GLOBAL DATA NETWORKS IN THE ENVIRONMENTAL AND LIFE SCIENCES

Wilson M.D.
School of Geographical Sciences, University of Bristol, Bristol, UK.

Keywords: Global data networks, data cooperation, data access, global observation

Contents

Summary

International data cooperation has travelled a long way since the World Data Center system was set up to serve the International Geophysical Year of 1952-1953. As environmental awareness has increased and pushed environmental issues up the political agenda, the importance of data cooperation has steadily risen due to the need for international, multivariate datasets to address complex, global problems. The number of World Data Centres has rapidly expanded, and in 1985 the Global Resource Information Database was created to facilitate information exchange. Finally, the creation of the Integrated Global Observing Strategy in 1998 began the process of solidifying international data cooperation into a single, although complex framework, which users can access on the internet through the Global Observing Systems Information Centre.

1. A background to global data networks

The importance of cooperation for data acquisition and use is due to the value that is gained from data integration. In *International Cooperation for Data Acquisition and Use*, Clark gives six bases for data integration: comparison, the identification of process drivers, change detection, hypothesis testing, the generation of regional or global typologies, and impact evaluation and management response. International cooperation is necessary to integrate data at an appropriate level, with which to drive scientific study at the international scale. The importance of this is emphasized in Chapter 40,

"Information for Decision Making" of Agenda 21, produced following the United Nations Conference on Environment and Development held in Rio de Janeiro, June 1992:

"More and different types of data need to be collected, at the local, provincial, national and international levels, indicating the status and trends of the planet's ecosystem, natural resource, pollution and socio-economic variables. The gap in the availability, quality, coherence, standardization, and accessibility of data between the developed and the developing world has been increasing, seriously impairing the capacities of countries to make informed decisions concerning environment and development." Chapter 40, Agenda 21. UNEP, 1992.

There are many organizations cooperating to provide international data exchange. The internet now provides an ideal gateway for researchers and organizations to locate and exchange data and information as a matter of routine. This article provides an overview of the three major networks in the environmental and life sciences: the ISCU World Data Center system, the Global Resource Information Database, and the centres that form part of the Integrated Global Observing Strategy. Much of the information presented has been compiled from the institution's websites. Such is the complexity and scale of global data networks, it has only been possible to provide a broad, high-level overview. The reader is directed to relevant websites for further, more detailed information (Appendices 1 to 3).

1.1 A brief history of data cooperation

The first large-scale international scientific enterprises took place over a century before Agenda 21 was written, with the International Polar Years of 1882-1883, held again in 1932-1933. These led to the International Geophysical Year (IGY) of 1957-1958. Planning of the IGY was coordinated by the Comité Special de l'Année Geophysique Internationale (CSAGI), set up by the International Council of Scientific Unions (ICSU). The World Data Center system was set up by CSAGI to serve the IGY (Section 2). Following a successful IGY, the World Data Center system was made permanent. It has continued to support new programs such as the International Magnetospheric Study of 1976-1979.

Development of international data networks continued after the Conference on the Human Environment in Stockholm, 1972. This led to the Global Environmental Assessment Programme (Earthwatch), run by the newly established United Nations Environment Programme (UNEP). The action plan of Earthwatch included the two functions "monitoring" and "information exchange", as defined below.

- *"Monitoring: to gather certain data on specific environmental variables and to evaluate such data in order to determine and predict important environmental conditions and trends."*
- *"Information exchange: to disseminate knowledge within the scientific and technological communities and to ensure that decision-makers at all levels shall have the benefit of the best knowledge that can be made available in the forms and at the times in which it can be useful."* Action plan of the Report of the United Nations Conference on the Human Environment, UNEP 1972.

These principles led to the creation of the Global Resource Information Database (GRID) in 1985, to manage and distribute data that had been gathered by UNEP for more than a decade (Section 3).
Agenda 21 emphasized the need to strengthen existing cooperative systems such as GRID. Accordingly, in 1998 the Integrated Global Observing Strategy (IGOS) was created, which unites the major satellite and surface-based systems for global environmental observations of the atmosphere, oceans and land. It involves a strategic planning process linking research, long-term monitoring and operational programs, as well as data producers and users, in a structure that helps determine observation gaps and identifies resources to fill observation needs. IGOS is intended to cover all forms of data acquisition concerning the physical, chemical, biological and human environment including the associated impacts. Data acquisition within IGOS is user driven, leading to results that increase scientific understanding and guide early warning, policy-setting and decision-making for sustainable development and environmental protection.

