Sharing Environmental Data through GEOSS

Giuliani, Gregory; Ray, Nicolas; Schwarzer, Stefan; De Bono, Andréa; Peduzzi, Pascal; Dao, Quoc-Hy; Van Woerden, Jaap; Witt, Ron; Beniston, Martin; Lehmann, Anthony

How to cite

GIULIANI, Gregory et al. Sharing Environmental Data through GEOSS. In: International journal of applied geospatial research, 2011, vol. 2, n° 1, p. 1–17. doi: 10.4018/jagr.2011010101

This publication URL: https://archive-ouverte.unige.ch//unige:18359

Publication DOI: 10.4018/jagr.2011010101

Sharing Environmental Data through GEOSS

Gregory Giuliani, University of Geneva and UNEP, Switzerland
Nicolas Ray, University of Geneva and UNEP, Switzerland
Stefan Schwarzer, UNEP, Switzerland
Andrea De Bono, University of Geneva and UNEP, Switzerland
Pascal Peduzzi, UNEP, Switzerland
Hy Dao, University of Geneva and UNEP, Switzerland
Jaap Van Woerden, UNEP, Switzerland
Ron Witt, UNEP, Switzerland
Martin Beniston, University of Geneva, Switzerland
Anthony Lehmann, University of Geneva and UNEP, Switzerland

ABSTRACT

Understanding the complexity of earth-system processes is crucial to convey improved information on the environment to decision-makers and the general public. Addressing this need by sharing environmental data is challenging because it requires a common agreed framework that allows easy and seamless integration of data from different sources. In this regard, the Global Earth Observation System of Systems (GEOSS) portends major benefits through various sharing mechanisms and by giving access to services that could be linked together to process and generate new understandable knowledge and information. Various United Nations projects could greatly benefit from the GEOSS approach.

Keywords: *Capacity Building, Data Sharing, GEOSS, Grid Computing, Interoperability, SDI, Service Chaining*

INTRODUCTION

Today we are living in a globalized world with rapidly evolving processes including climate change, population growth or environmental degradation. In parallel, means of communication have expanded to take on a remarkable place in our society, allowing us to access an enormous and continuous flow of information.

In the last 30 years, the availability of geospatial data has grown dramatically following

DOI: 10.4018/jagr.2011010101

the evolution of communication technologies supported by the rapid development of spatial data capture means such as remote sensing imagery, sensors and GPS (Philips, Williamson, & Ezigbalike, 1999). One of the challenges we are facing today is to make sense of this vast amount of data in order to turn them into understandable knowledge (Gore, 1998). Concrete actions can be taken only on the basis of knowledge and understanding, but often we know too little about the state of our planet's environment to take informed and sound decisions about how it should be managed.

Our planet is a multi-dimensional system made of complex interactions highly interconnected and continuously evolving at many spatial and temporal scales (GEO secretariat, 2007b). This means that to understand these interactions, we need to gather and integrate different sets of data about physical, chemical and biological systems. Altogether, these sets of data constitute environmental data sets or data related to the environment. These data are often georeferenced, describing a geographical location through a set of attributes and thus could be understood as being part of geospatial data. An environmental data set is seldom interesting in itself, but rather displays its full information potential when used in conjunction with other data sets, allowing one to monitor and assess the actual status of the global, regional or local environments, to discover complex relationships between them and to model future changes.

In 1998, the former vice-president of the United States, Al Gore, presented his visionary concept of a Digital Earth (Gore, 1998), a representation of the Earth embedding a vast amount of geospatial data and allowing to make better sense of it. To achieve this vision, Gore highlighted the need for a collaborative effort (from government, industry, academia and citizens) and pointed out the different technologies required: computational power, mass storage, satellite imagery, broadband network, interoperability and metadata.

Despite the fact that administrations and governments are recognizing that geospatial data are an important component of an information infrastructure (such as e-governement) that needs to be efficiently coordinated and managed for the interest of all citizens (Ryttersgaard, 2001), this huge amount of geospatial data is stored in different places, by different organizations and the vast majority of these data are not being used as effectively as they should. In consequence, a framework allowing one to discover, access, publish, share, maintain and integrate geospatial data appears to be essential. Such a framework is commonly known as a Spatial Data Infrastructure (SDI).

Different initiatives at the regional and global levels are influencing and promoting the creation of SDIs allowing data providers to share and publish their data in an interoperable manner. These initiatives coordinate actions that promote awareness and implementation of complementary policies, common standards and effective mechanisms for the development and availability of interoperable geospatial data and technologies to support decision making at all scales for multiple purposes. These initiatives are related to data access, harmonization, standardization, interoperability, seamless integration and services. Such an initiative is the Global Earth Observation System of Systems (GEOSS) which is a worldwide voluntary effort, coordinated by the Group on Earth Observation secretariat, to connect already existing SDIs and Earth Observation infrastructures. GEOSS is foreseen to act as a gateway between producers of geospatial data and end users, with the aim of enhancing the relevance of Earth observations for the global issues and offering public access to comprehensive information and analyses on the environment (GEO secretariat, 2005, 2007a). The GEOSS Common Infrastructure (GCI) provides core capabilities that allow users to search, access and use data, information, tools and services, and is made of five components: GEO portal (web portal to access GEOSS and search registries), GEOSS clearinghouse (connects the different components), GEOSS components and services registry (catalogue of services and components), GEOSS standards and interoperability registry (catalogue of standards to use allowing users to set up and configure an interoperable system), and a best practices wiki (offers a single space to share, discuss, propose and exchange ideas and best practices within the community). These components are dependent on the voluntary contributions of members and participating organizations. To support the nine defined Societal Benefit Areas (SBAs) (disasters, health, energy, climate, water, weather, ecosystems, agriculture, biodiversity), the mechanisms for data sharing and dissemination are presented in a 10-year Implementation Plan Reference Document (GEO secretariat, 2005) providing data sharing principles that any volunteer member must endorse. The key element o share data through GEOSS is to agree on "interoperability arrangements" (GEO secretariat, 2007a) allowing different components of the system to communicate with each other.

Turning data into understandable knowledge requires that data coming from different sources be easily and seamlessly integrated. With the capabilities offered by standards like the one proposed by the Open Geospatial Consortium (OGC), geospatial community can not only discover, access and publish interoperable geospatial data but also services that can be linked together, in chains of services, to process data and generate new information. Moreover, by registering services into GEOSS, these different resources are now accessible in a standardized way and are reusable for many different purposes.

The aim of this paper is to present experiences gathered through different United Nations (UN) and European research projects and to discuss promises and challenges envisioned in participating to an initiative like GEOSS, both in term of building chains of services and sharing data.

