CODATA
SPECIAL ISSUE

Committee on Data for Science and Technology

CODATA TASK GROUPS 2013-2014

Through its Task Groups, CODATA executes an ambitious international scientific agenda, addressing major data needs and policy issues in a broad range of subjects. These activities are selected at the biennial CODATA General Assemblies. Last November in Taipei the 28th CODATA General Assembly approved 12 Task Groups for the period 2013-2014.

We have therefore decided to dedicate a Special Issue of the Newsletter on these Task Groups.

We have invited each Task Group to consider what they see as the major scientific data challenge facing their activity and what role CODATA International could play in addressing this challenge.

March 2013

CONTENT :

If you want to publish a short article in a CODATA newsletter please contact Sally Davies at **codata.office@gmail.com**

ADVANCING INFORMATICS FOR MICROBIOLOGY
(http://www.codata.org/taskgroups/TGmicrobiology/index.html)

Juncai Ma, Co-Chair

Hideaki Sugawara, Co-Chair

Microbes, when isolated and preserved in culture collections, are associated with valuable information that should also be kept and published to promote further utilization of the material. By the end of 2012, 633 culture collections in 72 countries had registered with WFCC-MIRCEN World Data Centre for Microorganisms (WDCM), submitting 224 online catalogues. A large amount of microbial resource information, however, remains unregistered and is not available online. This may be due to the digital divide between developing and developed country's collections. Furthermore, there is no existing uniform information platform for microbial resource information and this even hinders the information sharing among developed country's collections and their users. The AIM Task group will benefit by communicating with other CODATA Task groups as well as with representatives from many countries and scientific societies to develop a solution to the challenge of the digital divide and data sharing of microbial resources information.


Kevin McCluskey, Co-Chair

In addition to the information retained by culture collections, AIM proposes to integrate with efforts to deal with huge datasets coming out of genomic, metagenomic and other "omics" projects. The AIM Task Group will apply emerging information technologies such as Cloud Computing technology, Big Data integration, and Data Mining to microbial information across multiple disciplines. Communication with experts on very large database (VLDB), high performance computing (HPC), and algorithms through CODATA conferences and other opportunities will synergistically contribute to each of these efforts.

Wu Linhuan

ANTHROPOMETRIC DATA AND ENGINEERING
(http://www.codata.org/taskgroups/TGanthro/index.html)

Daisy Veitch, Co-Chair
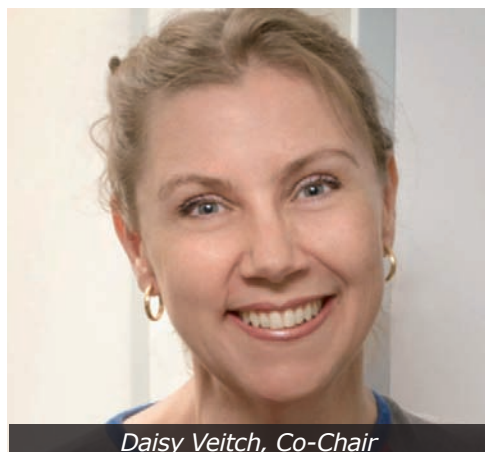
Kathleen Robinette, Co-Chair

is to find the underlying fit theory so predictions can be made for new products without extensive testing. People, especially designers, use visual information to help solve design problems. To increase functionality, anthropometric data need graphic interfaces to assist their intuitive use thus helping users make sense of data without knowing all the terminology. In addition, visualising will assist the use of 3D data by helping realise their full potential. However, comparing and searching 3D data is not straightforward.

CODATA International provides an interdisciplinary forum with potential to learn how other people approach some of these problems as sharing facilitates learning from each other.

There are several scientific challenges currently being faced. Fitmapping is the relationship between anthropometric data and products. Currently, each product has a specific relationship but the challenge

One example is a specialist meeting, wherein an invited speaker panel explores a specific issue such as the barriers faced using data to their full potential.

Daisy Veitch

DATA CITATION STANDARDS AND PRACTICES
(http://www.codata.org/taskgroups/TGdatacitation/index.html)

Jan Brase, Co-Chair

Sarah Callaghan, Co-Chair

Christine Borgman, Co-Chair

In the last decade, the amount of data created has outpaced our ability to process and analyze those data. Nevertheless, there is still a lack of consensus regarding the treatment of data, especially reference data sets in order to maintain and strengthen the scholarly record.

