ICSU World Data System (WDS)

Members’ Forum - PUBLIC EVENT

Academy of Sciences located in Taipei, Taipei
15:30–17:30, 28 October 2012
Room 101, Institute of Information Science (IIS)

Programme

Part 1

Poster Session: 15:30–15:45

1. The International Mathematical Union - IMU
   O. Caprotti
   (IMU: CEIC)

2. Data Services for Long Tail Science at the Integrated Earth Data Applications (IEDA) Data Facility
   K. Lehnert
   (IEDA)

3. Present Status of WDC for Geomagnetism, Kyoto as a Member of WDS
   H. Toh, et al.
   (WDC-Geomagnetism, Kyoto)

Talk Session #1: 15:45–16:15

Chair: M. Mokrane (WDS-IPO)

15:45–15:50 Welcome
   B. Minster
   (WDS-SC)

15:50–15:55 Proceedings of 1st WDS Conference
   T. Watanabe
   (WDS-IPO)

15:55–16:05 The Polar Data Center, in the National Institute of Polar Research, Japan
   M. Kanao, et al.
   (NIPR)

16:05–16:15 PANGAEA®—Data Publisher for Earth & Environmental Science
   M. Diepenbroek
   (MARUM)

Demonstration: 16:15–16:45

16:15–16:45 GeoMapApp
   K. Lehnert
   (IEDA)
Part 2

Talk Session #2: 16:45–17:30

16:45–16:55  Open Access and Data Usability  
S. Graves  
(GHRC)

16:55–17:05  Progress in Establishment of WDS-linked Data Centers in Africa  
W. Hugo  
(SAEON)
D. Selematsela*  
(WDC-BHH)

17:05–17:15  International Networking of the Data Centre  
F. Hernandez  
(Flanders Marine Institute)

17:15–17:25  IVOA, Astronomy On-line Resources and CDS  
F. Genova  
(CDS/IVOA)

17:25–17:30  Closing Remarks  
B. Minster  
(WDS-SC)

Date for the diary: 16:30–17:30 Wednesday, 31 October

WDS Town Hall Meeting - PUBLIC EVENT

Hall, 2nd Conference Room, Main CODATA Conference Venue
Meet members of the WDS Scientific Committee in an informal environment

*Presenting author
Confirmed Participants

WDS Scientific Committee (SC)
- Bernard Minster  Chair WDS-SC
- Michael Diepenbroek  WDS-SC / Center for Marine Environmental Sciences (MARUM)
- Françoise Genova  WDS-SC / Centre de Donnees astronomiques de Strasbourg (CDS) / International Virtual Observatory Alliance (IVOA)
- Wim Hugo  South African Environmental Observation Network (SAEON)

WDS International Programme Office (IPO)
- Mustapha Mokrane  WDS-IPO Executive Director
- Takashi Watanabe  WDS-IPO Senior Advisor / World Date Centre (WDC) for Cosmic Rays
- Rorie Edmunds  WDS-IPO Programme Officer

WDS Members (Current & ‘In Progress’)
- Olga Caprotti  International Mathematical Union (IMU), Committee on Electronic Information and Communication (CEIC)
- Sara Graves  Global Hydrology Resource Center (GHRC)
- Francisco Hernandez  Flanders Marine Institute, Data Centre
- Masaki Kanao  National Institute of Polar Research (NIPR)
- Yukinobu Koyama  WDC-Geomagnetism, Kyoto
- Kerstin Lehnert  Integrated Earth Data Applications
- Masahito Nosé  WDC-Geomagnetism, Kyoto
- Daisy Selematsela  WDC-Biodiversity and Human Health (BHH)
- Kwang-Tsao Shao  Fish Database of Taiwan
- Vladimir Tikunov  Data Centre for Geography, Moscow
- Hiroaki Toh  WDC-Geomagnetism, Kyoto

Other
- Richard Hartshorn  International Union of Pure and Applied Chemistry (IUPAC)

†This is a public event. Other participants are expected to attend.
Directions to WDS Members’ Forum from Main Conference Venue
(Approx. 3 min on foot)

Key
Red circle: Main CODATA Conference Venue—Building for Humanities and Social Sciences
Green circle: WDS Members’ Forum Venue—Room 101, Institute of Information Science
Black dashed line: Route between venues
Demonstration: GeoMapApp

K. Lehnert (Integrated Earth Data Applications)

GeoMapApp is a free map-based data discovery, visualization, and analysis tool that is increasingly used in the earth sciences community. It provides direct quantitative access to a wide range of geoscience data sets from a variety of disciplines, and enables users to import their own geospatial data in several common formats. The built-in functionality of GeoMapApp enables users to interact with and analyze geospatial data, and to produce high quality maps and images in a variety of formats. This presentation will demonstrate GeoMapApp's functionality and how you can use it in your own research.