THE NEED FOR DATA SHARING AND INTEGRATION

Until very recently, the different systems used to acquire environmental data were mostly operating in isolation, which made it difficult to easily discover, access and use the data content of these systems due to incompatibilities and inconsistencies of formats and data models (Bernard & Craglia, 2005). In addition, there is typically insufficient data exchange among different stakeholders, which is partially due to differing data policies. Other important impediments to the flow of data are the delays in accessing data that prevent timely use of information, duplication and redundancy of data acquisition, potential high costs associated with data creation and access, and unclear access rights and licensing policies (GEO secretariat, 2005). Altogether, these difficulties lead to a fragmentation of data sources, impeding their effective and efficient use, requiring much more time than necessary for data collection (Open Geospatial Consortium, 2004). All the previous considerations highlight the growing need to share data in an interoperable way and to ensure that data are easily accessible and discoverable, so that they can be used as often and widely as possible (Arzberger et al., 2004). Moreover, the adoption of the Agenda 21 resolution, a United Nations initiative proposing a set of actions to be taken at different scales to promote a sustainable development, fostered the importance of geospatial data to support decision-making and management related to degradation and threats affecting the environment (Nebert, 2005). Availability and access to appropriate information, and the related development of interoperable databases, are the necessary conditions for creating the basis for supporting the information management needs of implementing and monitoring sustainable development policies and goals, such as the United Nations Millennium Development Goals (MDGs) (Henricksen, 2007). The MDGs are eight development objectives (eradicate extreme poverty and hunger, achieve universal primary education, promote gender equality, reduce child mortality, improve maternal health, combat different diseases, ensure environmental sustainability, and develop and global partnership for development) that all UN members have agreed to achieve by 2015.

Over the last twenty years, the emergence and evolution of Geographic Information Systems (GIS) technology and the advent of applications such as Google Earth or OpenStreetMap (Craglia et al., 2008), allowed for a clear change on how geospatial data are handled and incorporated into regular workflows of organizations and agencies in the governmental, private and public sectors (Booz, Allen, & Hamilton, 2005). Highlighting these changes, Masser (2007) stated that to realize the full potential and benefits of geospatial data, access must be maximized with the help of Spatial Data Infrastructures (SDIs), that allow users to share, discover, visualize, evaluate and retrieve geospatial data. Moreover, the vast amount of data needed to run a complex model (e.g., in climatology or ecology), and the recognition that organizations and/or agencies need more data than they can afford financially (Rajabifard & Williamson, 2001), reinforce the concept that once a particular set of geospatial data has been created, it should be accessible to potential users in both the public and private sectors (Ryttersgaard, 2001). This reinforces the need to store such data in databases that are made widely accessible for various purposes (Philips et al., 1999). As a consequence, geospatial data can be seen as a shared resource which is maintained continuously.

To remove the barriers that block and impede a wide use of geospatial data and related information, Masser (2005, 2007) identified different needs such as eliminating or reducing restrictions on data access and availability (but protecting intellectual property rights), promoting interoperability between different data sets and different systems, and disseminating the information about data (metadata). Altogether these objectives are designed to create an environment that fosters activities for using, managing, producing and sharing geospatial data in which all stakeholders can cooperate with each other and interact with technology, to better achieve their objectives at different political/institutional levels (Rajabifard & Williamson, 2004). In this sense, interoperability appears to be a key element enhancing data sharing, communication and efficiency.

The great advantage of interoperability is that it describes the ability of locally managed and distributed heterogeneous systems (different operating systems, different databases, different data formats) to exchange data in real time to provide a service (OGC, 2004). The shift towards a processed-based infrastructure offering reusable and standardized components responsive to user needs and requests is supported by the Service Oriented Architecture (SOA) concept. In a SOA, services are the elementary components representing a set of operations that could be invoked by users allowing them to access, in the case of the geospatial community, distributed geospatial data as well as geoprocessing services. To implement and deploy geo-enabled services, the OGC proposes a suite of standards that use services over the Internet, so-called web services, giving access to distributed data and services through Uniform Resource Locators (URLs). This allows data providers to publish standardized services independently on how it is implemented and on which platform it is executed. This emphasizes the full potential of interoperability allowing an organization to maximize the value and reusability of data under its control and giving the ability to exchange these data with other interoperable systems. Using such OGC web services offers the possibility to seamlessly couple and reuse them in a variety of applications. By chaining together a series of web services, users can perform a set of operations to process data whereby new knowledge emerges from relationships that were not envisioned before (Open Geospatial Consortium, 2004). Granell et al. (2009) define service chaining as a mechanism for combining individual geospatial web services to create customized web applications. Although current SDIs mostly offer the abilities to search, view and access data, with the support of interoperable services and SOA related concepts it is now possible to build new applications based on distributed services (Friis-Christensen, et al., 2007; Diaz et al., 2008). When services are organized through a coherent chain, combined services can achieve a larger task (Di, 2004). The International Organization for Standardization (ISO) through its ISO 19119 standard (ISO, 2005) defines three types of chaining:

- Transparent (user-defined): the workflow is defined and managed by users.
- Translucent (workflow-managed): users invoke a service that manages the chain. Users are aware of atomic services that constitute the chain.
- Opaque (aggregated): users invoke an aggregated service that carries out the chain. Users have no awareness of the atomic services that constitute the chain.

In this paper, we will focus on the transparent chaining either by hard coding or by using OGC Web Processing Service (WPS) specification (OGC, 2007).

Through its online catalogue of registered services, GEOSS is an interesting and promising entry-point to discover and access services that could be integrated into service chaining process. It offers a framework to share data, expose them through interoperable services and allow the production and dissemination of timely and accurate data needed by decision makers and the public (GEO secretariat, 2005).

SERVING DATA INTO GEOSS

In 1985, the United Nations Environment Programme (UNEP)/Division of Early Warning and Assessments/GRID-Europe was founded as one of the first two centres of the Global Resource Information Database (GRID) network to support environmental decision-making within UNEP and the UN system as a whole, by generating and disseminating information about the state of the world's environment in a timely and understandable manner. To provide reliable environmental assessments and early warnings, GRID-Europe specialized in handling and analyzing spatial and statistical data on environmental and natural resource issues through computerized GIS and remotely-sensed imagery. Over the years, GRID-Europe has compiled an archive of global, European and other geospatial databases as part of its information management function. The experience and in-house capabilities of GRID-Europe offer a great potential to make geospatial and tabular databases compiled over the years available to a large array of users. Since its foundation, the Geneva office has received considerable support from Swiss and local authorities as well. This supporting was significantly reinforced, and GRID-Europe's institutional base broadened, with the signing of a "Partnership Agreement" between UNEP, the Federal Office for the Environment (FOEN) and the University of Geneva in June 1998.

