arise. Furthermore, geographic entities may be in the midst of political dispute (e.g., Taiwan), or they may change names during the span of time-series data (e.g., Germany, Yugoslavia). To come up with consistent time series, do you combine figures, divide them based on percentage  area or population, or just ignore the issue?

Finally, ways in which data for countries are aggregated varies from dataset to dataset (and sometimes variable to variable). For example, sometimes data sets aggregate to “developing nations” or “rest of Southeast Asia” (excluding some country in the region). It is difficult to normalize and compare such data. Meta-data and footnotes are often incomplete and scattered in separate documents. And the sum of subnational parts may not match the total provided by a different source. In such instances the researcher must decide how to resolve the problem, even though there may not be a clearcut solution. Although many challenges remain, he concluded by saying that as a consumer he is happy to see refinement of the available data.

Next, **Mr. Chris Chiesa** of the Pacific Disaster Center spoke on the Asia Pacific Natural Hazards Information Network (APNHIN). The APNHIN vision is to engage participants in formation of a distributed information network to foster sharing of geospatial data resources supporting natural hazard risk and vulnerability assessment and mitigation applications in the Asia-Pacific region. The APNHIN leverages advances in Internet and spatial data technologies, advances in commercial-off-the-shelf (COTS) software and other standards, and investment in development of data resources in order to develop and apply a common geospatial data framework for hazard data. It also fosters improved understanding of availability of data resources, improved understanding of hazards, vulnerabilities, and risk, and improved development and mitigation strategies. APNHIN provides dynamic mapping of real-time data through automated acquisition and processing, as well as a searchable metadata catalog for access to downloadable data. Among the current layers available are data on tropical storm tracks, earthquakes, volcanoes, wildfires, Hydronet raingages, and global cloud cover. He encouraged centers based in the Asia-Pacific region to become members of APNHIN.

In the final presentation of this session, **Mr. Kees Klein Goldewijk** of the Dutch National Institute of Public Health and the Environment (RIVM) spoke on global spatial data sets and user needs at RIVM. He noted that RIVM serves in a number of international data networks. Key areas of interest to the modelers and researchers working at RIVM include future land use, food, sea-level rise, climate scenarios (e.g., IMAGE, FAIR), historical land use and population (HYDE data base), and global gridded emissions (EDGAR database). They publish their data to Geography Network.

**Box 2. Presentation of a test bed activity in southern Africa**




FAO is seeking to strengthen information systems for environmental and natural resources management for food security and poverty alleviation in Africa. The current focus is on the southern African region, which faces a variety of food security challenges, ranging from chronic poverty, climate variability, market failure, to HIV/AIDS pandemic. The effective targeting of food security interventions in this context requires a consistent, open, and transparent decision-making framework ensuring sound and high quality decisions in accordance with the various and often conflicting stakeholder objectives. The FAO decision-making framework is supported by a variety of information tools and methodologies allowing the integration of qualitative and quantitative information describing the multidisciplinary and multidimensional aspects of food security. GeoNetwork is a key component of the FAO proposed framework. GeoNetwork, jointly developed by FAO and WFP, is a standardized spatial information management infrastructure, that (1) provides users with a global data library, metadata catalog, system for searching, editing, publishing geospatial information, and web-based mapping applications, and (2) implements the International Standard for Geographic Metadata (ISO19115) and the Open Geospatial Consortium Specifications. A practical application of this framework in Mozambique demonstrated that GeoNetwork and other FAO information tools and technologies could enhance collaboration and communication across sectors by allowing national institutions to access, manage, analyze, and publish data to better achieve the specific mandate of their institutions.

*Based on a presentation by France Lamy (FAO).*

He suggested a fruitful focus for collaboration and standards would be on availability of data and keeping it simple. For example, the Interrupted Goode Homolosine projection is not simple by comparison with maps in latitude/longitude coordinates. More than half a researcher's time can be consumed on data search, data conversion, and handling. Furthermore, he suggested that the global data community should agree upon some basic datasets such as global administrative boundaries and use of International Standards Organization (ISO) codes. It is also important to let other people know what you have by publishing metadata and data in the most commonly used catalogs. Creating one portal for people working in the thematic areas addressed by the workshop would help, since currently multiple versions of the same data set can be found in different locations.

### *Plenary Panel 4. Potential Roles for Global Data Coordination Mechanisms*

In all of the plenary presentations there was consistently mentioned a need for greater coordination, harmonization and standardization. This plenary addressed the roles – current and potential – of a number of different global data coordination organizations.

**Dr. Steeve Ebener** of the WHO began by presenting the United Nations Geographic Information Working Group (UNGIWG). The overarching objective of UNGIWG is to promote the use of geographic information within the United Nations System and Member States for better decision-making. Among the products being worked on in the context of UNGIWG are an agreement on standard international and second level administrative boundaries, an inventory of core global databases (see Dooley presentation in Plenary 3), and interoperable services such as GeoNetwork. Dr. Ebener heads the Second Administrative Level Boundaries data set project (SALB), the objective of which is the creation of a redistributable second administrative level boundaries global data set (a level below provincial or state level boundaries), covering the changes that have occurred since 1990 at the 1st level and since 2000 at the second level to be used with GIS technologies. Thus far, the SALB project has produced January 2000 administrative unit name/codes tables for 157 countries, historic changes tables for more than 100 countries and obtained validated digital maps for a total of 24 countries (http://www3.who.int/whosis/gis/salb/salb_home.htm).

**Dr. Hiromichi Fukui** of the Faculty of Policy Management at Keio University in Japan, presented on Digital Earth (DE). DE was launched in 1998 with the vision of creating a multi-resolution, three-dimensional representation of the planet, into which vast quantities of geo-referenced data can be embedded. The user interface would be a browsable, 3-D version of the planet available at various levels of resolution, built on a rapidly growing universe of networked geospatial information (see Figure 6 for a prototype). Vice President Al Gore was a major early proponent.

Subsequent to its launching in 1998, DE has hosted several international conferences in China (1999), Canada (2001), the Czech Republic (2003), and soon in Japan (2005). The technology behind DE, much of it incipient, includes visualization, high-speed computing, artificial intelligence, real-time computing, intelligent systems, search engines, data fusion, dynamic modeling, 3D rendering, and grid computing. The concept borrows from many of the standards, networks, and content that have been presented at the workshop.

Keio University is also promoting Digital Asia, an initiative to provide people and communities with easy access to geospatial information over the Internet at multiple scales by establishing a scheme to integrate and share the GIS and remote sensing data among all the countries of Asia. The focus is on practical applications, global change studies, sustainable development, and global environmental protection. The aim is to learn from this test bed activity to eventually make the DE a reality. In conclusion, Dr. Fukui noted that DE and Participatory Web GIS are a Gateway to link the real world and cyberspace, and to capture local knowledge and combine it with global data from satellites and other sources.

**Figure 6. The Digital Earth Viewer at Keio University.** *This tool provides an impressive array of tools for visualizing global data sets (http://geoinfo1.sfc.keio.ac.jp/MyMap/MyMap/SFC-JAXA).*


**Dr. Harlan Onsrud** of the University of Maine presented on the Global Spatial Data Infrastructure (GSDI) Association. The GSDI Association is a global forum to support the exchange of ideas and encourage joint activities. It has had seven conferences since 1996 at various locations around the world, with a forthcoming conference in Cairo, Egypt. The goals of the GSDI Association are to:

- support the establishment and expansion of local, national, and regional (multi-nation) spatial data infrastructures that are globally compatible;
- provide an organization to foster international communication and collaborative efforts for advancing spatial data infrastructure innovations;
- support interdisciplinary research and education activities that advance spatial data infrastructure concepts, theories and methods;
- enable better public policy and scientific decision-making through spatial data infrastructure advancements;
- promote the ethical use of and access to geographic information; and
- foster spatial data infrastructure developments in support of important world-wide needs.

The GSDI Association has a number of current projects. The publication 'Developing spatial data infrastructure' (i.e., the 'SDI Cookbook') describes how to implement SDI concepts within any nation adhering to interoperable standards. There is currently a Version 2 in several languages. Universal Description, Discovery, and Integration (UDDI) involves connecting islands of interoperability. The creation of hundreds of catalogues of data and services is inevitable but difficult to navigate, so GSDI is promoting a registry of all services at a global level. Finally, there is the Global Commons and Global Marketplace in Geographic Data, which redirects technological and legal approaches towards providing incentives for sharing locally collected data and facilitating data sharing.

In closing, Dr. Onsrud noted that the basic principles in establishing a GSDI Association were to: be inclusive; enhance democratic decision making processes; be as simple, transparent and open as possible; be flexible and adaptable to change; support sustainable development; build on and support existing initiatives; and engender partnerships.

**Dr. Karen Kline** of the University of California Santa Barbara and the Secretariat of the ISCGM spoke on the Global Map project, which is a globally consistent 1:1,000,000/1km data set based on the contributions of the national mapping organizations (NMOs) of the world. Its layers include elevation, vegetation, land use, and land cover. It will be updated regularly and available free-of-charge. Version 0 was based on previously available data (GTOPO30, IGBP land cover, DCW/VMap0) and distributed to participating NMOs. Version 1, released in November 2000, is based on data developed by NMOs or, depending on the country, an updated Version 0. For Version 2, there are discussions about adding additional layers and creating maps at different scales/resolutions. There are voting members, who represent NMOs, and nonvoting members such as NGOs and industry groups. Working groups develop specifications and evaluate the possibility of creating additional raster data layers.

