

# CODATA

## Newsletter 100

Committee on Data for Science and Technology

### CODATA welcomes the U.S. National Archives and Records Administration (NARA) as a new Supporting Organisation

May 2011

Preserving US Government Records of the Past, Present and future..... p.1&2

ICSU World Data System..... p.3

Why join WDS ?..... p.3

Artificial Intelligence in the Earth's Magnetic Field Study INTERMAGNET Russian Segment..... p.4

Forthcoming Meetings..... p.4



National Archives Building, Washington, DC

If you want to publish a short article in a CODATA newsletter please contact Cécile Carbonell at [codata@dial.oleane.com](mailto:codata@dial.oleane.com)

The U.S. National Archives and Records Administration (NARA) is the Federal government agency responsible for preserving and making available to the public

#### PRESERVING UNITED STATES GOVERNMENT RECORDS OF THE PAST, PRESENT AND FUTURE

By Vivek Navale and Ross Cameron (NARA)

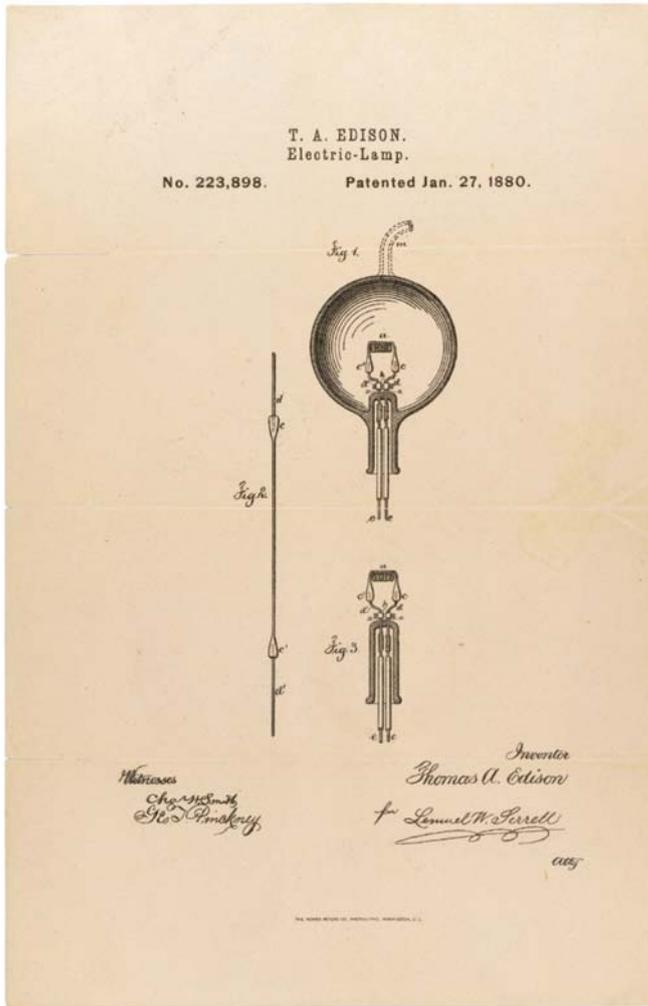
the records created in the course of business conducted by the United States Federal government. The National Archives was established in 1934, but its major holdings date back to 1775. They capture the sweep of the past: slave ship manifests and the Emancipation Proclamation; captured German records and the Japanese surrender documents from World War II; jour-

nals of polar expeditions and photographs of Dust Bowl farmers; Indian treaties making transitory promises; and a richly bound document bearing the bold signature «Bonaparte»—the Louisiana Purchase Treaty that doubled the territory of the young republic.

NARA keeps only those Federal records that are judged to have continuing value—about 2 to 5 percent of those generated in any given year. They now add up to a formidable number, diverse in form as well as in content. There .../...