GRID-Europe closely monitors developments in information technologies and examines their utility for environmental monitoring and policy formulation and thus is extending and developing its field of activities using SDIs. Moreover, the "Partnership Agreement" provides a major opportunity to work at different geographic scales ranging from global, to regional (Europe) and national (Swiss) and finally local (Geneva). Such a specificity allows GRID-Europe to participate to different applied research projects funded either by the United Nations or the European Commission. A common ground for these projects is to serve and share data through the European Directive on Infrastructure for Spatial Information in the European Community (INSPIRE) (European Commission, 2007), the United Nations Spatial Data Infrastructure (UNSDI) (Henricksen, 2007), as well as GEOSS.

PREVIEW Global Risk Data Platform

The PREVIEW (Project of Risk Evaluation, Vulnerability, Information, and Early Warning) Global Risk Data Platform (http://preview.grid.unep.ch) is a collaborative effort of UNEP, United Nations Development Programme (UNDP/BCPR), United Nations International

Strategy for Disaster Reduction (UNISDR) and the World Bank to share geospatial data on global risk from natural hazards. Users can freely visualize, download or extract data on past hazardous events, human and economical hazard exposure and risk from natural hazards. The platform covers nine types of natural hazard: tropical cyclones and related storm surges, drought, earthquakes, biomass fires, floods, landslides, tsunamis and volcanic eruptions. The collection of data is made via a wide range of partners. This geoportal was developed as a support to the 2009 Global Assessment Report on Disaster Risk Reduction (United Nations International Strategy for Disaster Reduction Secretariat, 2009), replacing the previous PREVIEW platform initially designed by UNEP/GRID-Europe and already available since 2000. The new PREVIEW platform is fully compliant with the OGC Web Services (OWS) to access data using Web Map Service (WMS), Web Feature Service (WFS), Web Coverage Service (WCS), geo-enabled Really Simple Syndication (GeoRSS) or Keyhole Markup Language (KML) as well as metadata using Catalogue Service for the Web (CS-W).

GEO Data Portal

The GEO Data Portal (http://geodata.grid.unep.ch) is the authoritative source for data sets used by UNEP and its partners in the Global Environment Outlook (GEO) report and other integrated environment assessments. Its online database holds more than 550 different variables, as national, sub-regional, regional and global statistics or as geospatial data sets (maps), covering themes such as Freshwater, Population, Forests, Emissions, Climate, Disasters, Health and Gross Domestic Product (GDP). The data can be displayed and explored on-the-fly through maps, graphs, data tables, downloaded in various popular formats, or copied and pasted into word processors. All information products in the GEO Data Portal can be accessed and used as web services as well. The retrieval of statistical and country-wide information has been enabled via a Simple Object Access Protocol (SOAP) connection; data from the database can be retrieved as maps via WMS or WFS; graphs can be displayed via a direct Uniform Resource Locator (URL) usage.

enviroGRIDS

EnviroGRIDS (http://www.envirogrids.net) is a European research project that will last from 2009 until 2013 and is funded under the seventh framework programme (FP7). The Black Sea Catchment is largely following an ecologically unsustainable pathway based on inadequate resource management that could lead to severe environmental, social and economical problems, especially in a changing climate (WWF, 2008). The aim of the project is to build capacities in the Black Sea region to use new international standards to gather, store, distribute, analyze, visualize and disseminate crucial information on past, present and future states of this region, in order to assess its sustainability and vulnerability. EnviroGRIDS objective is to federate and strengthen existing Observation Systems to address several GEOSS Societal Benefit Areas within a changing climate framework. The expected result will be a shared information system that operates on the boundary of scientific/technical partners, stakeholders and the public. It will contain early warning systems able to inform in advance decision-makers and the public about risks to human health, biodiversity and ecosystems integrity, agriculture production or energy supply caused by climatic, demographic and land cover changes on a 50-year time horizon. To achieve and support the enviroGRIDS vision and objectives, a grid-enabled Spatial Data Infrastructure (gSDI) is under construction. The aim of the gSDI is to host and analyze the data for the assessment of GEOSS Societal Benefit Areas, as well as the data produced within the project. These data must be gathered and stored in an organized form and accessible in an interoperable way on the grid infrastructure in order to provide a high performance and reliable access through standardized interfaces.

ACQWA

ACQWA (http://www.acqwa.ch) stands for Assessing Climate impacts on the Quantity and quality of Water. It is also a FP7 European research project lasting from 2008 until 2013. As the evidence for human induced climate change becomes clearer, so does the realization that its effects will have impacts on natural environment and socio-economic systems. Some regions are more vulnerable than others, both to physical changes and to the consequences for ways of life. According to the description of work, the project will assess the impacts of a changing climate on the quantity and quality of water in mountain regions which are particularly affected by rapidly rising temperatures, prolonged droughts and extreme precipitation. Modeling techniques will be used to project the influence of climatic change on the major determinants of river discharge at various time and space scales. Regional climate models will provide the essential information on shifting precipitation and temperature patterns. Snow, ice, and biosphere models will feed into hydrological models in order to assess the changes in seasonality, amount, and incidence of extreme events in various catchment areas. Environmental and socio-economic responses to changes in hydrological regimes will be analyzed in terms of hazards, aquatic ecosystems, hydropower, tourism, agriculture, and the health implications of changing water quality. Attention will also be devoted to the interactions between land use/cover changes, and changing or conflicting water resource demands. Adaptation and policy options will be elaborated on the basis of the results. The chain of processes involved in climatic, cryospheric and hydrologic models is complex because each process impacts on different compartments of human and natural systems. Different types of data covering various geographical regions are therefore necessary to build different sets of scenarios, which translates into substantial amount of data.

TECHNICAL COMPARISON AND COMMON GROUNDS

All these projects have in common that they already share (or will share in a near future) their data and metadata into the GCI. As a prerequisite all the registered services have to be interoperable using mainly standards proposed by the OGC, but also other protocols like the Simple Access Object Protocol (SOAP). A short comparison of these different projects is

Table 1. Technical comparison of enviroGRIDS, ACQWA, GEO Data Portal and PREVIEW projects

Project name	enviroGRIDS	ACQWA	GEO Data Portal	PREVIEW
Services	WMS, WFS, WCS, CS-W, KML, GeoRSS, WPS, grid services	WMS, WFS, WCS, CS-W, WPS, KML, GeoRSS	WMS, WFS, WCS, CS-W, SOAP	WMS, WFS, WCS, CS-W, KML, GeoRSS
Software	GeoServer, ArcGIS Server, PyWPS, GeoNetwork, gLite	GeoServer, GeoNetwork, PyWPS	GeoServer, GeoNetwork, MapServer	GeoServer, GeoNetwork, MapServer
Type of models	- Hydrological models	- Snow cover mapping	- providing base layers (socio-economic, ...)	- providing base layers (events, risk, ...)
Challenges & difficulties	- linking SDI and grid infrastructure - capacity building - authorization/authentication - portal integration	- capacity building - data integration	- data integration - data/metadata harmonization - different standard implementation - capacity building	- data integration - data/metadata harmonization - capacity building

given in Table 1, indicating which services are available, which software are used to publish these services, what are the types of models used to chain these layers, and finally what are the challenges and difficulties raised while integrating these services.