Abstracts: Talks

PANGAEA®—Data Publisher for Earth & Environmental Science

M. Diepenbroek (Center for Marine Environmental Sciences, University of Bremen)

PANGAEA is an information system for acquisition, processing, long term storage, and publication of georeferenced data related to earth science fields. Essential services supplied by PANGAEA are project data management, as well as the organization and implementation of data infrastructures. Storing about half a million datasets from all fields of geosciences, it belongs to the largest archive for observational earth science data. Standard conform interfaces (ISO, OGC, W3C, and OAI) enable access from a variety of data and information portals, including the search engine of PANGAEA itself (www.pangaea.de) and, for example, Google. All datasets in PANGAEA are citable, fully documented, and can be referenced via a persistent identifier (a digital object identifier or DOI)—a premise for data publication. Together with other International Council for Science (ICSU) World Data Centers (WDCs) and the Technical Information Library in Germany, PANGAEA had a share in the implementation of a DOI-based registry for scientific data, which now is supported by a worldwide consortium of libraries (www.datacite.org).

Seeing the overall positive impact on the quality and availability of scientific data, the PANGAEA group has been offering publication services for several years. They were the main initiators of the "Earth System Science Journal" at Copernicus (http://earth-system-science-data.net/), which is used for standalone peer-reviewed data publications. A further milestone was building up strong cooperation with commercial publishers such as Elsevier, Springer, Wiley, AGU, and Nature. A common web service enables supplementary data in PANGAEA to be directly referenced from an articles abstract page (e.g., Science Direct). The next step with commercial publishers is to further integrate the editorial process for the publication of supplementary data with the publication procedures on the journal side, which practically means an extension of the peer review and a synchronization of editorial processes.
PANGAEA is jointly operated as a long-term facility by MARUM at the University of Bremen and the Alfred Wegener Institute for Polar and Marine Research. More than 80% of the funding results from project data management and the implementation of spatial data infrastructures (more than 160 national and international projects over the last 15 years—www.pangaea.de/projects).

IVOA, Astronomy On-line Resources and CDS

F. Genova (Centre de Donnees astronomiques de Strasbourg/International Virtual Observatory Alliance)

The concept of Virtual Observatory (VO) emerged at the turn of the century. It aims at giving seamless access to the wealth of astronomical resources available on-line, and its infrastructure is composed of interoperability standards and access tools. It has been seen from the beginning as a global initiative, and the International Virtual Observatory Alliance (IVOA) was founded in 2002 as an alliance of VO national projects, also including the European VO among the founding parents. The IVOA coordinates VO activities around the world, and in particular, the technical work of establishing the standards. Participation is on a best effort basis.

Representatives of the VO projects form the Executive Board, and Working Groups are in charge of the technical work. There are currently seven Working Groups covering different aspects of the standards—Applications, Data Access Layer, Data Modelling, Grid & Web Services, Semantics, VO Event, Resource Registry; four Interest Groups—Theory, OGF Astro-RG, Data Curation & Preservation, Knowledge Discovery in Databases; two Standing Committees on Standards & Processes and Science Priorities; and a Liaison Committee (dealing with Liaison with other disciplines). Technical activities are coordinated by the Technical Coordination Group.

IVOA standardization procedure is derived from the W3C one. Most of the basic standards have now been defined, and activity is shifting towards standards devoted to more advanced usage and updates of existing standards to take into account feedback from implementation and usage and the evolution of astronomy.

I will briefly describe IVOA, which just joined WDS as a Network Member and is willing to share lessons learnt and practices. The VO concept is to allow any interested data centre or service to join, and the ‘VO layer’ comes on top of the existing data holdings. I will discuss the large diversity of astronomical data centre and services, from archives of large space and ground-based telescopes to focused services provided by scientific teams, and give the example of Strasbourg astronomical data centre (CDS), which provides added-value services that are widely used by the astronomical community. CDS, which is celebrating its 40th
birthday this year—quite a milestone for a data centre!—is also a WDS Member, and is one of the key actors of the VO development, providing several of its building blocks.