**CODATA Secretariat**  
5 rue Auguste Vacquerie,  
75016 Paris, France  
Tel: +33 1 45250496  
Email: [codata@dial.oleane.com](mailto:codata@dial.oleane.com)  
Website: <http://www.codata.org>

Coordinator : Cécile Carbonell  
Editor : Gordon Wood  
Design and Layout:  
JF Macaigne for Eo Conseil



Patents – Edison's Electric Lamp housed in NARA

is the strategic response to the challenge of preserving, managing, and providing access to electronic records. ERA will keep essential electronic Federal records retrievable, readable, and authentic for as long as they remain valuable—whether that is a few years or a few hundred years. NARA currently has 97.4 terabytes of data in billions of files from over 200 departments and agencies containing data from 1819-2010. There are a variety of record formats, including: data files and databases, electronic documents, e-mail messages with attachments, scanned images of textual records, Portable Document Format (PDF) records, digital photographic records, geospatial data records and web content records. The records are received from agencies on a variety of media types including CDs, DVDs, DLTs, 3480-class cartridges, open-reel tapes, external and internal hard drives, diskettes, mini-cartridges and even keypunch cards as well as by electronic transfer and web downloads.

NARA actively collaborates with several federal agency partners, including the United States Geological Survey (USGS), National Oceanic and Atmospheric Administration (NOAA) and National Aeronautics and Space Administration (NASA) for development and implementation of long-term preservation and access strategies of scientific electronic records. Recently, in collaboration with various federal partners, NARA developed a discussion guide based on the ISO Open Archival Information System (OAIS) Standard Reference Model, that can help Federal agencies identify and determine high-level data management policies, procedures, and processes needed to ensure long term preservation and access of digital assets for all stakeholders. A complete discussion is available in NARA's Toolkit for Managing Electronic Records at <http://www.archives.gov/records-mgmt/toolkit/>

More recently, as a supporting member of the International Council of Science: Committee on Data for Science and Technology (CODATA), NARA staff is also involved in creating an inventory of scientific data that are at risk, and whose unique scientific information is in danger of being lost to posterity. A discussion of the task group and members can be found at <http://www.codata.org/taskgroups/TGdataatrisk/index.html>

are approximately 9 billion pages of textual records; 7.2 million maps, charts, and architectural drawings; more than 20 million still photographs; billions of electronic files; and more than 365,000 reels of film and 110,000 videotapes. All of these materials are preserved because they are important to the workings of Government, have long-term research worth, or provide information of value to citizens. NARA not only preserves these records, but also makes them available to anyone who is conducting research or who is simply interested in viewing these records. Records help us claim our rights and entitlements, hold our elected officials accountable for their actions, and document our history as a nation. In short, NARA ensures continuing access to the essential documentation of the rights of American citizens and the actions of their Government.

In addition, NARA must also manage the rapidly growing number of electronic Government records. Now being developed, the Electronic Records Archives (ERA)

## THE 56TH CODATA EXECUTIVE COMMITTEE MEETING TOOK PLACE IN PARIS ON 26TH AND 27TH FEBRUARY 2011



From left to right : Hiro-o Hamaguchi, Mark Thorley, Françoise Genova, Antony Cooper, Kathleen Cass, John Broome, Huadong Guo, Sara Graves, Niv Ahituv, Robert Chen, Dora Ann Lange Canhos and Michel Sabourin.  
(Absent : Krishan Lal, Fedor Kuznetsov, Takashi Gojobori.)

## ICSU WORLD DATA SYSTEM (WDS)

WDS has been created through a decision of the 29th General Assembly of the International Council for Science (ICSU). It builds on the 50-year legacy of the ICSU World Data Centre system (WDC) and the ICSU Federation of Astronomical and Geophysical data-analysis Services. The WDS concept aims at a transition from existing stand-alone WDCs and individual Services to a common globally interoperable distributed data system that incorporates emerging technologies and new scientific data activities. Over 100 organizations have responded to a survey of organizations that manage scientific data and data services, showing their interest in joining the system. Candidatures are evaluated by the WDS Scientific Committee on criteria such as access to high quality data, data stewardship, and participation in broad harmonization and interoperability efforts. For more information on the system see <http://www.icsu-wds.org> and read the article in the Newsletter on "Why Join WDS ?"