Most of these projects make use of free and open source software (PostgreSQL/PostGIS, MapServer, GeoServer, GeoNetwork and PyWPS) because it can ease the portability and replicability of tools developed. Indeed, many countries with low to moderate incomes are often also affected by natural hazards, environmental threats or degradation, and these countries are especially interested to manage and share their geospatial data using free and open sources software. Having tools readily available to be deployed in these countries is a strong incentive for capacity building, knowledge transfer, and sharing of expertise.

These projects are also strongly related to capacity building in order to enhance an "open and sharing spirit". It is necessary to show and prove the benefits of data sharing through appropriate examples, to communicate best practices as much as possible and to develop guidelines and policies. Altogether this will help to reach agreement and endorsement on the use of new standards. Such a participative approach will certainly stimulate data providers to be more "open" and in consequence to share their data. The different projects presented before will organize different workshops and develop various teaching material allowing participants, ranging from students to members of government, to learn how to use the specific applications to share large amount of data. Rajabifard and Williamson (2004) believe that building capacities is an important challenge for SDIs concepts to be accepted and adopted at a large extent. For these authors, the best way to reach this objective is to establish a long-term commitment to education and research: otherwise the SDI vision will remain unclear and unachievable. Through these projects, the objective is to build the capacity of scientists to share and document their data in order to strengthen existing observation systems, the capacity of decision-makers to use it, and the capacity of the general public to understand the important environmental, social and economic issues at stake.

Through simple data integrating scenarios (integration into other web portals or applications) GEO Data Portal and PREVIEW have pointed out different issues. Integrating some socio-economic data sets coming from the GEO Data Portal with natural hazards maps of the PREVIEW project to compute, for example, economical exposition of a country to a specific hazard, was impossible. This problem comes from the different implementations of OGC specifications between Mapserver (used by the GEO Data Portal) and Geoserver (used by PREVIEW). Indeed, it appears that Mapserver use an argument "MAP" that is not standardized and not recognized by all clients. This problem will be solved by migrating to Geoserver all the data services of the GEO Data Portal. Thus implementation of a same specification can differ from one software to another and can impend a consistent integration of services.

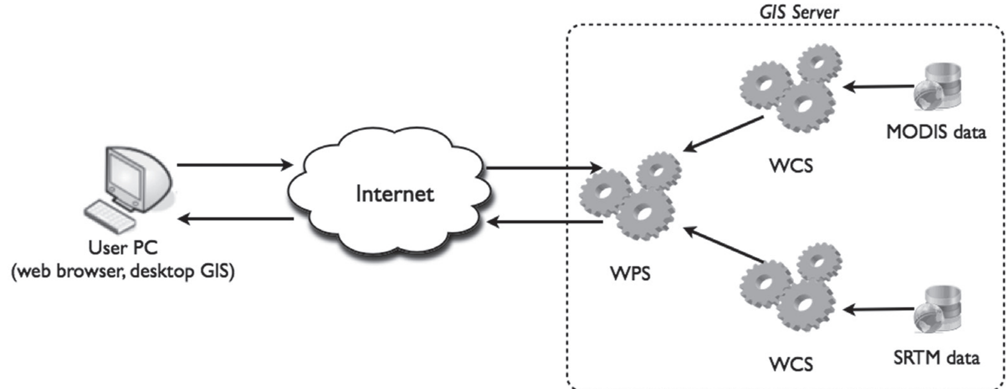
Another issue raised by data integration process was raised by the United Nations High Commissioner for Refugees (UNHCR) while trying to integrate WMS data coming from the PREVIEW project in order to identify area that are not suitable to install refugees camps. It appears that the only projection available was EPSG:4326 (Geographic) whilst UNHCR geoportal makes use of Google maps in EPSG:900913 (Spherical Mercator). This experience showed us that it is important, while publishing data services, to support at least the most frequent projection types. Geoserver supports natively all types of projections and it is easy to reproject on-the-fly data stored in another projection so that it can be integrated with data with other projections. In addition, following the size of the data set, an important processing overload has been observed caused by the on-the-fly reprojection process. This can slow the service chain and impend and efficient data integration.

In the ACQWA project, a specific constraint is the important number of partners involved and their different scientific backgrounds (climatologists, hydrologists, glaciologists, ecologists). It is quite challenging to raise awareness on new tools and way of gathering and exchanging data without strongly influencing the way these different communities are working with geospatial data. For that reason, the aim is to concentrate on the promotion of GEOSS as an interesting and useful framework to handle and discover scientific data. Obviously, a dedicated geoportal is under development to register the main outputs of the ACQWA project into GEOSS using OGC web services. Nevertheless to show the benefits of working with interoperable services, we are currently developing a scenario to make estimation of snow cover from remote sensing imagery using data coming from the Moderate Resolution Imaging Spectrometer (MODIS) and Shuttle Radar Topography Mission (SRTM). Project partners that are currently working to produce such estimations are working with PCI Geomatica, doing all the process chain manually. Our objective is to help our partners by publishing a WPS geoprocessing service that allows them to automatize this analysis (Figure 1). Once retrieved by FTP, MODIS images are saved on a server that store also SRTM tiles. All data are in EPSG:4326 and will be available using WCS standard published by Geoserver. Finally, the WPS service, currently under development using PyWPS, will implement the different steps to process the data. A major difficulty encountered until now is to "translate" the PCI functionalities by finding the equivalent in Geographic Resources Analysis Support System (GRASS) software. Indeed, PyWPS does not process data by itself and instead uses GRASS as a backend to access all the geoprocessing functionalities.

Once the snow cover process is successfully achieved, our hope is to convince other communities within the project to benefit from such an approach and to develop other scenarios especially making use of climate data.