ISCGM/Global Map have sought to develop capacity in many countries through education and training and software/hardware grants. They are building an ongoing relationship with NMOs, and the ISCGM works to facilitate participation and to raise funding and support. In terms of validation, only the IGBP land cover data sets have been validated. They are aware of the need for a method to determine accuracy, and they plan to establish a methodology utilizing the resources of NMOs to validate the Global Map.

**Mr. Jeffrey Tschirley** of the FAO supplemented his presentation on SIMAG, made during the opening plenary, to say that the idea is to create a “coalition of the willing” – in other words, an informal network that would not get bogged down in the creation of organizational structures but begin real work on applications. He said that he hoped terms of reference could be developed and circulated over the coming weeks.

### *Discussion*

Ron Weaver of the National Snow and Ice Data Center (NSIDC) asked if the SDI cookbooks say anything about best practices for developing portals. Alan Doyle of EOGEO responded that they could certainly be mined for information. Other models include the U.S. Federal Geographic Data Committee (FGDC) reference model and the Open Geospatial Consortium (OGC) portal model.

Alan Doyle posed a question about whether the distinctions between groups are getting lost because there are so many groups. He suggested the participants might wish to think how confusing it might appear to someone on the outside looking in. Robert Ford responded that the mandates of most of the organizations are fairly specific, but still the organizations need to look at how they might better serve those who are developing portals, data or capacity.

Robert Chen noted that participants may have detected that the organizers have a range of interests. This highlights the need for coordination and participation. Jeff Tschirley’s approach, based on a more informal process, may work. However, an informal group still needs to have links back to the organizations they represent. This is an opportunity to start a dialog.

## *Plenary Panel 5. Data Documentation, Quality, Preservation, and Intellectual Property Rights*

Institutional efforts related to data development, management, and dissemination need to be carried out in a coordinated way, or the potential for poor interoperability between information systems, incompatibilities between key datasets, confusing or even conflicting data documentation, and complex constraints on data re-dissemination and use is likely to increase. This plenary featured two presentations representing initiatives in the biodiversity community to develop institutional arrangements for data sharing.

**Mr. Jean-Louis Ecochard** of The Nature Conservancy presented on the Conservation Commons, an initiative among large NGOs in the biodiversity conservation community to facilitate data and information sharing. Adherents to the Commons basically subscribe to three principles. Principle 1 states that the Conservation Commons promotes free and open access to data, information and knowledge for conservation purposes. Principle 2 states that the Conservation Commons welcomes and encourages participants to both use resources and to contribute data, information and knowledge. Principle 3 states that contributors to the Conservation Commons have full right to attribution for any uses of their data, information, or knowledge, and the right to ensure that the original integrity of their contribution to the Commons is preserved. Users of the Conservation Commons are expected to comply, in good faith, with terms of uses specified by contributors and in accordance with these principles.

Contributors and users recognize that the creation of data, information, and knowledge has real costs, and that meeting such costs may require a range of compensatory mechanisms. However, contributors to the Conservation Commons seek to ensure that cost barriers of any kind will not prevent either contributions of resources or fair access to (or use of) conservation-related data, information and knowledge. They acknowledge that some data and information are restricted in terms of their use (e.g., because they are confidential or sensitive) and thus will not be considered part of the Conservation Commons. Providers and custodians of data and information have complete control of their data/information and have complete liberty to decide what data or information will be part of the Conservation Commons.

**Dr. James Edwards** presented on the Global Biodiversity Information Facility (GBIF). GBIF's goals are to make the world's biodiversity data freely and universally available via the Internet, to share primary scientific biodiversity data (especially georeferenced data), and to promote the development of biodiversity informatics around the world. Primary biodiversity data includes label data on some 1.5 to 3 billion specimens in natural history collections, herbaria, and botanical gardens, and their associated notes, recordings, etc. It also includes observational data such as bird banding data. Most of these data have been amassed over the past 300 years and are not digital.

Data providers make their data known through GBIF's Registry of Shared Biodiversity data, which currently serves more than 41 million specimen and observation records from more than 250 collections. The electronic catalogue of scientific names (ECAT) will be available as an authority file to any user.

In terms of georeferenced data, most new data meet the recommended standards, but legacy data are a huge problem. They are supporting the development of automated tools to handle the legacy data.

In terms of intellectual property, GBIF claims no intellectual property rights on data served through its portal. Control of data remains with data providers. Each provider decides which data to serve, and sets its own policy for data re-use. Most providers do not allow commercial use of data, but they differ in what this means to them. GBIF has held one IPR workshop and is planning others.

**Mr. Paul Uhlir** of the National Research Council deferred his comments on creating an information commons for a global spatial data infrastructure because they were almost identical to his presentation of the Breakout Group 1 report (see below).

# Summary of Breakout Group Discussions

A significant portion of the workshop was dedicated to breakout group discussions and reports back to plenary. Each breakout group was given a concrete goal (or set of goals) as well as a list of discussion questions. These are provided in Annex 2, along with the list of participants for each group. At the end of each breakout group report there is a summary of the discussion that took place in plenary following the oral presentation of each group's report.

## *Group 1: Principles for Data Sharing and Access*

This group focused on the elaboration of principles for data sharing, attribution, and dissemination, including possible commitment to dissemination of data sets as global public goods. A secondary agenda item was to develop agreement on the role and mandate of the recently initiated Spatial Information Management Advisory Group (SIMAG).

There are two main categories of global spatial data users:

1. Public Users (non-commercial): These include academic researchers; decision-makers; international organizations; inter-governmental organizations; national, regional, state/provinces, local governments; NGOs; donors; and the education community.
2. Private Users (commercial): multinational firms; businesses that have operations all over the world (e.g., global transportation); value added geospatial technology companies.

Within these, there are both end users (consumers) and value-added users (data integrators, re-users and re-disseminators).

There are different characteristics of data that are relevant to data sharing or access policies. These characteristics include the spatial resolution, temporal resolution (real time versus retrospective data), single event versus longitudinal data, and original source data versus amalgamated processed data.

Concerning the rights to use and re-disseminate data and derived products, different kinds of users may need different pricing/licensing. Data generation is demand-driven though not necessarily market-driven (especially from public sector sources). Different users have different requirements depending on the nature of their work or research questions. Users generally want the lowest possible restrictions, both in terms of the purchase price and legal restrictions on obtaining and re-using the data. Users in developing countries have the greatest need for free and unrestricted access, because even a small fee for access can be prohibitive for them.

A major question is, what aspects of a global dataset's intellectual property need to be protected? There is a distinction between intellectual property protection (statutory public law such as copyright or database protection laws in the E.U.) and contracts and licenses (private law). A further distinction is whether the data source is governmental, commercial, or non-commercial (NGO or educational sector). Hybrid, multi-source databases, which are typical in value-added geospatial databases or data products, have more complex information law considerations.

Data produced by the United States government have no intellectual property (IP) rights, which means that the data are in the public domain. Other countries also have more limited IP protection for government data and information (e.g., the European Union Directive on Public Environmental Information). Every data source operates within a set of norms unique to its disciplinary community, program, institutional and national context, which determine whether the source wishes to assert IP rights or provide access on more liberal terms and conditions. Data sources that seek cost recovery or commercial exploitation of their data generally wish to apply the maximum legal protection available. Typically, most data sources wish to have attribution whenever possible.

Other reasons for a data source to protect its data include national security, privacy, and confidentiality reasons. Data sources that are not in the public domain do not want the entire database misappropriated or misused.

Data distributors (i.e., third-party intermediaries) need to have sharing agreements with data sources and data users. Motivations and needs of data distributors typically parallel data sources. Data distributors may add additional restrictions beyond the source requirements.

The breakout group was tasked with addressing two important questions:

1. What are the pros and cons of making data available in the public domain or under open access terms?
2. Are there alternative approaches that might provide suitable intellectual property protections to data sources and/or owners, yet at the same time promote widespread data use and application?

The group's consideration of these issues was limited to the public sector – i.e., data and information sources produced by or funded by the government. There are a number of compelling reasons for placing government-generated geospatial (and other) data and information in the public domain with open access.[1] Firstly, a government entity needs no legal incentives from exclusive property rights to create information. Both the activities that the government undertakes and the information produced by it in the course of those activities are a public good. Secondly, the taxpayer has already paid for the production of the information. Thirdly, transparency of governance is undermined by restricting citizens from access to and use of public data and information. Similarly, individual rights are compromised by restrictions on re-dissemination of public information, particularly of factual data. Fourthly, numerous economic and non-economic positive externalities – especially through network effects – can be realized on an exponential basis through the open dissemination of public-domain data and information, e.g., through the Internet. Finally, many public geographic data and information resources are global public goods.