The IGOS Partners include the following:

- The Committee on Earth Observation Satellites (CEOS).
- Integrated research programs on global change within the World Climate Research Programme (WCRP) and the International Geosphere-Biosphere Programme (IGBP).
- The International Group of Funding Agencies for Global Change Research (IGFA) and international agencies sponsoring global observations, including the Food and Agriculture Organization of the United Nations (FAO), United Nations Educational, Scientific and Cultural Organization (UNESCO), Intergovernmental Oceanographic Commission of UNESCO (IOC), International Council for Science (ICSU), United Nations Environment Programme (UNEP), and World Meteorological Organization (WMO).
- The Global Observing Systems (G3OS), comprising the Global Climate Observing System (GCOS), the Global Ocean Observing System (GOOS), and the Global Terrestrial Observing System (GTOS).

As a key part of IGOS, the Global Observing Systems (Section 4) make use of and expand the existing network of global data centres, strengthening the link between ground based observing systems and satellite remote sensing. Collectively they are known as the G3OS, and provide easy access and exchange of observation data.

2. World Data Centres

The ICSU World Data Center system mission statement illustrates its broad remit:

"Data constitute the raw material of scientific understanding. The World Data Centre system works to guarantee access to solar, geophysical and related environmental data. It serves the whole scientific community by assembling, scrutinizing, organising and disseminating data and information."

Mission Statement of the ICSU World Data Centre system (Guide to the World Data Centre system, ICSU 1996).

Since it was established in 1957, the number of ICSU World Data Centres (WDCs) has expanded to a total of around 50 (Appendix 1). They are operated for the benefit of the international scientific community, and make information on data holdings freely available, usually through their respective websites. Confidential or security-classified data are not held in WDCs. Data are provided to scientists free of charge, or on an exchange basis, or provided for the cost of the media and delivery.

Data are acquired by WDCs from individual scientists, projects, institutions, local and national data centres, and other WDCs. Sources for data include routine monitoring programs and ICSU-sponsored scientific programs. In addition, the data rescue program safeguards or digitizes older data sets, making them available for modern scientific use. Where the physical size of a dataset is large (such as those acquired through remote sensing), a WDC may provide metadata but not the full dataset. The WDC will instead refer users to the data, which is often archived at the original acquiring institution.

To access data held in a WDC, scientists should contact the relevant center. Contact details are given in Appendix 1 of this article. Information on data holdings for each centre may be found on the WDC website, or in the Guide to World Data Centres.

2.1 WMO World Data Centres

Several organizations exist, which refer to themselves as World Data Centres, but which are not affiliated with the ICSU, and may operate under different terms. The most notable are those operated by World Meteorological Organisation (WMO) the majority of which form part of the Global Atmosphere Watch (GAW). The functions of these centres are to acquire and archive data on the background chemical composition and related physical characteristics of the atmosphere from all parts of the globe, and to manage the resulting database. Data and derived products and information are distributed, and services provided to users in collaboration with the stations and laboratories submitting the data, other GAW and non-GAW centres and the WMO Secretariat.

3. Global Resource Information Database

GRID is a worldwide network of centres collaborating to provide, analyze and exchange geographic datasets, with a focus on the nature, distribution, and use of environmental and natural resources. GRID is one of three elements in the Environment Information Networking (EIN) Unit of UNEP. The activities of GRID complement UNEP-Infoterra and the Environmental and Natural Resource Information Network (ENRIN).

GRID is an environmental data management service within the United Nations (UN), maintaining a global archive of environmental information for use by analysts and decision makers at all levels. Each GRID centre is responsible for the acquisition, management and distribution of data for either a region or a specific theme. Centres also participate in collaborative projects with other environmental bodies in their region, offer technical advice for institutions and support training courses and workshops. A list of GRID centres is given in Appendix 2.