**Open Access and Data Usability**  
*S. Graves (Global Hydrology Resource Center)*

Can researchers and others only ‘use’ data if the data provider allows for open access to the data or are there ways to increase data usability with ‘open data use’ tools? Traditionally, data centres have been concerned with the functions of ingesting, organizing, and storing data for archival purposes. Creating metadata, of varying complexity and completeness, provided for easier access to data through user interfaces of various types, assuming the data was open for access. This presentation will address the challenge of providing mechanisms for better data usability in different types of data centres.

**International Networking of the Data Centre**  
*F. Hernandez (Flanders Marine Institute, Data Centre)*

Since the establishment of the Flanders Marine Institute in 2000, the Flanders Marine Data Centre (VMDC) has been closely involved in international data management networks. Scientific research, and especially marine research, is highly dependent on international cooperation. Since time at sea is expensive and the seas are limitless, and thanks to a multidisciplinary approach, the marine research community has built strong international partnerships for several decades.

Facilitated by their good contacts with Flemish scientists, the VMDC started a national oceanographic data centre (NODC) within the scope of the International Data and Information Exchange programme (IODE) of UNESCO’s International Oceanographic Committee.

Each NODC within this network supports local marine scientists through data management activities and takes part in the exchange of data, information, and expertise with the other nodes. These activities are still at the basis of the data centre’s and the VLIZ library’s data management services: the documentation, archiving, integration, and distribution of measurement data, as well as the disclosure of scientific publications. Our scope is the output of Flemish marine research groups, the Southern Bight of the North Sea, including the Flemish coast and tidal rivers.

In addition, cooperation within the framework of the IOC has resulted in hosting the IODE project office, the development of an ASFA input centre, dozens of participations in IODE workshops, the organisation of training courses, bilateral cooperation with Bulgaria and Ukraine, and the development of a real-time sea level station monitoring system as part of the GLOSS programme.
The sea level station monitoring system registers the data and status of 616 tide gauges, managed by 108 institutions, in real time.

Within the scope of the MARBEF ‘network of excellence’ project we were invited by NIOO-CHEME and Ghent University to take care of the data management package. The data systems created over a period of five years, afterwards developed into the European node of the Ocean Biogeographic Information System (EurOBIS) and the World Register of Marine Species (WoRMS).

EurOBIS has collected 13 million observations supplied by 382 data suppliers (datasets).
WoRMS describes 210,000 biological species (436,000 names) and is managed by 270 taxonomic experts in 38 countries.

At the request of the Catholic University of Leuven, we have developed, as part of the ENCORA network, Coastal WIKI, a system which has been supplemented with WIKI contexts regarding biodiversity, ecotoxicology and alien species.

- The WIKI currently contains over 1500 informative pages, validated by dozens of European experts.

Within the scope of the European Marine Observation and Data Network (EMODnet) and funded by DG-MARE, VLIZ has been appointed to develop the biological data portal. This collaboration between the SeaDataNet, MARS, OBIS, and IODE networks is entering its second phase in 2012.

In 2012 the ESFRI LifeWatch project is starting the development of a large-scale, distributed e-infrastructure for the study of biodiversity and ecosystems. VLIZ coordinates the Belgian participation of INBO, UCL, RBINS and BBPf, builds a marine observatory and is responsible for the creation of the taxonomic backbone.
Progress in Establishment of WDS-linked Data Centers in Africa

W. Hugo (South African Environmental Observation Network) and D. Selematsela* (World Data Centre for Biodiversity and Human Health)

ICSU awarded a grant to the Regional Office in Africa in May 2012 to assist with the establishment of a networked data centre for socio-economic and related subjects serving African countries (NDCSEA). Previously, ICSU and the United States Geological Survey’s National Biological Information Infrastructure (NBII), through the WDC for Biodiversity and Ecology grants awarded in 2009/10, assisted with the establishment of a prototype data centre for biodiversity and human health in Africa.

The South African Environmental Observation Network, as part of the National Research Foundation, is assisting with the implementation of both of these initiatives. The technical platform for these centres has now been firmly established at the South African Centre for High Performance Computing, which provides these centres with high-performance computing, petabyte range storage facilities, and fast connectivity through the South African National Research Network. In addition, the centres make use of a shared software platform that is maintained and continuously improved by participating initiatives. This software platform is available as a public good, is open source, and focuses on the main technical functions required of a well-managed data centre.