There are countervailing polices and practices that limit the free and unrestricted access to and use of government information, including public geospatial data and info. These include: (1) Legitimate statutory exemptions to public-domain access and use, and to Freedom of Information Act (FOIA) statutes (e.g., based on national security concerns, the need to protect privacy, and to respect confidential information, among others). (2) Government-generated data are not necessarily provided without cost, even if there are no restrictions on reuse. (3) Government agencies sometimes may be prohibited from competing directly with the private sector in providing information products and services. (4) Government agencies typically enforce the proprietary rights in data and information originating from the private sector that are made available for government use or, more generally, for regulatory and other purposes, unless expressly exempted. (5) Many science agencies and research programs allow their researchers to have exclusive use of data for a specified period. (6) Despite mandates, government agencies may fail to actively disseminate data and information, or to preserve them for long-term availability.

There are several implications of requiring IP restrictions and high fees for access to public geospatial data. Firstly there would be a general diminution in the scope of geospatial data and databases in the public domain that can be openly accessed and used in downstream applications. Secondly, sole-source provider (monopolization) problems would be exacerbated in geospatial database markets. Thirdly, there would be higher administrative transaction costs. Fourthly, there would be less data-intensive research and significant opportunity costs, including reduced social and economic benefits from public investments in geospatial

[1] Note: the remainder of this sub-section, up to the next two questions, is based on Paul Uhlir's presentation materials, which themselves were based on Uhlhir (2003).

data resources. Lastly, there would be less effective international, inter-institutional, and interdisciplinary cooperation using digital networks.

For geospatial data and information produced with government funding, government-funded databases and other forms of information in the non-commercial public sector (e.g., academia and NGOs) should be assumed to be protected by intellectual property laws unless funding sources require sharing or open access. Sources of data and information in the non-governmental and non-commercial sector typically will want to promote and enforce attribution for their work, using IP or licensing terms.

The ideal legal and policy regime would support the open availability and unfettered use of public data, and place a premium on the broadest possible dissemination and use of geospatial data and information produced by government and government-funded sources by (1) expressly prohibiting or reducing intellectual property protection of geospatial data and information produced by government; (2) actively promoting and contractually reinforcing the cooperative, sharing norms of science through open access terms for government-funded data and information in government research grants and contracts; (3) carving out a large and robust statutory public domain for non-copyrightable data, as well as other immunities and exceptions favoring science, education, and other public interest uses; and (4) disseminating data and information freely or at no more than the marginal cost of reproduction and distribution.

Another set of questions the group considered were the following:

1. What kind of bi-lateral or multi-lateral agreements might be developed among major data developers and distributors that would obviate the need for case-by-case arrangements for data sharing?
2. Might these agreements cover alpha versions or data still under development?

Intergovernmental organizations (e.g., UN specialized agencies and regional organizations, such as the European Commission, the Organization of American States, the Organization for Economic Co-Operation and Development, etc.) have a major role to play in setting data and information policy standards relating to access under least restrictive conditions, as outlined above. So do major operational and research programs, such as the proposed Global Earth Observing System of Systems (GEOSS), the World Climate Research Programmme (WCRP), and many others. Finally, organizations focused on geospatial data and information activities and issues have a significant role to play. Because they have many government members and related stakeholders that are both producers and users of geospatial data and information, they have an interest in adopting rules and mechanisms that: (1) maximize the value of such information, and (2) minimize the transaction costs and other deadweight social costs associated with over protection and unnecessary restriction on access to and use of these public information resources. They also have the ability to consider specifically the needs of developing countries. Some notable examples of data exchange policies and practices already exist, such as those developed by the World Meteorological Organization (WMO). The best practices and optimal terms and conditions from the existing information regimes can be compiled and extracted for consideration and adoption by other multilateral organizations that have a recognized need for developing or updating their information policies and practices.

Lastly the group considered if it is possible to establish clearer and more consistent definitions and practices regarding use and dissemination rights with respect to data re-dissemination; attribution of data sources, integrators, owners, and disseminators; and use of derived imagery and statistics vs. raw data. Many examples of good or best practices already exist, as noted above, which address these different elements. Some of these have already been compiled and analyzed in recent reports like *Licensing Geographic Data and Services* published by the National Research Council (2004). Model license terms are also available on the Creative Commons Project Website (www.creativecommons.org)

Finally, the group agreed that international coordination mechanisms such as CODATA, SIMAG, GSDI or UNGIWG are essential to promoting data sharing and access, and to developing agreements on how the geospatial data community will work together.

***Plenary Discussion***

Following the presentation of the working group results, Alex de Sherbinin of CIESIN asked if there are similarities between national governments and inter-governmental organizations (IGOs) in data access policies. Mr. Uhlir responded that since IGO members are nation-states, there are. Policies within IGOs are treaty-based and specify the data available. Uwe Deichmann of the World Bank observed that taxpayers should not pay twice for data, but the problem within developing countries is that there are no stable funding mechanisms to produce data, so producers rely on data sales to fund data collection. Mr. Uhlir agreed that this is a thorny problem. There are different mechanisms for creating sustainable data collection/data centers, but it is difficult to get donors to pay attention to data distribution issues. Regional data consortia are an option.

Alan Doyle of EOGEO suggested that data users need to make the argument that open data access is better. Perhaps studies should be conducted that show the efficiencies gained from open access, or economic externalities of having basic data available free of charge. Joe Dooley of Spatial Data Services & Mapping asked if there are mechanisms like the GNU general public license that restrict commercial uses of data. Mr. Uhlir responded affirmatively, mentioning the 2004 National Research Council report on licensing cited above and the http://www.creativecommons.org website.

### *Group 2. Gaps in Global Data and Collaboration in Data Development*

The group decided not to cover framework data, including boundaries, coastlines, roads, rivers, and gazetteers. The need for high quality global data sets on these topics is undisputed, and work is already being carried out in several of these areas. Examples are the Second Administrative Level Boundaries (SALB) of UNGIWG and Global Map (described in Plenary Panel 4). Roads, however, were repeatedly mentioned as one of the priority data sets in the plenary. This will require a special collaborative emphasis. Significant private sector involvement with roads data may provide a useful model, but there are also major intellectual property concerns to overcome. A separate workshop for planning roads data development was discussed among several of the organizations present at the meeting.

The data gaps group reiterated the need for generating a comprehensive baseline inventory of global data sets. All current examples of such inventories are either not comprehensive in scope (e.g., the list drawn up by the UNGIWG, which does not include data from outside the UN system), or lack mechanisms for update (e.g., the Millennium Ecosystem Assessment). It was agreed that a comprehensive annotated database of datasets, with a realistic update mechanism, must be a priority of the community, and one of the larger organizations involved in this field (e.g., UNEP or FAO) are potential leaders in pursuing such an effort.

The group organized the discussions by asking point persons to suggest priority data sets in the following core areas: Health, Food and Agriculture, Land Use and Land Cover, Population and Poverty, Hydrology, Institutions, and Biodiversity. The group member suggested specific data gaps to be considered, followed by a group discussion. In order to focus discussions, the group defined several criteria for priority datasets, including:

- Are the data set definitions sufficiently precise to be operational, but not so narrow that they do not offer a broad range of uses?
- Do they have a wide potential user group with a high level of expressed interest?
- Do they support broader initiatives such as the MDGs or the Millennium Ecosystem Assessment?

- Are they feasible and updatable, within a realistic budget and time frame?
- Are the data relatively simple and unprocessed rather than the result of significant downstream modeling?

Aspects to be considered for each data set are:

- Data set description.
- Who cares about the data?
- Data development model (e.g., recovery of existing national data, remote sensing, etc.).
- Potential champions.
- Potential collaborators.
- National-level buy-ins and links (e.g., through mapping or statistical agencies).

### *Health*

The main priorities in the health sector are facility locations and health district boundaries. Facilities are a much higher priority, but a global dataset could take many years to produce. The main challenges include creating a culture of long-term data maintenance and dissemination in the health community, as well as international standards and nomenclature, and addressing issues of confidentiality and national security. Possible partnerships include survey organizations such as the Measure DHS program and the United Nations Childrens Fund (UNICEF) in addition to WHO, who already have activities in this area. Health districts data are a logical extension of the SALB process, and are likely achievable on a five-year timescale with the appropriate resources. WHO is an obvious champion for both products.

### *Food and Agriculture*

Agricultural production systems data are critical for a wide group of users. While there is already a coordinated effort to address livestock data (see Box 1, page 8), crop data and related information for crop models are in much greater need of improvement. A specific need is better agricultural production systems classifications and maps, including a more detailed analysis of the mixed production systems that are too often lumped together in ways that hinder analysis. Working towards this is a realistic medium term objective, and to that end pilot studies are being undertaken to classify and map agricultural production systems in five countries (Uganda, Senegal, Vietnam, Peru and Andhra Pradesh state of India) to feed into a workshop to explore how best to harmonize and extend these to a global coverage. FAO and CGIAR are involved in this effort.

Additional agricultural data requirements include crop price and disease data. Such information allows analysis of where crop production is occurring and how it is likely to change, for instance, as a result of climate change or price shocks in traded agricultural commodities. AGRIMAP and FAOSTAT are promising initiatives, but the former currently operates on a small budget and FAOSTAT requires significant involvement from national actors, which is necessarily longer term. FAOSTAT should also move towards sub-national data compilation. FAO and the GIS groups in CGIAR centers are in the best position to provide leadership in this area.