In the process of turning data into understandable information and knowledge by chaining data services a new challenge has emerged. The ever-increasing spatial and temporal resolution of geospatial data are causing a tremendous increase in term of data volumes and the limits of the processing capacities of traditional GIS and SDI are being reached. With the advent of grid computing and the progressive deployment of large grid infrastructure projects (e.g., Enabling Grids for E-sciencE) many scientific disciplines now have access to sizable computing resources and new opportunities are emerging. For Foster et al. (2008) grid aims to federate resource sharing in a dynamic and distributed environment across a network allowing to access unused CPUs and storage space to all participating computers. Currently, SDIs are lacking processing power and should therefore be made interoperable with grid infra-

Figure 1. Data sources and processing steps for a geoprocessing service estimating snow cover



structures, which are offering large storage and computing capacities. Recent studies (Muresan et al., 2008; Di, Chen, Yang, & Zhao, 2003) applied a successful approach to extend grid computing to the remote sensing community and to make OGC web services grid-enabled. Both studies considered that the grid has a great potential for the geospatial disciplines. Padeberg and Greve (2009) have identified several differences between OGC-compliant SDIs and grid infrastructures concerning service description, service interface, service state and security. In particular, grid infrastructures are based on SOAP messaging protocol to invoke operations and Web Service Description Language (WSDL) to describe services. OGC-compliant does not support neither SOAP nor WSDL, except WPS, and thus chaining geospatial services with grid services could be problematic. In addition, OGC standards do not provide any security mechanisms (authentication, encrypted communication between resources) which is a major concern in grid infrastructures. Finally, Di et al. (2003) showed that the current grid metadata catalog system is not good enough to answer the needs of the geospatial community, especially the requirements of the ISO19115 standard. All these differences must be overcome in order to allow traditional SDIs to benefit from the power of grid computing, and consequently to offer new services to GEOSS.

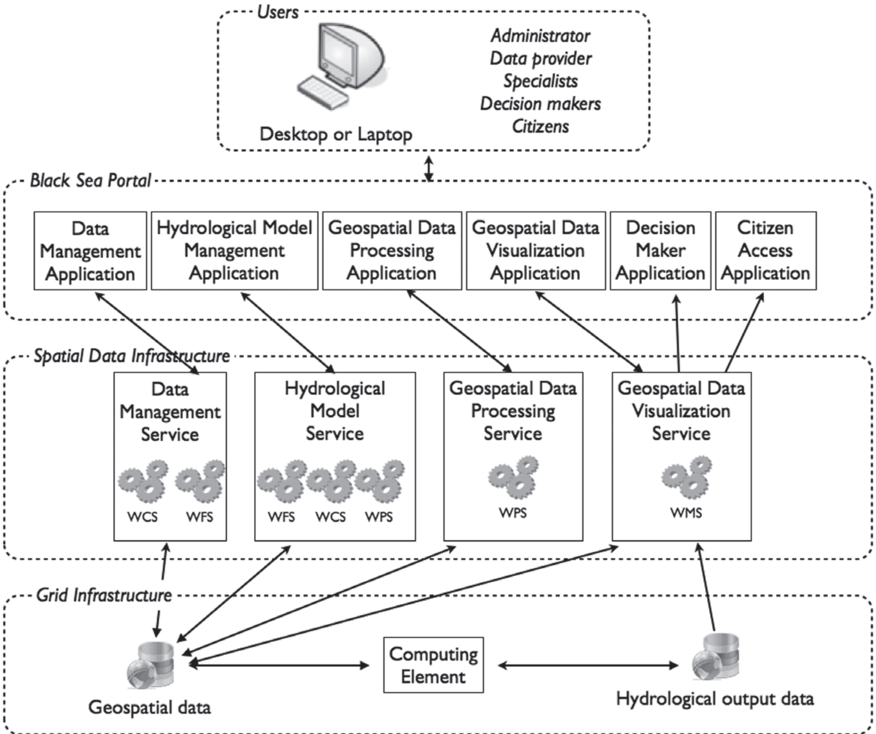
The main scientific and technological challenge of the enviroGRIDS project will be to link an SDI with a grid infrastructure to benefit from the processing capacities offered by grids. Indeed, WPS appears to be an adequate candidate to be grid-enabled because, first, it supports SOAP protocol and, second, geospatial community has a growing processing need that current SDIs cannot deliver. A grid-enabled SDI will allow users to model high resolution hydrological models (e.g., Soil and Water Assessment Tool) of the Black Sea catchment under various climate, land cover and demographic scenarios. In order to develop such a gSDI to support the development of Black Sea portal functionalities, the different components of the enviroGRIDS architecture are currently being defined to highlight the main issues emerging from different conceptual and technological solutions (Figure 2).

These issues concern the choices of software components, data repositories, data management, grid-oriented processing, grid portal, and interoperability between SDIs and grid infrastructures. Although the use of grid-enabled web services to access data sets stored in the SDI will also be explored (Maué & Kiehle, 2009), bridging architectural gaps between grids and SDIs remains very challenging (Padberg & Greve, 2009) without extensions and customizations. For example, an important question concerns the location of geospatial data repositories: inside or outside the grid? The answer is not trivial and will greatly influence services and in particular chains of services to process data. In the one hand, being outside the grid, all OGC-compliant services functionalities remain the same and grid services are only used to process the data. On the other hand, being inside the grid, all OGC web services have to be modified to support grid environment, becoming grid-enabled. The latter would allow benefiting from all the advantages of the grid (security, replicability, scalability, storage and processing capacities) but would obviously require a lot of developments for adapting already existing SDIs. In consequence, an incremental development and implementation strategy will be developed taking into account different integration scenarios aiming to hide the complexity of the grid while preserving OGC interfaces.

CHALLENGES AND PROMISES

From the experience acquired, or being acquired, through these different projects, it is obvious that many challenges remain both tangible (e.g., technology) or less tangible (e.g., culture, behavior). Nevertheless, it is critical to overcome them in order to improve our knowledge, share our experience and attempt to strive towards a society that is better informed. Achieving the goal of sustainable

Figure 2. EnviroGRIDS grid-enabled SDI components supporting Black Sea portal



development requires the integration of a large number of different data from various sources. Through agreed common standards and a clear political will, these data can be integrated in an interoperable way, leading to a new collaborative approach to decision-making.

Having environmental data in digital form allows easy storage and dissemination, facilitate data exchange and sharing, faster and easier update and corrections, ability to integrate data from multiple source (see Figure 1), and customization of products and services (Henricksen, 2007). In this sense SDIs appear to be a good choice to encompass the sources, systems, network linkage standards and institutional issues involved in delivering geospatial data from many data sources to the widest possible group of potential users (Coleman, McLaughlin, & Nichols, 1997). The fact that, during the last years, multiple SDIs initiatives have been developed all around the world, ranging from local to regional levels, is a good sign. It appears that there is a growing recognition that geospatial data is a critical element underpinning decision making in many disciplines (Rajabifard & Williamson, 2001) and as such needs to be effectively managed.