The GRID network has three long-term objectives (GRID website, 1996):

1. To enhance availability and open exchange of global and regional environmental geo-referenced data sets.
2. To provide the United Nations and intergovernmental bodies with access to more accurate environmental data management technologies.
3. To enable all countries in the world to make use of GRID-compatible technology for national environmental assessment and management.

Datasets from the GRID archive are distributed cost free to users around the world. The majority of datasets may be duplicated and distributed freely. Datasets are classified into the following general categories (GRID data release policy, November 1992):

A. **Global and regional datasets from UN, non-governmental or intergovernmental sources:** Public domain datasets which have been released by a UN organization or an intergovernmental organization, either as maps, atlases, books or in other published form, or which are available as digital data.
B. **Global and regional datasets from national organizations:** Developed within national agencies, typically for analysis of global processes or as an aid to international communication. This category could include datasets derived by cooperating GRID-compatible national centres. For most practical purposes they are public domain.
C. **National datasets from international development activities and non-governmental organizations:** Obtained from the same sources as type A data, these data are limited in scope to one country or a part thereof. They have been generated through the activities of an internationally financed development project, and are generally considered by the country in question to be suitable for circulation to users outside that country.
D. **National datasets from national organizations:** National or sub-national datasets arising out of a national organization project activity, often generated from GRID supported case studies. Such datasets are usually privileged by the terms of the case study agreement. Most national and sub-national datasets in this category may be released only with the expressed permission of the relevant authority. Spatially re-sampled versions of the data may be used to update other GRID datasets.
E. **Datasets from commercial organizations or centres subject to cost recovery:** Datasets produced by profit-making organizations, either for internal use or for public sale, or by national institutions which are bound by legislation to recover costs of data acquisition and distribution. These reside in the GRID archive by arrangement and are for in-house use only. Many of these datasets are satellite sensor images sold by commercial vendors.

4. Global Observing Systems Information Centre

A key part of the Integrated Global Observing Strategy, the Global Observing Systems Information Centre (GOSIC) offers open access to the G3OS via the internet: the Global Climate Observing System (GCOS), the Global Ocean Observing System (GOOS), and the Global Terrestrial Observing System (GTOS). GOSIC provides information on the

G3OS centres, and the data sets, products, and services that are available. After being set up initially as a pilot project by the Joint Data and Information Management Panel (JDIMP) of the G3OS in 1997, GOSIC is now established and in 2000 the role of guidance and evaluation passed to the G3OS steering committees.

G3OS data and information management is achieved using a highly distributed system of operational centres, data centres, and scientific organizations. Information, data, and data products, are prepared and maintained in the G3OS centres and GOSIC provides links to them (Figure 1). GOSIC also provides links to the G3OS planning and design documents wherever they may be held. This minimizes the difficulties of navigating many complex and vastly different organization web sites.

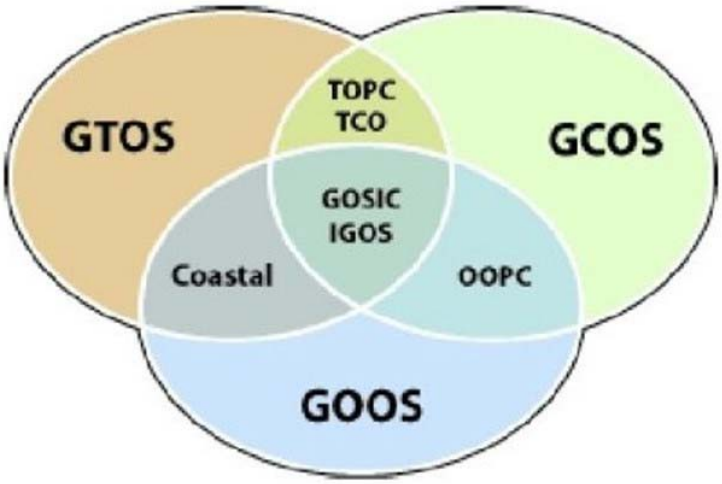

Figure 1. The G3OS centres, reproduced from http://www.fao.org/gtos/Orgpartnrs.html