### *Land Use and Land Cover*

Global land cover data would be made more useful chiefly with increased temporal and classification resolution. There is an urgent need for a globally consistent time series of land use information. Land use data also require more specific categories. This is especially true for agricultural land use classes that differentiate among individual crops and farming systems, fallow areas, and urban subclasses. These improvements will require more funding for interpretation of remotely sensed data. UNEP, the Committee on Earth Observation Satellites (CEOS), the International Society for Photogrammetry and Remote Sensing (ISPRS), and the Joint Research Centre (JRC) of the European Commission are potential leaders in this area. An interesting difference in priorities exists between global climate and ecosystems modelers, who work with relatively coarse (one or half degree) data, and the agricultural research community and similar applications that have a preference for much higher resolution data.

### *Population and Poverty*

In the short term, population data could be made more flexible if a consortium of providers of underlying "clean" data were able to produce custom derivatives and aggregations. Specific population attributes, such as broad age distributions, would also be valuable, but are only feasible in the longer term. CIESIN could continue to play a leadership role in this area.

In the short term, sub-national estimates of Gross Domestic Product or similar indicators of economic activity (e.g., agricultural output) would be useful. Initial data sets have been compiled by the World Bank, among others, but there is no systematic, longer term effort to put such data sets together.

Small area estimates of poverty are valuable in a broad range of applications. However, the data requirements are extensive (high quality census micro data and household surveys close in time), and differences in data collection conventions and definitions between countries are significant. So global coverage is unlikely in the short term. Cataloguing of these efforts has begun, with involvement from the World Bank, CIESIN, and the World Resources Institute, and their continuing work will focus on compiling more data and determining its comparability and utility in an international context. The use of survey-based proxies of poverty, including hunger and mortality, could continue to be useful, but they also pose challenges of coverage and comparability. In the long run, attention must be paid to the continued modification of survey instruments providing information related to poverty at ever higher spatial resolution.

Because data in this sector need to be compiled from the bottom up—based on national data sets—the realistic goal in the foreseeable future should be compilation of data sets with partial global coverage.

### *Hydrology*

More comprehensive river flow data would be valuable for agricultural and environmental uses. However, there are many distribution restrictions to overcome. One of the largest databases by the Global Runoff Data Centre (GRDC) is only partially accessible to the public. GRDC and the WMO are in a position to take the lead on this.

Information on still water sources and related infrastructure is also valuable. Large scale infrastructure, like reservoirs, aqueducts and dams could be captured relatively easily given a modest investment in time and resources. Earthsat's water bodies layer derived from Geocover Landsat data would be very useful both as a framework data set and for hydrological applications. This will require a lobbying effort to place it in the public domain. Groundwater is trickier, but potentially tractable on a five-year timescale. Small point sources like wells are far more problematic.

***Institutions***

Datasets of regional and environmental institutions and conflict could be valuable to a variety of policy actors. An example is organizations dedicated to regional economic development or river basin management. The University of Oregon, Uppsala University and CIESIN have done work in this area, but it is not as well-developed as many of the other areas. Priorities are therefore less clear. Given the nature of this type of information, a proper model for geographic referencing that is useful for global applications would need to be developed.

***Biodiversity***

Work on data gaps in this sector is ongoing in the context of the Convention on Biological Diversity (CBD), and the discussion here stemmed from that work. The CBD defines three broad categories of data: species, habitat, and protection. Efforts to create data on species extent and richness, especially among endangered species, have generally been successful. Conservation NGOs such as Conservation International and the World Wildlife Fund (WWF) have taken the lead in this. Work on habitat has progressed more slowly, and is likely to require increased coordination. Data on the protection of biodiversity is extensive and coordinated, but important attributes of areas, including the quality of protection, as measured, for example, by budgets or the existence of effective regulations, are often lacking. UNEP-GRID and the UNEP World Conservation Monitoring Center (WCMC) have provided leadership in these areas, but outputs could be much improved.

***Plenary Discussion***

Robert Sandev of the UN noted that ocean and maritime spatial data sets seemed to be completely absent from the group's report. Glen Hyman of CIAT observed that data development efforts are largely *ad hoc*, and that there is a need to find a way to put real resources behind the efforts. The Millennium Ecosystem Assessment might provide one kind of model.

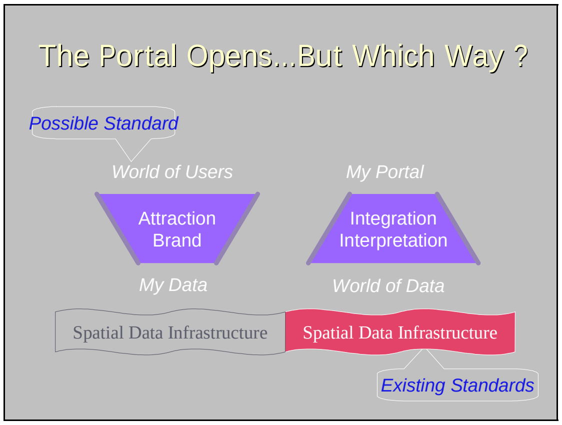
## *Group 3. Data Search, Discovery and Documentation: Role of Portals*

Data portals are dynamic and evolutionary. This means that they can come and go. The ephemeral nature of data portals themselves is a major issue in the sense that an 'abandoned' portal represents resources no longer available to the community. Often data portals fail because sponsors have unrealistic expectations of use, technologies, or costs of maintenance. Organizations need to carefully weigh the pros, cons, and costs of establishing or participating in portal activities. Organizations need to support stability in data archives and continuous data availability.

Data portals are and will continue to be developed primarily with user needs in mind. Sometimes these needs are very specialized. Discussions were largely on the Internet side of the digital divide – that is, in the part of the world that has ready access to the Internet. However, it was recognized by the group that we need to develop ways to make the data and information within geospatial data portals available to those organizations and individuals without reliable Internet connectivity.

The breakout group developed a conceptual 'model' (Figure 7) that differentiates between two data portal designs. On the left side of the figure, the general user community or 'world of users' is funneled to a specialized set of data owned or distributed by the portal custodian.

**Figure 7. Two data portal designs.**




On the right, the user is presented with a specialized portal, tuned to their user community's specific data view and data lexicon, but with a connection to a 'world of data', available through distributed catalogs. The breakout group recommends the 'opening funnel' approach (on the right), since it provides the user with a broad spectrum of data, through the tools and language of their specific discipline field. In order for this model to be successful, metadata and data standards must be applied across a broad spectrum of geospatial, socioeconomic, and environmental data.

### *Background on Data Portals*

Several members of the breakout group provided background materials that help define data portals, in both general usage of the term, and in the specific geospatial community.

In addition, the group refers readers to the following publication *Geospatial Portal Reference Model* (Open Geospatial Consortium, 2004). This document provides high level definitions of portal attributes and functions.

The group's working definition is: "an amalgamation of software applications that consolidate, manage, analyze and distribute information across and outside of an enterprise (including Business, Intelligence, Content Management, Data Warehouse and Mart and Data Management applications)" (Shilakes and Tilman, 1998).

A data portal in the broader sense comprises a few or more of the services as listed in Box 1. Although this definition describes a complete system from human interface (client) to systems storing and serving data and information, the group defined a *data portal* as the thin client that sits on top of other services that provide metadata, data, and information to the client.

### *Recommendations*

The geospatial data community (henceforth the 'community') must work with groups developing interoperability layers for metadata and interoperable services. As described in the group's 'model,' the data portals that serve a specialized user base but still provide discovery and access to a wide range of data provide better value to the user. The community should encourage portals that provide a specialized access framework to diverse data, as opposed to portals that only provide access to specific data sets.

The community must identify and promote incentives that encourage interoperable metadata, data and services, although the breakout group did not identify specifics.

Data portals require metadata about the portal that characterize the uses for and data accessed through the portal. Such metadata will improve the efficiency of the discovery process by the end user.

The community must work to develop a set of 'best practices' criteria for geospatial data portals. The breakout could not identify existing groups or documents that specifically address best practices. Such guidance could improve effectiveness of all geospatial portals, thus providing improved services to the user community. The community needs a recognized thesauri or keyword list. This process could start with groups like the Global Change Master Directory (GCMD) or the FGDC.

Developers of data portals need a mechanism to gather feedback on user satisfaction. Such metrics are required by the portal operators to motivate continued improvement of the portal.

### *Action Items*

The group recommends the following action items:

1. Evaluate GCMD and other candidate thesauri development efforts as potential standards for data access.
2. Encourage OGC/Digital Library communities to develop metadata standards that distinguish content, format, capabilities, potential uses.
3. Identify an organization willing to assemble and publish a best practices manual for data portal operation.
4. Develop a network of 'honest broker' advisors who can help organizations with little technical and information management expertise in assessing their portal needs and strategies.

**Box 3. Characteristics of portals**

Wayne Eckerson, analyst of the Patricia Seybold Group, took a closer look at the Enterprise Information Portal (EIP) marketplace and has developed ten key requirements for an enterprise portal. All quotes are from Wiseth (1999).

1. Easy to Use. 'An enterprise portal must be geared to the skills of the broadest range of users in order to promote self service.' As a consequence the enterprise portal has a graphical interface and uses public browser like consumer portals in the internet.