These centres use of a different establishment and operations model compared with the majority of WDS members. This model relies heavily on a shared infrastructure to meet the technical aspects of WDS criteria, and then seeks and promotes opportunities to establish the governance and scientific oversight components and content provision arrangements that constitute a viable service.

In respect of the WDS membership of these centres, some challenges remain:

1. The centres operate in a distributed and virtualized manner due to being distributed across national boundaries. No ready model of a hosting organization exists that is equivalent to such a centre, nor is there clarity on the management model.
2. Funding of the longer-term scientific oversight and governance of the centres needs to be confirmed. The best model remains, in our view, to align the grant funding that may be available from national and international institutions to identify and compensate individuals active in the appropriate domains.
3. Content provision in a virtual centre, and its quality assurance processes require new thinking and approaches to allow community-based review of data, given that full-time personnel to perform this function are unlikely to be available.

The ICSU funding for establishment of NDCSEA recognizes several of these impediments, and some of the funding will go towards workshops and expert inputs in respect of the
questions of governance, management, quality assurance, and funding that have been highlighted. The output from this effort is likely to be useful for the future establishment of other centres of a similar nature.

The Polar Data Center, in the National Institute of Polar Research, Japan
M. Kanao*, M. Okada, and A. Kadokura (National Institute of Polar Research)

Recent, rapid, technological improvement and development of Earth observation by satellites and ground observation networks both in the Arctic and the Antarctic has led to a large quantity of polar observation data being collected every day. The processing and utilization of these data is an important issue to promote polar science. At the 22nd Antarctic Treaty Consultative Meeting in 1998, affiliate countries were obliged to ensure that scientific data collected from Antarctic programs could be freely exchanged and used. Following Article No. III.1.c of the 1998 Antarctic Treaty, each country is required to establish a National Antarctic Data Centre (NADC) and to properly provide the data collected from involved scientists.

The Polar Data Center (PDC; http://www.nipr.ac.jp/english/polar-information01.html) of the National Institute of Polar Research (NIPR) has served as the Japanese NADC with a strong relationship with the Scientific Committee on Antarctic Research (SCAR) under ICSU. Our mission is twofold: scientific data management and management of information infrastructure. PDC established a data policy in February 2007, based on the requirements of the Standing Committee on Antarctic Data Management of SCAR. This contributed to the subsequent SCAR Data and Information Management Strategy (de Bruin & Finney, 2011). During the International Polar Year (IPY), PDC compiled much of the polar data from the endorsed projects involving Japanese activities.

PDC has the significant task to archive and deliver the digital data obtained from the polar region. Summary information of all the archived data (metadata) is available to the polar science community, as well as more general interests. The compiled metadata describe all types of observed/collected science disciplines (space and upper atmospheric sciences, meteorology and glaciology, geoscience and bioscience) from both long- and short-term projects in the Arctic and Antarctic, particularly data collected by the Japanese Antarctic Research Expedition. In the science meta-database provided by PDC, a total of 150 metadata records had been accumulated as of October 2011, including metadata from IPY endorsed projects (http://scidbase.nipr.ac.jp/). A new content management system for providing the metadata has been in place since April 2011.

In addition to the polar communities, strong links have been formed with other science bodies of ICSU, such as Committee on Data for Science and Technology and the new WDS. Regarding aurora data, in particular, we have administered the WDC for Aurora since 1981.
Now the Aurora Data Center in PDC has taken over the tasks of the WDC for Aurora. The Aurora Data Center (http://polaris.nipr.ac.jp/~aurora/) is responsible for data archiving and dissemination of all-sky camera observations, and other aurora-related observations, especially in the Antarctic region.

Beyond these obligations, PDC is responsible for the archiving and analysis of Earth observation satellite data (from the Polar Operational Environmental Satellite of the National Oceanic and Atmospheric Administration), seismological data (short-period and broadband seismometers) and crustal movement data (obtained from the Global Positioning System and Very Long Baseline Interferometry) around the Syowa Station (SYO; 69S, 30E), East Antarctica. Finally, PDC manages various information infrastructures such as (1) a mainframe and a workstation system, (2) the network systems of domestic and related facilities (e.g., SYO), and (3) the reception facilities for Earth observation satellite data.