CODATA is delighted to welcome the WDS as a Co-opted Member of its community !

### WHY JOIN WDS ?



*ICSU WDS meeting in Stellenbosch (South Africa) October 2010.*

#### **For improving your data facility's local and international recognition**

- Joining this international and interdisciplinary 'community of excellence' under the auspices of ICSU will heighten the profile of your organization without compromising its own identity, gaining worldwide exposure to potential users of the data, products and services offered. ICSU and its member Unions provide prestigious auspices under which to operate. This is because ICSU members—121 National Members mostly represented by scientific academies covering 141 countries and 30 Scientific Unions (and their Associations)—provide the main forum for international scientific collaboration and exchange of information in the major areas of science.

- Improved visibility and engagement in international activities, programs, and projects will improve your facility's reputation and peer recognition and in turn might also improve its standing within the host organization which will in many cases enhance funding opportunities: WDS membership will be an effective means of bringing international recognition of your data facility to the attention of decision-making people within the host institution (i.e. "WDS" is a "good name"). In addition, the non-governmental nature of ICSU allows WDS members to interact with organizations which otherwise might not be possible.

#### **For adhering to and promoting open data sharing principles for the benefit of all**

- A WDS Member adheres to the WDS Data Policy which is built around the concepts of full and open access, and of data sharing provided at no cost or for no more than the cost of reproduction (i.e. "COFUR": Cost of Fulfilling User Request). The full benefits of data sharing are only

realized if someone initiates the sharing process as a matter of principle. In this way WDS members initiate a "win-win" process that benefits not only themselves but also their partners in the scientific community and the scientific community at large, which can be particularly useful when some are still reluctant to make their data available to the international community.

#### **For displaying your commitment to data and service quality and long term data curation**

- Joining the WDS means becoming part of a community recognized to have met accreditation criteria on scientific relevance, governance, data management, technical infrastructure and security, as well as open and equitable access to quality-assured scientific data, data services, products and information.

#### **For improving your products and performance, as well as global data management and data exchange**

- WDS opens worldwide prospects for collaborations that enhance products, services, or performance (e.g. easier sharing of data and information with other systems). This is especially the case when members of existing disciplinary groupings of services/centres join collectively (e.g. geophysics, oceanography, astronomy), thereby extending the scope of whole fields of science to others.

- WDS membership provides fertile ground onto which new, emerging international data exchange and management standards and activities could naturally embed themselves.

#### **For facilitating the citation of your data**

- WDS places a strong focus on data publication services as a new component of scholarly publishing. Data citability has an overall positive impact on the quality and availability of scientific data and is a major incentive for data providers. Data facilities frequently find that their data are used in scientific papers without any citation or acknowledgement of the source. WDS data publishing services will provide these facilities with an important avenue to highlight their contributions to the scientific community.

#### **For increasing your disciplinary and interdisciplinary user base**

- WDS data holdings, WDS services output, and WDS-provided information will be easily searchable, rapidly accessible, and satisfy national and international standards for data and data services. Enhanced interoperability through international standards could in turn be implemented at other levels, internally within the host organization and externally with its partners.

## ARTIFICIAL INTELLIGENCE IN THE EARTH'S MAGNETIC FIELD STUDY INTERMAGNET RUSSIAN SEGMENT

*Anatoly Soloviev, Chief of Laboratory of Mathematical Analysis of Geomagnetic Data  
Institution of the Russian Academy of Sciences Geophysical Center RAS*

The objectives of the CODATA Task Group "Earth and Space Science Data Interoperability" (TG) are to promote the establishment and usage of modern and open data access standards and to help develop efficient ways to achieve inter-operability. The TG promotes free access to metadata exchange and integration worldwide, including goal-oriented co-operation. Concrete applications of this approach promoted by the TG include the World Atlas of the Earth's Magnetic Field from 1500-2010 and the launch of a new geomagnetic data node for the CIS countries.