GTOS: Global Terrestrial Observing System: http://www.fao.org/gtos/
GCOS: Global Climate Observing System:
http://www.wmo.ch/web/gcos/gcoshome.html
GOOS: Global Ocean Observing System: http://ioc.unesco.org/goos/
TOPC: Terrestrial Observation Panel for Climate:
http://www.wmo.ch/web/gcos/topc.htm
OOPC: Ocean Observation Panel for Climate:
http://www.bom.gov.au/bmrc/ocean/OOPC/
IGOS: Integrated Global Observing Strategy: http://www.igospartners.org
GOSIC: Global Observing System Information Centre: http://www.gos.udel.edu/
Coastal: Collaboration on coastal zones between GOOS and GTOS.

There are four methods provided to navigate G3OS data and information using GOSIC (GOSIC website, 2001):

1. View and link via the elements of the G3OS. GOSIC provides online access to data and information held by the agencies that operate the subsystems of G3OS. These agencies contribute data and information management functions according to their

capability and expertise. Examples of subsystems include the GCOS Surface Network (GSN), the GOOS Global Temperature and Salinity Profile Project (GTSPP), and the GTOS Thematic Network - Ecology (GTN-E).

2. Search the Data Set Registry for historical and operational data sets. The "Data Set Registry" is a directory of historical and operational data sets that have been identified as meeting a requirement of one of the observing systems. The Registry also includes entries for historical data sets that are not part of an initial observing system. The registry is based on the fields and structure of the NASA Global Change Master Directory (GCMD). It is operated by NASA and features both G3OS and NASA views of the contents. It thus serves both the G3OS and NASA user communities. The Registry is searchable by spatial extent, temporal range and keywords (Figure 2).

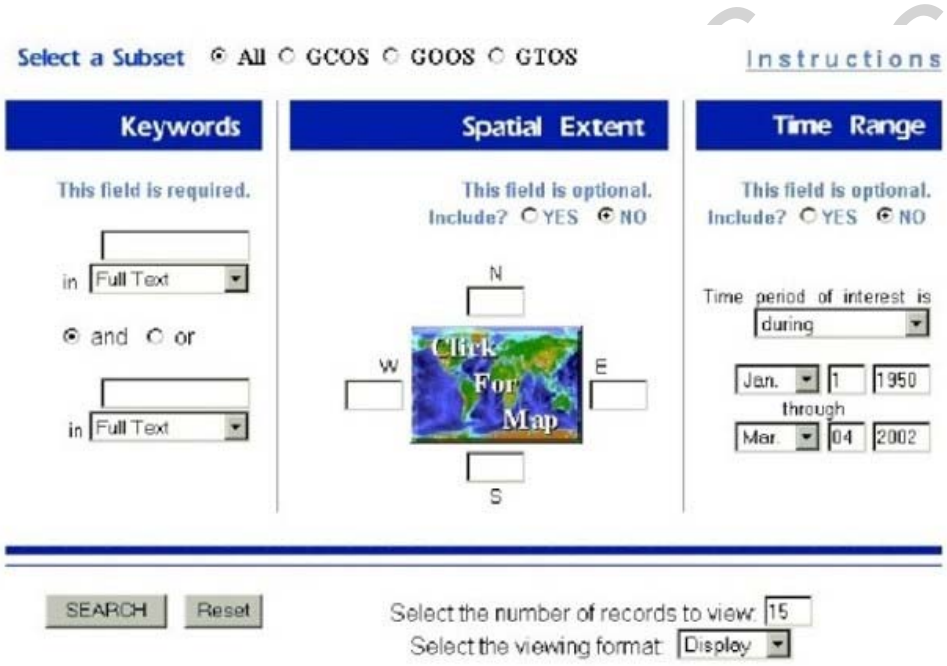

Figure 2. The GOSIC Data Set Registry search interface. Users may search the registry using keywords only, or restrict their search to a particular geographical location or time period. This search facility may be found at: http://gcmd.gsfc.nasa.gov/Data/portals/gosic/freetext/ft_search.html. Alternatively, the Hierarchical Keyword Search (http://gcmd.nasa.gov/Data/portals/gosic/) may be used to rapidly search for all data sets associated with given keywords.