2. Universal Information Access. An EIP must provide broad access to structured and unstructured information from 'a variety of sources--intranet, internet and extranet.' Portals require comprehensive metadata sources to describe the content in the right context so 'the user can easily find and access it.'

3. Dynamic Resource Access. The user must be able to 'search by category, publish information, subscribe to new content, query and analyse information, and plan and execute activities.'

4. Extensible. The enterprise portal can provide access to all sources, only if it includes a published application programming interface that ``developers can use to hook in existing and future applications."

5. Collaborative. Users should not only be able to publish documents, but also should be able to annotate existing documents and 'create and participate [in] threaded discussions.' When users subscribe to objects, such as reports, spreadsheets and messages, they must have the obligation to 'define the format, delivery channel, and alert method.' Only publishers and administrators should be able to give access rights to objects to users or groups.

6. Customizable. Administrators should have the ability to 'configure different permissions for different' users and groups. Nonetheless users must have the possibility to 'configure settings appropriate to their own needs.'

7. Proactive. 'The enterprise portal can be truly empowering only if it provides an infrastructure for proactive activities.' There must be the ability to 'subscribe to alert mechanisms, create key-performance-indicator monitors, and create agents for automatic searches, or queries' to keep the user informed.

8. Secure. As the portal is a bridge between internal and external interactions it 'should provide security mechanisms to ensure the privacy and integrity of data.' In fact the organization must 'control access at a very granular level--by user, by group, or even by object--and should provide security mechanisms to ensure the privacy and integrity of data.'

9. Scalable. Most enterprises that use the portal technology are very big and are growing every year, consequently the portal must support 'thousands of concurrent requests, hundreds of information sources, and dynamic generation of web pages by thousands of users.' Therefore the architecture behind portals must be very robust and provide capabilities such as 'load balancing across multiple servers, intelligent caching, pooled connections, or other performance-enhancing techniques.'

10. Manageable. 'Simple graphical tools must enable administrators to set rapidly up the user interface, establish permissions, and integrate with other resources.' Monitoring, tuning, and content-management tools should also be part of the portal solution.

*Source: http://www2.iicm.edu/cguetl/education/projects/rscheuch/seminar/node21.html.*

### *Plenary Discussion*

Marc Levy of CIESIN asked the group if they identified any practical first steps for using existing portals. Ron Weaver of NSIDC responded that the group did not get to those specifics. Someone also mentioned that keywords are very important, and that using the GCMD key words might save a lot of effort.

## *Group 4. Technical Data Interoperability*

The Technical Data Interoperability Group was tasked to address technical data interoperability with regards to the production, dissemination, and use of global data sets, and asked to focus primarily on "flexible user access through interoperability and open standards for both cataloged data and Internet mapping." In pursuit of this task, ample consideration was also given to the following: (1) Standardization of spatial data and information such as consistent projections, scales, and boundary files, (2) Identification of data set information that would facilitate appropriate use of the data, and (3) Integration of biophysical and socioeconomic data, including substantive and methodological issues.

This group explored and described standards, conventions, and processes that can promote both technical and substantive data interoperability. Creating mechanisms to facilitate communication and cooperation among data developers, distributors, and users in the global data/research community is one key practice.

### *Discussion*

System interoperability issues extend beyond the network, and include both interface protocols for data retrieval, as well as data standards like projection, datum, and content standards. Documentation needs grow as audiences broaden, and the documentation process becomes an onerous task unless primary focus areas are defined. Global data documentation must meet the needs of at least four primary user types: the researcher, the data technician, the systems analyst, and the information manager. Researchers must be able to determine if the data and the methodology behind the data are applicable for the task at hand. The data technician needs to get the data into a GIS, computer model, or other software program and visually display it. The data must be documented such that reformatting and accurate map projections are possible and verifiable. Global documentation must also provide the information needed by computer systems administrators and architects so that data can be utilized on different platforms. Finally, the documentation must provide information managers with cataloging information so that entries can be found readily by search engines and other access tools.

Standards and conventions are needed and must be flexible. A modular approach for defining standards must be followed, and validation and conformance tools must be developed to insure interoperability. The basic framework for standards has been addressed and implemented by OGC, FGDC, and ISO. These standards meet many data interoperability needs, but not all. At the present time, the FGDC metadata standard is merging with ISO, and many OGC standards are becoming ISO standards. Once the ISO standards are published, a good start would be to implement them and then expand them to address unmet needs. Data content standards need to be defined; for example, we should develop new data after evaluating the FGDC, UNGIWG, or GSDI frame-work data layers and Global Map products as possible models. Currently a universal spatial data model does not exist. The OGC models approach this, but have yet to be implemented, and commercially available products do not conform to interoperability standards because they are most often vendor based.

The scope of the data during the research and collection stage of their development ultimately defines itself spatially, so there will always be differences preventing the definition of a single map projection for global data. Projection standards do not describe all possible coordinate systems. Currently there are no open source software products using projection standards.

Standards alone are not enough to support the interoperability needs of global spatial data.

Tools are needed that implement these standards. Ideally, the determination of global tie points, or known locations for spatially referencing data sets, must be defined. Tools that are developed should not separate the content of the data in presentation form. Portals are a good way to promote shared understanding and to encourage conformance to global standards. The tools and infrastructure that we define must enable automated data discovery and access in order to improve metadata, vocabularies, and data structures.

There are manifold technologies due to adoption time frames. What is old to 'cutting edge' developers may be new to many. The role of GRID computing for large data sets is fast developing, as are new search technologies. The training needs of all stakeholders must be considered to develop working standards. Thesauri need to be implemented to allow inter-community translation. Multi-lingual applications, software, and training technologies need to be part of any interoperability strategy. Mechanisms for sharing experiences and devising standards, and strategies for development and deployment are critical.

***Plenary Discussion***

Steeve Ebener of WHO commented on the need for a global ground reference on which to base any global layer. Greg Yetman of CIESIN responded that it was implicit in framework layer specifications. Jeff Tschirley of FAO observed that data integration remains a major problem.

## Group 5. Science Data Integration

Two frameworks were discussed as starting points towards identifying the issues related to the integration of earth science and social science data. First, the scientific method was used as a tool to identify the points at which data integration issues arise. And second, the thematic areas identified for the workshop highlight specific areas where more research needs to be done and those that are working well. Although discussion covered mostly the scientific method framework, it was felt that the thematic areas could be used to assist in developing a series of case studies highlighting the current best practices in integrating different types of data.

The scientific method has several major steps, broken down as follows for purposes of this discussion:

- Project Design
- Methodology, Experimental Design, Data Availability
- Data Understanding
- Analysis and Feedback Loop/Discovery

The first main step in any project is defining the problem and the user needs. Who funds the project, and the end users of the data impact the definition of the problem. After defining the problem, a literature search is usually conducted to identify previously existing research that may be relevant, including methodology and data. The methodology to be used is then defined, as well as the experimental design. The data that are available are then determined, and evaluated. Once the appropriate data have been obtained, they often require extensive cleaning. In other words, the data need to be entered, checked, conditioned, and transformed, with multiple data sets integrated into a useful whole. During this entire process, a feedback loop is in place, causing adjustments to be made to the methodology and the data used. After the experiment is completed, the results must then be interpreted and validated; then disseminated and reviewed.

The key points during this process outlined above in which data integration is integral are the data availability and experimental design phase, an evaluation of the data, and the feedback loop/discovery phase.

During the experiment design phase, the data appropriate to the project, if even available, have been identified. Most times, the data are not ideal. There may be missing data values, variable data collection methods, and data collated from different sources. Such idiosyncrasies must be accommodated. How are missing data points or values handled? How have the methods to accommodate the missing data changed over time? Data are often aggregated or disaggregated depending upon the purpose for which they will be used. Different methods currently exist for aggregation or disaggregation for different types of data (e.g., socioeconomic data collected by census district or household level versus environmental data collected for specific points or at the watershed level). The spatial extent, or the study area, has an impact. Data collected at the global level are occasionally used for regional or local level studies. Although these data may not be appropriate, researchers may have no other data options available to them.

The unit of analysis (the resolution at which the project or problem is being looked at) does not always equal the unit of measurement (the resolution at which the data were collected). This difference impacts the design of experiments – the ideal data resolution may simply not be available, and therefore aggregated or disaggregated data may have to be used. This is not just a spatial issue, but also a temporal one. Also, what is being investigated? Flows, pools, or fluxes?

An additional problem identified is that of data understanding. Better understanding or easier to use "meta-search" engines for spatial data would allow easier discovery of appropriate data sets. Metadata, while cursorily used, still do not meet the needs of researchers – what is the data set's pedigree? How were the data first collected? What transformations were performed? Metadata often do not have enough information on the quality of the data. And what error effects have been propagated via transforming the data, i.e., from one projection to another, or by changing the units of measurement?

During the discovery phase, or analysis and feedback phase, the experiment and methodology are 'tweaked' based upon the knowledge gained during the previous stages of the project. Questions raised here:

- How will the appropriate unit of measurement be determined – what is the scale at which the processes of interest are operating?
- How can changes in the unit of reporting (i.e., change in boundaries) over time be incorporated?
- How can error values be placed on the results?