The SDI hierarchy model proposed by Rajabifard (2002) is composed of inter-connected SDIs developed at different levels (from local to global). Each SDI of a higher level is formed by the integration of data developed and made available by the lower level. Such a hierarchy can be approached though two views: on one hand, it is an umbrella in which the SDI at a higher level encompasses all SDI components from lower levels. On the other hand, it can be seen as the building blocks supporting the access of data needed by SDIs at higher levels. This hierarchy allows creating an environment in which users working at any level can rely on data from other levels and integrate data from different sources (Mohammadi & Rajabifard, 2009). Such a hierarchy is clearly envisioned in the concept of the system of systems on which GEOSS relies, integrating systems together into an information highway which both links together environmental, socio-economic and institutional databases and provides a movement of information from local to global levels. For Masser (2006), the SDI hierarchy poses the challenge of multi-stakeholder participation in SDI implementation, because the bottom-up approach differs a lot from the top-down approach. The top-down vision, common in the SDI literature, emphasizes the need for standardization and uniformity while the bottom-up view stresses the importance of diversity and heterogeneity caused by the different needs of the various stakeholders. As a consequence, it is necessary to find a consensus ensuring sufficient standardization and uniformity while recognizing the diversity and heterogeneity of the different stakeholders acting at different levels. In particular, building a system of systems like GEOSS is highly dependent on a clear governance structure that is understandable and acceptable by the volunteer participants in order to develop a shared vision of the system and to allow users to feel a common sense of ownership (Masser, 2007). As it is reminded in the Strategic Guidance document (GEO secretariat, 2007a), the success of GEOSS will depend on interoperability arrangements that data providers agree to endorse.

As a provider of environmental data, GRID-Europe is continuously facing the challenge of encouraging data providers to go "open" and to share their data in an interoperable and OGC-compliant way. At present, technology is no longer a problem because solutions based on a variety of software can be proposed and/or developed depending on the requirements and the technical capabilities available. The most difficult task is to create an environment allowing wide agreement on data sharing principles. In this particular regard, the GEOSS "best practices wiki" could be of great benefit to help people promote sharing principles. A lesson learned from our experience is that once users can discover data they need, their most important preoccupation is to know what is the quality of the data they are going to access and whether they can trust this data. We are convinced that sharing data is an efficient way to eventually recognize whether this data is of sufficient quality. By submitting/exposing the data to the judgment of the broader community, one can know if it is useful or not. Through data sharing, one can also benefit from the interaction with end users by receiving feedbacks and then improve the data sets accordingly. Sharing data and participating to GEOSS can therefore contribute to the improvement of data, which in turn allows better information and eventually better decisions.

In the current climate of economic constraints, interoperability and standardization have never been so important because a non-interoperable system impedes the sharing of data, information and resources, which increase the risk for a system to fail in delivering its expected benefits and to remain unused (Open Geospatial Consortium, 2004). Geospatial data can be an expensive and time consuming resource to produce, and for this reason, it is of high importance to improve accessibility and availability and promote its reuse. Many decisions that organizations need to make depend on good quality and consistent data, readily available and accessible (Rajabifard & Williamson, 2001). The process of reuse does not only concern the data itself, but also encompasses the capabilities, skills developed, invested effort and capital. This process allows an organization to share the costs of data, people, and technology, which helps realize more rapid returns on investment. By reusing data, one can avoid duplication of efforts and expenses and enable users to save resources, time and effort when trying to acquire or maintain data sets (Rajabifard & Williamson, 2001).

Percivall (2006) claimed that in a distributed environment, the help of open standards such as OGC can help scientists to rapidly find and evaluate a lot of different data sets and processing approaches, providing a flexible and cooperative environment that foster collaboration in the different scientific communities that work with geospatial data. Thus, organizing the workflows using standard-based web services could provide a great benefit in term of productivity to address the nine SBAs of GEOSS. OGC standards provide a solid ground for interoperability between services within distributed geoprocessing environment offered by SDIs (Friis-Christensen et al., 2007). In particular, the fact that these services can be reused and chained within other applications is a very useful aspect offering the opportunity to solve specific problems in a more flexible way than with stand-alone applications. Nevertheless, some performance issues can appear with services that need to access and move large amount of data. This can negatively impact the execution time of this service (e.g., huge overload in gathering necessary data) especially if this service is chained with other services.

Consequently, GEOSS represents a very promising and potentially powerful framework to share and expose data. In particular, the fact that a good governance structure is already in place allows a clear vision that can be easily shared and endorsed by the participants. The fact that participating to GEOSS is on a voluntary basis could be seen either as a great opportunity or as a risk. Indeed, the voluntary aspect poses the threat that only a few data providers join such an initiative and, as a consequence, the system could miss its objectives. Nevertheless, the growing number of components and services registered through GEOSS is a good sign for optimism. In particular, we think that international organizations such as UNEP could play a major role by paving the way toward a broader acceptance by similar organizations. The fact that GEOSS is based on distributed systems that can operate, evolve and be managed in a relative independence appears to be a good choice to find a consensus ensuring sufficient standardization and uniformity, while recognizing the diversity and heterogeneity of the different stakeholders. Finally, GEOSS offers a unique characteristic that justifies by itself its existence, which is the possibility to see emergent properties. For Béjar et al. (2009), this emergence is the main objective of a system of systems, where users perform functions that cannot be made with any single component. This means that such a system is more than the sum of its parts and offers the possibility to better understand the complex relationships between the different components of the Earth system.

CONCLUSIONS

Geospatial data is a critical element underpinning decision-making for many disciplines and is indispensable to make sound decisions at all levels, from global to local. Experiences from developed countries show that more than two-thirds of human decision-making are affected by spatially-referenced data (Ryttersgaard, 2001). Even if the technology exists, organizations and agencies around the world are still spending billions of dollars every year to produce, manage and use geospatial data, but they still do not have the information they need to answer the challenges our world is facing (Rajabifard & Williamson, 2001).

The web service model proposed by the OGC appears to be suitable to allow users to combine different services to solve a specific problem in a scalable and flexible way. Nevertheless, through simple examples of services chaining, we have highlighted different issues that could potentially impede an easy integration: problems with different implementation of a same specification, problems regarding different projections used in different web applications, overload caused by on-the-fly reprojection using large data sets. Moreover, working with different communities that are not necessarily aware of the possibilities offered by OGC web services could limit the diffusion of such approach outside the geospatial community. These communities need to be convinced, through simple examples, which working with chained services can bring benefits in their own working flows. Finally, grid computing appears to be a promising complement of traditional SDIs capabilities to build WPS services for processing large data sets. To achieve this objective, implementation of SOAP protocol into OGC specifications is a pre-requisite in order to allow the two types of infrastructures to communicate (interoperability) and to ease the combination of OGC and grid services in efficient chains.