3. Study the detailed end-to-end systems and link via the data flow diagrams. Data flow diagrams are used to document the components of the operational observing systems (Figure 3). The diagrams identify the participating centres and show the flow of the data from acquisition to final archival. The diagrams have hotlinks that allow the user to connect to the centres, view the data, products, and services available, and download or order data or products.

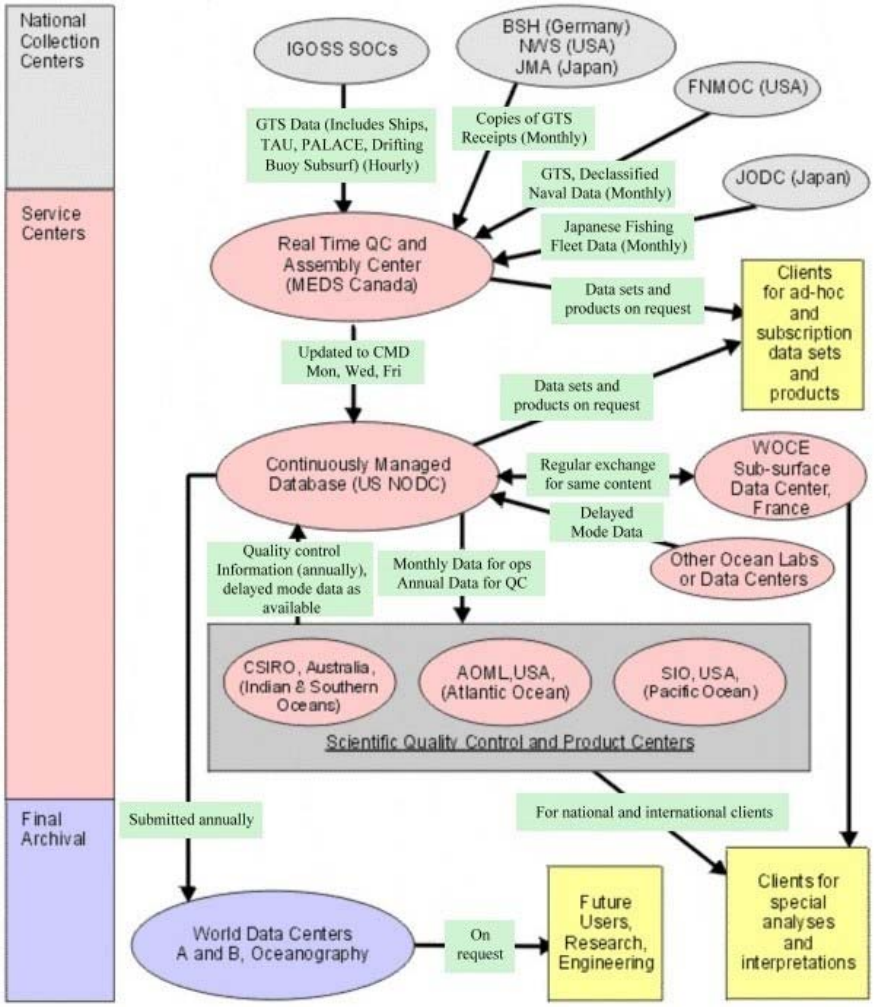

Figure 3. The data flow diagram for the Global Temperature-Salinity Profile Program (GTSPP) of GOOS, reproduced from http://www.gos.udel.edu/goos/GTSPP_dflow.htm. When accessed on the GOSIC website, users may click on each part of the diagram to access more information or to link to the GTSPP Data Centres. A full list of data flow diagrams may be found at http://www.gos.udel.edu/ios/G3OS_data_flows.htm.