Geomagnetic field study is one of the fundamental problems of geophysics. The Earth's magnetic field is the most important source of information about the inner structure of the Earth and its evolutionary processes. The most accurate and rapid data on the geomagnetic field are provided by INTERMAGNET (International Real-time Magnetic Observatory Network), which has been operating now for more than 20 years (<http://www.intermagnet.org/>). The main objectives of the network, which consists of more than 110 observatories all over the world, are observations, acquisition, accumulation, storage, dissemination, analysis and correction of magnetograms with minimum time delay. Huge volumes of recorded geomagnetic data lead to the necessity of fundamentally new methods of processing it. Such methods include pattern recognition and geoinformatics.

The international Workshop "Artificial Intelligence in the Earth's Magnetic Field Study. INTERMAGNET Russian Segment" was organized by the TG on 26-28 January 2011 in Uglich, Yaroslavl region, Russia. The main goal of the Workshop was to exchange knowledge and experience in the application of modern geomagnetic instruments, deployment of magnetic observatories as well as technical/scientific processing of magnetograms. The specific objective was to discuss plans for further INTERMAGNET development in Russia.

The Workshop brought together leading scientists and specialists in geomagnetism and magnetic observations

from Canada, France, Germany, Hungary, Russia, USA and Ukraine. Expert magnetologists expressed their opinions regarding deployment of future geomagnetic observatories in Russia. Manufacturers and suppliers of geomagnetic and modern computer equipment took part in the exhibition organized within the framework of the Workshop. At the Workshop there was a meeting of the institutions of the Russian Academy of Sciences (RAS) overseeing the work of the Russian observatories of INTERMAGNET standard. On the underpinning of several institutions, the installation of new INTERMAGNET observatories was planned. As a result of this meeting the following important steps for cooperation in creating and developing the INTERMAGNET Russian Segment were taken:

1. An agreement for joint work on the project of the INTERMAGNET Russian Segment deployment was concluded;
2. A decision to establish a Working group within the National Geophysical Committee of the Russian Federation for creation of the INTERMAGNET Russian Segment was taken;
3. An agreement to establish a national geomagnetic node, based at the Geophysical Center of RAS, for collecting and processing geomagnetic data from the INTERMAGNET Russian Segment with the perspective of becoming the Seventh INTERMAGNET GIN (Geomagnetic Information Node) was reached;
4. The deployment of an experimental training temporary INTERMAGNET observatory, based at the magnetic station "Moscow" of IZMIRAN (Pushkov Institute of Terrestrial Magnetism, Ionosphere and Radio Propagation of RAS) in Troitsk, Moscow Region, was initiated;
5. Leading geomagnetism scientists formulated specific recommendations on the deployment of new INTERMAGNET observatories in Russia.

Detailed information on the Workshop and the results obtained are available on the Web-site <http://uglich2011.gcras.ru/>.

### FORTHCOMING MEETINGS

- ICSTI Summer Conference, 7-8 June 2011 - Beijing, China (<http://www.icsti2011.org/eng/index.html>)
- 1st ICSU WDS Conference "Global Data for Global Science", 3-6 September 2011 - Kyoto, Japan (<http://wds-kyoto-2011.org/>)
- ICSU 30th General Assembly, 24th September-1st October 2011 - Rome, Italy (<http://www.icsu.org/general-assembly>)
- IRDR Conference "Disaster Risk: Integrating Science and Practice", 31st October-2nd November 2011 - Beijing, China (<http://www.irdrinternational.org/conference2011.php>)
- The PV-2011 Workshop, 15-17 November 2011 - Toulouse, France (<http://www.pv2011.com/>)
- 7th International Digital Curation Conference: "Public ? Private ? Personal ? Navigating the Open Data Landscape", 5-7 December 2011 - Bristol, UK (<http://www.dcc.ac.uk/node/9172>)
- IEA 2012, February 12-16, 2012 - Recife, Brazil (<http://www.iea2012.org/>)