Ten years after, GEOSS could be seen as an initial step to achieve Gore's vision, because the relevant technologies are available and there is growing recognition that countries can benefit both economically and environmentally from better access to data. GEOSS has the potential to support the achievement of sustainable development initiatives such as the UN Millennium Development Goals and to offer a unique framework to share data and collaborate for a better society. In this sense, organizations such as UNEP can act as a "catalyst", contributing to GEOSS, building capacities and ensuring that environmental data are easily accessible. This is a necessary step to ensure better-informed decision-making for the more sustainable development of our planet.

ACKNOWLEDGMENTS

The authors would like to acknowledge the European Commission "seventh framework programme" that funded the enviroGRIDS (Grant Agreement n° 226740) and ACQWA (Grant Agreement n°212250) projects, and UNEP for its support. The views expressed in the paper are those of the authors and do not necessarily reflect the views of the institutions they belong to.

REFERENCES

Arzberger, P., Schroeder, P., Beaulieu, A., Bowker, G., Casey, K., & Laaksonen, L. (2004). Promoting Access to Public Research Data for Scientific, Economic and Social Development. *Data Science Journal*, *3*, 17. doi:10.2481/dsj.3.135

Béjar, R., Latre, M. A., Nogueras-Iso, J., Muro-Medrano, P. R., & Zarazaga-Soria, F. J. (2009). Systems of Systems as a Conceptual Framework for Spatial Data Infrastructures. *International Journal of Spatial Data Infrastructures Research*, *4*, 17.

Bernard, L., & Craglia, M. (2005). *SDI - From Spatial Data Infrastructure to Service Driven Infrastructure*. Paper presented at the 1st Research Workshop on Cross-learning on Spatial Data Infrastructures (SDI) and Information Infrastructures (II), Enschende, The Netherlands.

Booz, A. (2005). *Geospatial Interoperability Return on Investment*. Hamilton: NASA.

Coleman, D. J., McLaughlin, J. D., & Nichols, S. (1997). *Building a Spatial Data Infrastructure*. Paper presented at the 64th Permanent Congress Meeting of the Fédération Internationale des Géomètres (FIG), Singapore.

Craglia, M., Goodchild, M. F., Annoni, A., Camara, G., Gould, M., & Kuhn, W. (2008). Next-Generation Digital Earth: A position paper from the Vespucci Initiative for the Advancement of Geographic Information Science. *International Journal of Spatial Data Infrastructures Research*, *3*, 22.

Di, L. (2004). Distributed Geospatial Information Services – Architectures, Standards, and Research Issues. In *Proceedings of the International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, Istanbul, Turkey (Vol. 35, Part 2, Commission II).

Di, L., Chen, A., Yang, W., & Zhao, P. (2003). *The Integration of Grid Technology with OGC Web Services (OWS) in NWGISS for NASA EOS Data*. Paper presented at the GGF8 & HPFC12, Seattle, WA.

Diaz, L., Granell, C., & Gould, M. (2008). Case Study: Geospatial Processing Services for Web-based Hydrological Applications. *Geospatial Services and Applications for the Internet*, 31-47.

European Commission. (2007). *Proceedings of the Directive 2007/2/EC of the European Parliament and the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)*, Brussels, Belgium.

Foster, I., Yong, Z., Raicu, I., & Lu, S. (2008). *Cloud computing and grid computing 360-degree compared*. Paper presented at the 2008 Grid Computing Environment Workshop, Austin, TX.

Friis-Christensen, A., Lutz, M., Ostländer, N., & Bernard, L. (2007). Designing Service Architecture for Distributed Geoprocessing: Challenges and Future Directions. *Transactions in GIS*, *11*(6), 799–818. doi:10.1111/j.1467-9671.2007.01075.x

GEO secretariat. (2005). *Global Earth Observation System of Systems 10-Year Implementation Plan Reference Document*. Geneva, Switzerland: GEO.

GEO secretariat. (2007a). *Strategic Guidance for Current and Potential Contributors to GEOSS*. Geneva, Switzerland: GEO.

GEO secretariat. (2007b). *The Full Picture*. Geneva, Switzerland: GEO.

Gore, A. (1998). *The Digital Earth: Understanding our planet in the 21st Century* (p. 4).

Gorgan, D., Bacu, V., Ray, N., & Maier, A. (2009). *EnviroGRIDS data storage guideline* (p. 67).

Granell, C., Gould, M., & Esbri, M. A. (2009). Geospatial Web Service Chaining. In Karimi, H. A. (Ed.), *Handbook of Research on Geoinformatics* (pp. 189–195). Hershey, PA: IGI Global.

Henricksen, B. (2007). *UNSDI Compendium: A UNSDI Vision, Implementation Strategy, and Reference Architecture*. UNGIWG.

ISO. (2005). *ISO 19119: Geographic Information – Services*. Gland, Switzerland: ISO.

Masser, I. (2005). *The Future of Spatial Data Infrastructures*. Paper presented at the ISPRS Workshop on Service and Application of Spatial Data Infrastructure, Hangzhou, China.

Masser, I. (2006). Multi-level Implementation of SDIs. *GIM International*, *20*, 4.

Masser, I. (2007). *Building European Spatial Data Infrastructure*. Redlands, CA: ESRI Press.

Maué, P., & Kiehle, C. (2009). Grid technologies for geospatial applications: an overview. *GIS Science*, *3*, 65–67.

Mohammadi, H., & Rajabifard, A. (2009). Multisource Spatial Data Integration within the Context of SDI Initiatives. *International Journal of Spatial Data Infrastructures Research*, *4*, 18.

Muresan, O., Pop, F., Gorgan, D., & Cristea, V. (2008). *Satellite Image Processing Applications in MedioGRID*. Paper presented at the 5th International Symposium on Parallel and Distributed Computing.

Nebert, D. D. (2005). *Developing Spatial Data Infrastructure: The SDI Cookbook*.

(2004). *Open Geospatial Consortium* (p. 7). The Havoc of Non-Interoperability.

Open Geospatial Consortium. (2007). *OpenGIS Web Processing Service* (p. 87).

Padberg, A., & Greve, K. (2009). Gridification of OGC web services: challenges and potential. *GIS Science*, *3*, 77–81.

Percivall, G. (2006). *GEOSS to Benefit from "Service Chaining" Based on OGC® Standards* (p. 5). Geoinformatics.

Philips, A., Williamson, I., & Ezigbalike, C. (1999). Spatial Data Infrastructure concepts. *The Australian Surveyor*, *44*(1), 8.

Rajabifard, A. (2002). *Proceedings of Diffusion of Regional Spatial Data Infrastructures: with Particular Reference to Asia and the Pacific*, Melbourne, Australia.

Rajabifard, A., & Williamson, I. P. (2001). *Spatial Data Infrastructures: Concept, SDI Hierarchy and Future Directions*. Paper presented at the Geomatics'80, Tehran, Iran.