Suggested recommendations include:

- Review work to date in the area of science and social science data integration for global-scale data
- Develop a detailed conceptual framework that describes the process of integrating global spatial data for scientific analysis
- Use the conceptual framework to develop a set of case studies on science data integration

***Plenary Discussion***

Dennis Ojima of Colorado State University suggested working closely with the data gaps group to produce better time series data. Ric Cicone of Isciences observed that there are often problems in working with data sets that are outside your area of expertise. He suggested that enhanced metadata might help users to know whether it is possible to resample the data or perform other operations.

# References

National Research Council (NRC). 2004. *Licensing Geographic Data and Services*. Washington, DC: National Academies Press. Available at http://books.nap.edu/catalog/11079.html

Open Geospatial Consortium (OGC). 2004. *Geospatial Portal Reference Model*. OpenGIS Discussion Paper OGC 04-039, Version: 0.2, edited by Louis C. Rose. Available at http://www.opengeospatial.org/

Sanchez, Pedro, M.S. Swaminathan, Philip Dobie, and Nalan Yuksel. 2005. *Halving Hunger: It Can be Done*. Sterling, VA: Earthscan and Millennium Project.

Shilakes, Christopher C., & Julie Tylman. 1998.. "Enterprise information portals: Move Over Yahoo!, the Enterprise Information Portal is on its Way." In-depth Report, November 16, 1998. Meririll Lynch, New York. Available at http://emarkets.grm.hia.no/gem/Topic7/eip_ind.pdf

Uhlir, Paul. 2003. "Discussion Framework," in: *The Role of Scientific and Technical Data and Information in the Public Domain*. Julie M. Esanu and Paul F. Uhlir, eds., Washington, DC: National Academies Press, pp. 3-9 and 119-124.

Wiseth, Kelli. 1999. "Portal Power," *Oracle Magazine*, November 1999.

# Annex 1. Workshop Agenda

*Note: Most of the presentations listed on the agenda are available for download from the workshop website: http://sedac.ciesin.columbia.edu/GSDworkshop/.*

| **Day 1 – Tuesday, 21 September**<br>*Identification of User Needs* | |
|---|---|
| 0900-0915 | **Welcome** by Roberta Balstad Miller, CIESIN |
| 0915-0945 | **Setting the Stage:** Principal Goals and Objectives of the Workshop<br>Robert Chen (CIESIN), Jeffrey Tschirley (FAO) |
| 0945-1045 | **Opening Plenary Panel #1:** Access to Global Spatial Data and Information from the User's Perspective<br>*During the last decade, technological advancements in remote sensing, geographic information systems (GIS), and other geospatial data management tools in tandem with development of the Internet have contributed to the creation of many global and regional spatial databases, online mapping services and geospatial data clearinghouses. However, the dramatic expansion of this wealth of geographic data and information has not necessarily been matched by increases in the ability of users to take full advantage of the available resources, or to make sense of similar products. This plenary panel sets the stage for the workshop by addressing user needs for data access and documentation.*<br>*Chair:* Stanley Wood (IFPRI)<br>*Panelists:*<br>1. Dennis Ojima (Colorado State University and Land Project), "Global Environmental Change Science"<br>2. Robert Ford (Loma Linda University), "Who Needs Spatial Data Infrastructure?"<br>3. Jillian Thonell (Millennium Ecosystem Assessment), "Global Data Challenges from an MA Perspective"<br>4. Craig Beech (Regional Peace Parks Foundation), "Peace Parks Foundation Requirements for Global Data" |
| 1045-1100 | *Coffee Break* |
| 1100-1230 | **Presentations:** Global Spatial Data and Information Dissemination Initiatives<br>*Short presentations on selected major data development and dissemination initiatives for main topical areas.*<br>*Chair:* Roberta Miller (CIESIN)<br>*Agriculture and Food Security:* Robert Zomer (IWMI) and Tim |

| | |
|---|---|
| | Robinson (FAO)<br>*Environment:* Ashbindu Singh (UNEP)<br>*Health:* Steeve Ebener (WHO)<br>*Population and Poverty:* Deborah Balk (CIESIN) |
| 1230-1400 | **Lunch:** *Optional presentations on various data development and dissemination initiatives are scheduled from 1315-1400 in Monell Auditorium. Please see separate lunchtime presentation program for titles and abstracts.* |
| 1400-1500 | **Open Discussion:** Selection of breakout topics and discussion on targeted outcomes of the workshop. Proposed breakout topics are:<br>*Chair:* Robert Chen (CIESIN)<br>*1. Principles for data sharing and access.* Chair: Paul Uhlir (NRC)<br>*2. Gaps in global data and collaboration in data development*. Chair: Uwe Deichmann (World Bank)<br>*3. Data search, discovery, and documentation; role of portals*. Chair: Ron Weaver (NSIDC)<br>*4. Technical data interoperability*. Chair: William Anderson (Praxis101)<br>*5. Science data integration.* Chair: Glenn Hyman (CIAT) |
| 1500-1700 | **Breakout Session 1.** Identification of major issues for each topic. (coffee break at 1600) |
| 1700-1730 | **Report to plenary** on issues identified in breakout session |
| 1745-1800 | *Transportation to dinner* |
| 1800 | *Dinner at the IBM Conference Center*<br>Speaker: Harlan Onsrud (University of Maine and GSDI) |
| 2000 | *Transportation to hotels* |
| **Day 2 – Wednesday, 22 September**<br>*Global Data Sets – Gaps, Standardization and Integration Issues* | |
| 0830-0930 | **Presentation** of test bed activities in Africa – France Lamy (FAO) and Paola de Salvo (WFP) |
| 0930-1045 | **Plenary Panel #2**: Inventories of Global Spatial Data, Standardization, Search, Portal Development<br>*The presentation on the first day highlighted many activities related to the development and dissemination of global data sets. This plenary will discuss core data sets for sustainable development research and operational* |

| | |
|---|---|
| | *programs, efforts to develop standards, and the development of portals and online mapping tools for data access.*<br>*Chair:* Steeve Ebener (WHO)<br>*Panelists:*<br>1. Joseph Dooley (Spatial Data Services & Mapping), "Core Global or Geo-Spatial Databases (CGDB)"<br>2. Allan Doyle (EOGEO), "Closing the Interoperability Gap"<br>3. Richard Cicone (ISciences, LLC), "Geospatial Data Fusion: Integration Challenges"<br>4. Chris Chiesa (Pacific Disaster Center), "The Asia Pacific Natural Hazards Information Network (APNHIN)"<br>5. Kees Klein Goldewijk (RIVM), "Global Spatial Data Sets and User Needs at RIVM" |
| 1045-1100 | *Coffee Break* |
| 1100-1230 | **Breakout Session 2.** Groups based on topics identified on day 1. Identification of challenges and opportunities. Recommendations for short- and medium-term actions and initiatives, including identification of possible sources of support, collaborative proposals. |
| 1230-1400 | **Lunch**<br>*Optional presentations on various data development and dissemination initiatives are scheduled from 1330-1400 in Monell Auditorium. Please see separate lunchtime presentation program for titles and abstracts.* |
| 1400-1530 | **Breakout Session 2** (continued). Focus on fleshing out candidate agreements; plans for new datasets, testbeds and portals; proposals to adopt or develop particular standards, etc. |
| 1530-1545 | *Coffee Break* |
| 1545-1700 | **Roundtable Discussion:** Potential Roles for Global Data Coordination Mechanisms<br>*Chair:* Alex de Sherbinin (CIESIN)<br>*1. UN Geographic Information Working Group* – Steeve Ebener<br>*2. Digital Earth and ISPRS* – Hiromichi Fukui<br>*3. Global Spatial Data Initiative* – Harlan Onsrud<br>*4. Spatial Information Management Advisory Group* – Jeffrey Tschirley<br>*5. International Steering Committee for Global Map* – Karen Kline<br>*6. CODATA* – Robert Chen |

| | |
|---|---|
| | *7. Open Geospatial Consortium* – George Percivall |
| 1710 | *Transportation to hotels, Morningside* |
| **Day 3 – Thursday, 23 September**<br>*Global Data Sets – The Way Forward* | |
| 0830-0930 | **Report out from Day 2 Breakout Sessions:** Identification of specific action items and initiatives. |
| 0930-1045 | **Plenary Panel #3**: Data Documentation, Quality, Preservation, and Intellectual Property Rights<br>*Institutional efforts related to data development, management, and dissemination need to be carried out in a coordinated way, or the potential for poor interoperability between information systems, incompatibilities between key datasets, confusing or even conflicting data documentation, and complex constraints on data re-dissemination and use is likely to increase. This plenary will address these issues.*<br>*Chair:* Robert Sandev (UN)<br>*Panelists:*<br>1. Jean-Louis Ecochard (The Nature Conservancy), "The Conservation Commons"<br>2. James Edwards (GBIF), "The Global Biodiversity Information Facility: Data Documentation, Quality, Preservation and IPR"<br>3. Paul Uhlir (National Research Council), "Creating a Global Information Commons for Public GSDI: Legal and Economic Policy Aspects" |
| 1045-1100 | *Coffee Break* |
| 1100-1220 | **Concluding Discussion**: Major outcomes and the way forward. Open discussion on the workshop statement.<br>*Chair:* Robert Chen (CIESIN) |
| 1220-1230 | **Close of Meeting** |
| 1230 | *Transportation to Manhattan, airports* |
| 1230-1400 | **Lunchtime bilateral meetings** |

# Annex 2. Breakout Group Guidelines

Each breakout group has a total of five hours of meeting time over days one and two of the workshop (not including the initial plenary discussion on the breakout group themes). At the end of day one, each group will have five minutes in plenary to describe *a)* how they have framed the main issues that they will address and *b)* what progress they have made in their initial discussions. The final report will be on day three, when each group will have ten minutes to summarize their discussions and describe the way forward.