The new WDS aims at a transition from existing standalone WDCs and individual services to a common, globally interoperable, distributed data system that incorporates emerging technologies and new scientific data activities, including polar data as a legacy of the IPY. The WDS policy of full and open access to data will benefit the international scientific community and ultimately society at large. Many concepts of data publication and data citation should be adopted and promoted by the WDS to facilitate timely release of data. The WDS has agreed to take the necessary steps to archive IPY data and to work with the Polar Information Commons (PIC) to preserve, curate, and add value to data in the PIC cloud in order to preserve the legacy of IPY data. Links among these data bodies need to be strengthened in the post-IPY era. PDC/NIPR hopes contribute to such WDS activities in the future.

**Abstracts: Posters**

**The International Mathematical Union – IMU**

O. Caprotti (International Mathematical Union, Committee on Electronic Information and Communication)

IMU, a member of the International Council for Science, is a nongovernmental, nonprofit, scientific organization of countries with a mathematical culture. At present, IMU includes 69 Members, 8 Associates, and 4 Affiliate Members.

IMU's objectives are to build, support, and strengthen international cooperation in mathematics through meetings, such as the International Congress of Mathematicians, and other activities for the development of mathematical sciences.
Data Services for Long Tail Science at the Integrated Earth Data Applications (IEDA) Data Facility

K. Lehnert (Integrated Earth Data Applications)

IEDA is a US-based data facility funded through a contract with the US National Science Foundation to operate data systems and data services for solid earth geoscience data. IEDA has developed a comprehensive suite of data services that are designed to address the concerns and needs of investigators, especially researchers working in the 'Long Tail of Science' (Heidorn, 2008). IEDA provides a data publication service, registering datasets with DOI to ensure their proper citation and attribution. IEDA works with publishers on advanced linkages between datasets in the IEDA repository and scientific online articles to facilitate access to the data, enhance their visibility, and augment their use and citation. IEDA also developed a comprehensive investigator support that includes tools, tutorials, and virtual or face-to-face workshops that guide and assist investigators with data management planning, data submission, and data documentation. A critical aspect of IEDA’s concept has been the disciplinary expertise within the team and its strong liaison with the science community, as well as community-based governance. These have been fundamental to gain the trust and support of the scientists and have led to significantly improved data preservation and access in the communities served by IEDA.

Present Status of WDC for Geomagnetism, Kyoto as a Member of WDS

H. Toh*, M. Nosé*, Y. Koyama*, M. Takeda, Y. Odagi, and T. Iyemori (World Data Center for Geomagnetism, Kyoto)

WDC for Geomagnetism, Kyoto is unique in the sense that it is part of an international organization (i.e., ICSU/WDS) and yet is an institution established within the Graduate School of Science, Kyoto University. As the latter, our official name is the Data Analysis Center for Geomagnetism and Space Magnetism (DACGSM), and in this capacity we are in charge of student education, as well as conducting scientific research on geomagnetism and space magnetism. As a part of ICSU-WDS, it is our task to collect geomagnetic data from all over the globe and make them open to the public, including various scientific communities. Our mission here in Kyoto therefore can be said three-fold, namely, providing ‘education’, ‘research’, and ‘data services’.

Among those activities ongoing in Kyoto, those classified as fitting to our role as WDC for Geomagnetism, Kyoto, are our development of the Inter-university Upper atmosphere Global Observation NETwork (IUGONET), derivation of auroral electroject (AE) and storm-time disturbance ($D_{st}$) indices, and acquisition of new geomagnetic data in logistically remote locations. IUGONET is a project among universities and research institutions—some of which host WDS-related data centers—that has the present aim of constructing a metadata database pertinent to aeronomy. The calculation and publication of AE and $D_{st}$
indices are two of the main tasks of WDC for Geomagnetism, Kyoto. Those indices are derived from raw vector geomagnetic time series on the Earth and are provided in three different forms: ‘real-time’, ‘provisional’, and ‘definite’. All of them are now published on our website (http://wdc.kugi.kyoto-u.ac.jp/index.html) and are accessible via the Internet. The acquisition of new geomagnetic data is a common task for geomagnetism WDCs worldwide. However, our efforts include geomagnetic data acquisition in logistically difficult regions such as the deep seafloor, which may be an activity individual to WDC for Geomagnetism, Kyoto.