-
-
-

Bibliography

Clark, M.J. (2002) 6.64.4 International Cooperation for Data Acquisition and Use. *In* "Encyclopaedia of Life Support Systems." EOLSS Publishing Company Limited, Paris. [Provides an outline of the importance of cooperation as aimed for by the global data networks introduced in this article]

GCOS (2001a) website http://www.wmo.ch/web/gcos/ [Website of the Global Climate Observing System, part of the G3OS]

GCOS (2001b), GCOS Implementation Strategy: Implementing GCOS in the new millennium, June 2001, GCOS-67 (WMO/TD No. 1072), http://www.wmo.ch/web/gcos/Publications/gcos-67.pdf [Illustrates the objectives of GCOS and the multi-disciplinary approach taken to provide comprehensive information on the whole climate system]

GCOS/GTOS (1997) GHOST - Global Hierarchical Observing Strategy (Brochure), March 1997, GCOS-33, http://www.wmo.ch/web/gcos/images/brochure/ghost.pdf [General overview of GHOST, which is designed to be an optimal sampling system]

GLOSS (2001) website, http://www.pol.ac.uk/psmsl/programmes/gloss.info.html [Website of the Global Sea-Level Observing System]

GOOS (1998) Strategic Plan and Principles for the Global Ocean Observing System, Version 1.0, Paris, January 1998, GOOS report No. 41, IOC/INF-1091, UNESCO, http://ioc.unesco.org/goos/docs/GOOS_041_GOOS_str_pl.pdf [The strategy of the Global Ocean Observing System]

GOOS (2001) website, http://ioc.unesco.org/goos/ [Website of the Global Ocean Observing System, part of the G3OS]

GTOS (2001a) GTOS Regional Implementation Plan for Southern Africa, February 2001, GTOS-21, http://www.fao.org/gtos/doc/pub21.pdf [An example of how GTOS is implemented, for Southern Africa]

GTOS (2001b) website, http://www.fao.org/gtos/ [Website of the Global Terrestrial Observing System, part of the G3OS]

GTOS (2002) Global Terrestrial Observing System Biennial Report 2001-2002, http://www.fao.org/gtos/doc/pub30.pdf [Updates on the development of GTOS]

GTSPP (2001) website, http://www.nodc.noaa.gov/GTSPP/gtspp-home.html [Global Temperature-Salinity Profile Program]

GoDAE (2001) website, http://www.bom.gov.au/bmrc/ocean/GODAE/ [Website of the Global Data Assimilation Experiment, part of the Ocean Observation Panel for Climate]

GOSIC (2001) website, http://www.gos.udel.edu/ [Website of the Global Observing System Information Centre: illustrates the four methods provided to navigate the G3OS]

GRID (1992), GRID's Data Release Policy, November 1992, United Nations Environment Programme, http://www.unep.org/unep/eia/ein/grid/datpol.htm [Provides information on dataset classification and the release policy of GRID]

GRID (1996) website, http://www.unep.org/newdraft/unep/eia/ein/grid/ [Overview of the Global Resource Information Database, including the three long-term objectives]

ICSU (1996) ICSU Guide to the World Data Centre System, ICSU Panel on World Data Centres, April 1996, http://www.ngdc.noaa.gov/wdc/guide/wdcguide.html [In particular, includes the mission statement of the World Data Centre system: "Data constitute the raw material of scientific understanding…"]

IGOS (2002) Integrated Global Observing Strategy: For the Monitoring of our Environment from Space and from Earth, http://ioc.unesco.org/igospartners/IGOSbrochure2002.pdf [Provides a general overview of IGOS, which includes the G3OS]

UNEP (1992) Agenda 21: Environment and Development, United Nations Environment Programme, http://www.unep.org/Documents/Default.asp?DocumentID=52 [See Chapter 40. Illustrates the importance of international cooperation in providing information for decision making]

UNEP (1972) Report of the United Nations Conference on the Human Environment, United Nations Environment Programme, 1972, http://www.unep.org/Documents/Default.asp?DocumentID=97 [Contains the action plan of Earthwatch, which includes the two functions "monitoring" and "information exchange"]

Biographical Sketch

Matthew D. Wilson is currently based in the School of Geographical Sciences, University of Bristol, and is conducting research into the prediction of flood inundation. He completed his Ph.D. entitled "Evaluating the effect of data and data uncertainty on predictions of flood inundation" in the School of Geography, University of Southampton, in 2003. He obtained his degree in Environmental Sciences from the University of Southampton in 1998. His research interests are in spatially-distributed dynamic modeling, the use of remotely sensed data in modeling and the assessment of the effect of data uncertainty on model predictions.