Each breakout group will have a rapporteur who will record minutes of the discussions on a laptop and help the chair to pull together a short PowerPoint presentation and a summary report from the group.

***1. Principles for data sharing and access.***

*Participants*

Paul Uhlir, US National Research Council, USNC for CODATA (chair)
Meredith Golden, CIESIN, Columbia University (rapporteur)
James Edwards, Global Biodiversity Information Facility (GBIF)
Norman Kerle, International Institute for Geo-Information Science and Earth Observation (ITC)
France Lamy, UN Food and Agriculture Organization, SDRN
Adriano Miele, UNESCO Institute of Statistics
Harlan Onsrud, University of Maine, GSDI
Jeff Tschirley, UN Food and Agriculture Organization (FAO)
Gordon Young, United Nations World Water Assessment Programme

*Goals*

Elaboration of principles for data sharing, attribution, and dissemination, including possible commitment to dissemination of data sets as global public goods. Agreement on the role and mandate of the recently initiated Spatial Information Management Advisory Group (SIMAG).

*Discussion Questions*

a) From a user perspective, what are key needs for: access to global-scale data; rights to use and redisseminate data and derived products; and clarity in providing attribution to data sources and owners?

b) From a data source perspective, what aspects of a global dataset's intellectual property need to be protected and what are the key needs for attribution, documentation, and reporting on data usage?

c) From a data distributor's perspective, what rights and protections are needed to support data distribution and access? What are key needs for documentation of these rights and protections on the part of both data sources and data users?

d) What kind of bi-lateral or multi-lateral agreements might be developed among major data developers and distributors that would obviate the need for case-by-case arrangements for data sharing? Might these agreements cover alpha versions or data still under development?

e) What are the pros and cons of making data available in the public domain? Are there alternative approaches that might provide suitable intellectual property protections to data sources and/or owners yet promote widespread data use and application?

f) Is it possible to establish clearer and more consistent definitions and practices regarding use and dissemination rights (e.g., with respect to data redissemination; commercial vs. educational vs. scientific use; attribution of data sources, integrators, owners, and disseminators; and use of derived imagery and statistics vs. raw data)?

g) What specific roles can international coordination mechanisms such as CODATA, SIMAG, GSDI or UNGIWG play in promoting data sharing and access?

## 2. *Identification of gaps in global data and collaboration in data development.*

*Participants*

Uwe Deichmann, World Bank (chair)
Adam Storeygard, CIESIN (rapporteur)
Deborah Balk, CIESIN
Doug Beard, United States Geological Survey
Brian Blankespoor, Oregon State University
Lorant Czaran, United Nations Cartographic Section
Joe Dooley, SDS Mapping
Steeve Ebener, World Health Organization
Hiromichi Fukui, Digital Earth and ISPRS
Yuri Gorokhovich, CIESIN
Chris Ogolo Ikporukpo, University of Ibadan
Kees Klein Goldewijk, RIVM
Marc Levy, Columbia University
Daniel Prager, World Resources Institute
Joerg Priess, Unversity of Kassel
Tim Robinson, United Nations Food and Agriculture Organization
Jillian Thonell, Millennium Ecosystem Assessment

*Goals*

Stock-taking of global data sets and identification of significant gaps and overlaps in the following thematic areas: environment; food and agriculture; population, poverty and health. Collaboration in the development of new global-scale spatial data sets.

*Discussion Questions*

a) What are major scientific gaps and weaknesses (e.g., spatial, temporal, qualitative, topical) in global-scale data for the four major thematic areas identified by this workshop: environment; food and agriculture; population, poverty and health?

b) Are there important gaps in cross-disciplinary, integrated datasets and indicators that are needed for policy, decision making, education, or other applications?

c) For known problems such as poor data coverage and quality in many developing countries, are there ongoing initiatives or new opportunities that could be harnessed to address global-scale data needs?

d) How can the global data community reduce duplication of effort in data creation, improvement, updating, and dissemination?

e) What new technologies and standards should be considered to facilitate development of higher quality, better integrated global-scale datasets?

f) What kind of collaborative efforts might be undertaken among major data developers, and how might funding be procured for implementation?

g) Could an initiative be developed for working with specific countries to intensify their data development efforts in these domains which might serve as a role model to other countries?

### 3. *Data search, discovery, and documentation; role of portals.*

Participants

Ron Weaver, NISDC (chair)
Chris Lenhardt, CIESIN (rapporteur)
Lars Bromley, American Association for the Advancement of Science
Chris Chiesa, Pacific Disaster Center
Paola DeSalvo, World Food Programme
Alex de Sherbinin, CIESIN
Robert Ford, Loma Linda University
Jeffrey Henigson, UN Office for the Coordination of Humanitarian Affairs
Dennis King, State Department
Joshua Lieberman, Geospatial Architect
Dan Phillips, US Geological Survey
Robert Zomer, International Water Management Institute

Goals

Identification of user needs for online services, and for education and capacity-building in how to use those services. Collaboration in the development of new web-mapping portals, search tools, and documentation resources.

Discussion Questions

a) Who are the most important users of global data, and what types of assistance in finding, accessing, and using data do they need?

b) How can flexible user access be promoted through interoperability and open standards of both catalog searching and Internet mapping? How can such access be tailored to the specific needs of different user communities (e.g., users with different topical interests; research vs. educational vs. applied users, users from different countries or linguistic backgrounds, users with different levels of sophistication and access to tools)?

c) What additional information on data sets (beyond discovery metadata) would facilitate their appropriate use (e.g., data set guides, tutorials, citation lists)?

d) Given the expected proliferation of sources of similar global-scale data that may be appropriate for different uses, what mechanisms could be established to promote the most appropriate use of data by different user types? For example, how can the global data community support both *a)* the needs of scientists and experts for detailed data that reflect current uncertainties and *b)* the needs of general users for generalized data and information?

e) Would it make sense to develop a global data web ring similar to the sustainability web ring?

f) Would it be useful to coordinate efforts in portal and web mapping development in order to reduce duplication in the services being provided?

g) Could joint initiatives/joint funding be developed to ensure that web portals are up-to-date and well funded (rather than set up only to languish from lack of attention and resources)?

### 4. *Technical data interoperability.*

<u>*Participants*</u>

William Anderson, Praxis101 (chair)
Greg Yetman, CIESIN (rapporteur)
Tammy Beaty, Oak Ridge National laboratory
Benno Blumenthal, International Research Institute for Climate Prediction
John Del Corral, International Research Institute for Climate Prediction
Allan Doyle, EOGEO

<u>*Goals*</u>

Identification of standards and conventions for storing, disseminating, and visualizing spatial data (e.g., consistent projections, scales, boundary files, etc.). Agreements on standards and conventions to promote technical data interoperability.

<u>*Discussion Questions*</u>

a. What are the short-term and long-term needs of the global data community for standards and conventions to promote data sharing and seamless user access?

b. Do current standards for data interoperability being developed and implemented by organizations such as the Open Geospatial Consortium (OGC), the Federal Geographic Data Committee (FGDC), and the International Standards Organization (ISO) meet these needs?

c. What additional standards or conventions are needed (e.g., agreement on particular datums, framework datasets, projections, terminology, thesauri)?

d. Given that users will have different thematic interests in global data (e.g., users interested in water vs. agriculture vs. energy; users interested in short-term climate variability vs. long-term climate change), are there ways to promote flexibility in developing customized user interfaces without excessive duplication of underlying common data?

e. What improvements are needed in metadata, vocabularies, and/or data structures to promote cross-disciplinary data discovery, access, and integration?

f. Are there new technologies in the near future on the horizon that could facilitate technical data interoperability if implemented in a coordinated way?

g. Are there needs for training, access to tools and documentation, and sharing of expertise, especially with regard to developing countries, that could be met through collaborative activities?

### 5. Science data integration.

*Participants*

Glenn Hyman, CIAT (chair)
Karen Kline, University of California at Santa Barbara (rapporteur)
Ric Cicone, ISciences LLC
Christopher Doll, CIESIN
Ellen Douglas, University of New Hampshire
Ron Janssen, Vrije Universiteit Amsterdam
Dennis Ojima, Colorado State University
John Rumble, CODATA
Noam Unger, Office of Humanitarian Affairs, Department of State
Stanley Wood, IFPRI

*Goals*

Discussion on the integration of biophysical and socioeconomic data, including substantive and methodological issues. Agreements on standards and conventions to promote substantive data interoperability.