Rajabifard, A., & Williamson, I. P. (2004). *SDI Development and Capacity Building*. Paper presented at the GSDI-7, Bangalore, India.

Ryttersgaard, J. (2001). *Spatial Data Infrastructure: Developing Trends and Challenges*. Paper presented at the International Conference on Spatial Information for Sustainable Development, Nairoibi.

United Nations International Strategy for Disaster Reduction Secretariat (UNISDR). (2009). *Global Assessment Report on Disaster Risk Reduction 2009.* Geneva, Switzerland: UN.

World Wide Fund. (2008). *Greening the Black Sea Synergy*. Brussels, Belgium: WWF.

Gregory Giuliani obtained a degree in Earth Sciences and then went on to complete a master in environmental sciences, specializing in remote sensing and GIS. He previously worked as a GIS Consultant for the World Health Organization, as a University tutor in remote sensing and GIS and as a GIS developer in a local Swiss GIS company. He works at UNEP/GRID-Europe since 2001 and is the focal point for Spatial Data Infrastructure (SDI). In 2008, he also started to collaborate closely with the envioSPACE laboratory where he started a PhD thesis on Spatial Data Infrastructures for environmental data. He is work package leader in the EU-FP7 enviroGRIDS project where he coordinates SDI development, and is also participating in the EU-FP7 ACQWA project. Finally, he developed the PREVIEW Global Risk Data Platform, a web application to share geospatial data on natural hazards.

Nicolas Ray, MSc environmental science, PhD biology, is a certified PRINCE2 project manager. He is currently the project manager of the FP7 enviroGRIDS project and coordinator of all grid-computing related activities, splitting his time between the enviroSPACE group and GRID-Europe. He is skilled in trandisciplinary approaches to solve complex spatially-explicit problems linked to the modeling of animal movement and habitat, with the development of several spatial and statistical analysis tools to integrate various data types.

Stefan Schwarzer started as an intern with GRID-Europe in 1998, and joined the GRID team in May 2000. With academic training in geography, Stefan holds a Diploma in Applied Physical Geography form the University of Trier/Germany. He is mainly responsible for the GEO Data Portal, an on-line database. Main areas of work include statistics, GIS, website design and programming.

Andrea De Bono, MSc and PhD in Geology, is a specialist in geomatics. He currently works on database management and global data sets process for the UNEP GEO Data Portal. He is also taking part in the integration of scenarios of demographic, climatic, and land use changes within the enviroGRIDS project. He previously worked for the UN GAR-Disaster Risk Reduction as GIS analyst mainly focused on population exposure to natural hazards.

Pascal Peduzzi is head of the Global Change & Vulnerability Unit at United Nations Environment Programme (UNEP/GRID-Europe). He has a master of science in remote sensing and leads a team of scientific researchers to quantify the effects of environmental degradations (deforestation, climate change, decline of ecosystems) on populations and ecosystems using satellite imagery, GIS modelling and terrain observations. He coordinated the Global Risk Analysis for the 2009 Global Assessment Report on Disaster Risk Reduction and is one of the lead Author of the next IPCC report on extreme events. He is the author of various scientific publications and provided numerous conferences to governments and the general public to raise awareness on the consequences of human impacts on the environment.

Hy Dao studied human geography at the University of Geneva and later got a post-graduate diploma from the International Institute for Aerospace Surveys and Earth Sciences (ITC, Enschede, The Netherlands). He holds a PhD in human geography from the University of Geneva. He is currently heading the Metadata and Socio-Economics unit at GRID-Europe where he works at half-time on secondment by the University of Geneva where he is also senior lecturer in geographic information. His fields of activities include sustainable development indicators, population mapping, vulnerability assessment, graphic semiology, and e-learning.

Jaap Van Woerden is environmental data and indicator specialist at UNEP's Division of Early Warning and Assessment office in Geneva. He has extensive experience in environmental data coordination & management and indicator development. Mr Van Woerden holds a Masters degree in (Human/Urban) Geography (University of Utrecht, Netherlands) and is author of various reports and publications on core environmental data and indicators in support of assessment and reporting on the global environment. He coordinates data and indicators for UNEP's Global Environment Outlook (GEO) assessment work and is lead author of the GEO Indicators published in the GEO Yearbook series, which includes glacier melting and other 'headline' global change indicators.

Ron Witt is the regional coordinator of the UNEP/DEWA/GRID-Europe. He has worked for UNEP both at its Nairobi Headquarters and the Geneva office for over 24 years, beginning as a GIS/remote sensing specialist at the GRID-Nairobi centre, and then in Geneva as database manager and manager of the centre, as well as the DEWA Regional Coordinator. Mr. Witt is responsible for UNEP/DEWA's broad programme of activities for the pan-European region, focusing on early warning, assessment, environment and security and capacity building activities for all of these subjects within (South-)Eastern Europe, the Caucasus and Central Asia. Prior to his career with UNEP, Mr. Witt worked for the U.S. National Aeronautics and Space Administration (NASA) at the Goddard Space Flight Center (GSFC) as a state programme manager and "technology transfer specialist" at the Eastern Regional Remote Sensing Applications Center (ERRSAC), and later carried out applied research in remote sensing applications within the Earth Resources Branch at NASA/GSFC in Greenbelt, Maryland, USA. Mr. Witt holds a Bachelor's degree in European history and Geography from Middlebury College, Vermont, and a Masters degree in Geography and remote sensing from the University of Utah.

Martin Beniston, is a full professor, holds the Chair for Climate Research, and is director of the Institute of Environmental Sciences at the University of Geneva. He holds a Bachelor Honours degree in Environmental Science from the University of East Anglia, a Masters degree in Atmospheric Science from the University of Reading, a doctorate from the University Pierre and Marie Curie (Paris) in atmospheric modelling, and a habilitation from ETH-Zurich in climate modelling. He has worked as a researcher in Paris, Canada, the Max-Planck-Institute in Hamburg, and at the Ecole Polytechnique Fédérale in Lausanne. He chaired the Swiss National Climate Program in Berne from 1990-1992, then shared his time between the IPCC (as one of the "Impacts" working group co-chairs from 1992-1997) and senior scientist at ETH-Zurich. He was appointed full professor and head of the Geosciences Department of the University of Fribourg in 1996 and held that position for 10 years until his latest appointment in Geneva in October, 2006.

Anthony Lehmann, MSc and PhD in biology, is a specialist in statistical analysis. He is the coordinator of the enviroGRIDS project and work package leader in the EC-FP7 ACQWA project. He is heading both the enviroSPACE laboratory at the University of Geneva and the Environment Monitoring and Modelling unit at UNEP/GRID-Europe. He has a long-term experience developing projects on biodiversity indicators and their spatial downscaling, and generalized regression and spatial predictions.