*Discussion Questions*

a) What are the needs of different types of users for integrated global datasets, e.g., with respect to resolution, quality, time averaging/resolution, presentation of uncertainty?

b) What are key pitfalls of haphazard data integration (e.g., using inconsistent values for land area or population in density calculations; inconsistent boundary datasets for mapping and analysis; mixing of low and high resolution data; mixing of environmental and socioeconomic data)? What steps could be taken to reduce the risk of such pitfalls?

c) For specific applications, what are examples of appropriate and inappropriate cross-disciplinary global data usage? (Examples: use of GPW vs. LandScan, differences among land cover datasets, limitations of various digital elevation datasets, widespread use of DCW/VMAP) What steps could be taken to encourage appropriate use and reduce inappropriate uses?

d) For ongoing or future efforts to develop and improve global-scale datasets, what steps could be taken at an early stage to improve the cross-disciplinary quality and use of such data (e.g., retention and quality control of georeferencing data and parameters; agreement on common framework datasets; securing of particular data rights)?

e) Are there new methods for analyzing, visualizing, and summarizing data uncertainty and accuracy that should be applied to key global datasets?

f) What emerging issues need to be addressed regarding data integration in new topical areas, for new applications, or at higher spatial and temporal resolution (e.g., confidentiality, national security, intellectual property rights, liability)?

g) Are there opportunities to promote training and sharing of expertise and experience related to science data integration?

# Annex 3. Participant List

| Name | Department or Division | Organization | Email |
|---|---|---|---|
| William Anderson | | Praxis101 | band@acm.org |
| Deborah Balk | CIESIN | Columbia University | dbalk@ciesin.columbia.edu |
| Roberta Balstad | CIESIN | Columbia University | roberta@ciesin.columbia.edu |
| Doug Beard | National Biological Information Infrastructure | US Geological Survey | dbeard@usgs.gov |
| Tammy Beaty | | Oak Ridge National Laboratory | beatytw@ornl.gov |
| Craig Beech | | Regional Peace Parks Foundation (PPF) | cbeech@ppf.org.za |
| Brian Blankespoor | Universities Partnership for Trans-boundary Waters | Oregon State University | bblank11@yahoo.com |
| Benno Blumenthal | IRI | Columbia University | benno@iri.columbia.edu |
| Helene Bray | Department of Peacekeeping Operations | United Nations | bray@un.org |
| Lars Bromley | | American Association for the Advancement of Science (AAAS) | lbromley@aaas.org |
| Pietro Ceccato | IRI | Columbia University | pceccato@iri.columbia.edu |
| Robert Chen | CIESIN | Columbia University | bchen@ciesin.columbia.edu |
| Chris Chiesa | | Pacific Disaster Center | cchiesa@pdc.org |
| Ric Cicone | | ISciences, LLC | rcicone@isciences.com |
| Lóránt Czárán | | United Nations Geographic Information Working Group | czaran@un.org |
| Howard Dale | | UNICEF | hdale@unicef.org |
| Paul Davis | Global Land Cover Facility | University of Maryland | pdavis@umiacs.umd.edu |
| Paola de Salvo | GeoSpatial Analyst VAM | UN World Food Programme | Paola.DeSalvo@wfp.org |
| Alex de Sherbinin | CIESIN | Columbia University | adesherbinin@ciesin.columbia.edu |
| Uwe Deichman | | World Bank | UDeichmann@worldbank.org |

| Name | Department or Division | Organization | Email |
|---|---|---|---|
| John Del Corral | IRI | Columbia University | jdcorral@iri.columbia.edu |
| Christopher Doll | CIESIN | Columbia University | cdoll@ciesin.columbia.edu |
| Joe Dooley | | Consultant | dooley@sdsmap.com |
| Ellen Douglas | Institute for the Study of Earth, Oceans and Space | University of New Hampshire | ellen.douglas@unh.edu |
| Allan Doyle | | EOGEO | adoyle@intl-interfaces.com |
| Steeve Ebener | Evidence and Information for Policy | UN World Health Organization | ebeners@who.int |
| James Edwards | | Global Biodiversity Information Facility | jedwards@gbif.org |
| Robert Ford | School of Science and Technology | Loma Linda University | rford@univ.llu.edu |
| Jeanne Foust | | Environmental Systems Research Institute | jfoust@esri.com |
| Hiromichi Fukui | | Digital Earth | hfukui@sfc.keio.ac.jp |
| Chandra Giri | | US Geological Survey EROS Data Center | cgiri@usgs.gov |
| Kees Klein Goldewijk | | The Netherlands National Institute of Public Health and the Environment (RIVM) | kees.klein.goldewijk@rivm.nl |
| Leah Goldfarb | | International Council for Science (ICSU) | leah@icsu.org |
| Anne Griggs | | Centers for Disease Control | agriggs@cdc.gov |
| Jeffrey Henigson | ReliefWeb | UN Office for the Coordination of Humani-tarian Affairs | henigson@un.org |
| Leen Hordijk | | International Institute for Applied Systems Analysis | hordijk@iiasa.ac.at |
| Glenn Hyman | | CGIAR-CIAT | g.hyman@cgiar.org |
| Chris Ogolo Ikporukpo | Department of Geography | University of Ibadan | coikporukpo@yahoo.com |
| Shuichi Iwata | | University of Tokyo | iwata@k.u-tokyo.ac.jp |
| Ron Janssen | Institute for Environmental Studies | Vrije Universiteit, Amsterdam | ron.janssen@ivm.vu.nl |
| Norman Kerle | | ITC - International Institute for Geo-Information Science and Earth Observation | kerle@itc.nl |
| Dennis King | Humanitarian Information Unit | State Department | kingd@state.gov |

| Name | Department or Division | Organization | Email |
| --- | --- | --- | --- |
| Karen Kline | | International Steering Committee for Global Mapping | kline@geog.ucsb.edu |
| Raffaela Kozar | Center for Tropical Agriculture | Columbia University | kozar@iri.columbia.edu |
| France Lamy | SDRN | UN Food and Agriculture Organization | France.Lamy@fao.org |
| Kerstin Lehnert | LDEO | Columbia University | lehnert@ldeo.columbia.edu |
| Christopher Lenhardt | CIESIN | Columbia University | clenhardt@ciesin.columbia.edu |
| Art Lerner-Lam | LDEO | Columbia University | lerner@ldeo.columbia.edu |
| Marc Levy | CIESIN | Columbia University | mlevy@ciesin.columbia.edu |
| Joshua Lieberman | | Geospatial Architect | josh@oklieb.net |
| Liu Chuang | GCIRC | Chinese Academy of Sciences | lchuang@igsnrr.ac.cn |
| Adriano Miele | | UNESCO Institute of Statistics | a.miele@uis.unesco.org |
| Ekkehard Mochmann | Zentralarchiv für Empirische Sozialforschung | University of Cologne | mochmann@za.uni-koeln.de |
| Thomas Moritz | | American Museum of Natural History | tmoritz@amnh.org |
| Dennis Ojima | | Colorado State University | dennis@nrel.colostate.edu |
| Harlan Onsrud | GSDI | University of Maine | onsrud@spatial.maine.edu |
| Juna Goda Papajorgji | | Urban and Regional Information Systems Association (URISA) | jpapajor@co.alachua.fl.us |
| Dan Phillips | National Biological Information Infrastructure | US Geological Survey | dphillips@usgs.gov |
| Francesca Pozzi | CIESIN | Columbia University | fpozzi@ciesin.columbia.edu |
| Daniel Prager | EarthTrends | World Resources Institute | dprager@wri.org |
| Jörg Priess | Center for Environmental Systems Research | University of Kassel | priess@usf.uni-kassel.de |
| Tim Robinson | AGAL | UN Food and Agriculture Organization | Tim.Robinson@fao.org |
| Chet Ropelewski | IRI | Columbia University | chet@iri.columbia.edu |
| John Rumble | | CODATA | jumbleusa@earthlink.net |
| Robert Sandev | Office of Legal Affairs | United Nations | sandev@un.org |

| Name | Department or Division | Organization | Email |
|---|---|---|---|
| David Schell | | Open Geospatial Consortium | dschell@opengeospatial.org |
| Ashbindu Singh | | United Nations Environment Program | Ashbindu.Singh@rona.unep.org |
| Adam Storeygard | CIESIN | Columbia University | astoreygard@ciesin.columbia.edu |
| Jillian Thonell | | Millennium Ecosystem Assessment | thonell@millenniumassessment.org |
| Jeff Tschirley | Sustainable Development Dept. | UN Food and Agriculture Organization | Jeff.Tschirley@fao.org |
| Paul Uhlir | USNC for CODATA | US National Research Council | PUhlir@nas.edu |
| Noam Unger | Humanitarian Information Unit | US Department of State | UngerNC@state.gov |
| Ron Weaver | | National Snow and Ice Data Center | weaverr@nsidc.org |
| Stan Wood | | International Food Policy Research Institute (IFPRI) | S.Wood@cgiar.org |
| Greg Yetman | CIESIN | Columbia University | gyetman@ciesin.columbia.edu |
| Gordon Young | Global Water Assessment Programme | UNESCO | g.young@unesco.org |
| Robert Zomer | | International Water Management Institute (IWMI) | r.zomer@cgiar.org |