For the first two years, the work of the DCSP has been to collect the different best-practice examples of current data citation in a systematic way and document common methods and

workflows. A report summarizing the current state of practice and requirements will be published this Spring. Over the the next 2 years, DCSP will work on detailed guidelines for data scientists and other stakeholders to define the different proper methods and systems of referencing and citation for routine needs. In addition, DCSP will explore opportunities for developing guidelines and will work with the standards communities to see if more formal standards might be considered.

The DCSP is extremely thankful to be part of the CODATA family for this, as CODATA International provides direct contact to data professionals from various disciplines from all over the world. The Task Group's work greatly benefits from having insights into the current practices and future needs of different communities and will also benefit from having CODATA as a disseminator of its recommendations.

Jan Brase

FORTHCOMING MEETINGS

2-3 May 2013 - Torino, Italy
The 6th annual workshop on "The Organisation, Economics and Policy of Scientific Research",

6-7 May 2013 - Florence, Italy
International conference on "Framing the Digital Curation Curriculum"

DATA AT RISK

The primary objectives of Data at Risk Task Group for 2013-14 are to:

(1) Continue building an Inventory of sources of scientific data likely to be "at risk" because of their fragile, non-digital state, and

(2) Establish an international forum ("Federation") that can provide increased visibility and mutual support for as many worldwide efforts towards data rescue as can be located.

In September 2012 DARTG participated at the major UNESCO conference (Vancouver) on "The Memory of the World in the Digital Age", where it organized a session of its own on "Data At Risk".

A vision of the Federation which DARTG wants to create was discussed with UNESCO. A formal request was submitted to UNESCO to consider the sponsorship of such a Federation of worldwide efforts towards data rescue

Elizabeth Griffin, Chair

by including it within its "Memory of the World".

The major challenges being confronted are in the areas of education and advertisement of what is needed and why. That is the main reason for ensuring the proposed Federation has well-recognized patronage.

CODATA International can assist by endorsing the efforts to create that Federation and by displaying relevant pages (when ready) on its website. Linking CODATA to UNESCO is a step that could be explored, as it will not only bring CODATA's work more deeply into the realms of UNESCO (as, indeed, did DAR's participation in UNESCO's Vancouver meeting) but also furnish UNESCO an outlet into fundamental and specialist science.

Elizabeth Griffin

EARTH AND SPACE SCIENCE DATA INTEROPERABILITY

The major scientific data challenges facing the Task Group's activity are:
- the absence of a unified Global Information System (GIS) on geosciences;
- the absence of a regional geomagnetic data center for CIS countries;
- the sparse coverage of the territory of Russia with geomagnetic observatories;
- the significant delay in preparing definitive geomagnetic data;
- the manual quality control and processing of geomagnetic data within the INTERMAGNET program; and
- the absence of a popular science publication on the Earth's magnetic field evolution for the last several centuries.

Alexei Gvishiani, Chair

CODATA International could play a role in addressing these challenges by:
- distributing information on the Task Group achievements;
- searching new data sources for a unified GIS on geosciences;
- facilitating exchange of experience on handling and storing huge geosciences data sets;
- monitoring progress in the field of intellectual processing of geoscience data;
- representing data in a GIS environment;
- attracting a wider range of specialists to work in this area.

Anatoly Soloviev

EXCHANGEABLE MATERIALS DATA REPRESENTATION TO SUPPORT SCIENTIFIC RESEARCH AND EDUCATION

One of the most pressing issues with respect to materials data today is the establishment of data repositories as part of a global data infrastructure to enable the capture, storage, exchange and dissemination of materials data and results. This need is especially acute given the expanded capabilities of modelling and simulation in materials science and engineering as well as major initiatives recently launched to shorten the time required to develop and commercialize new materials. While materials property data share many characteristics in common with other types of scientific and technical data, the continually changing nature of materials results in unique features that differentiate

Laura Bartolo, Co-Chair

the data themselves and management of those data from practices appropriate for other disciplines.

CODATA International with its strong ties to ICSU, its Unions and other major scientific associations, provides essential policy-making services for the Task Group to address this critical requirement. As a well-recognized international organization, CODATA will help the Task Group convene an international workshop during its 2013-14 term to develop guidelines for the adoption of common data formats and schemas as well as for the operation of materials data repositories.

Laura Bartolo

FUNDAMENTAL PHYSICAL CONSTANTS
(http://www.codata.org/taskgroups/TGfundconst/index.html)

The main purpose of the CODATA Task Group on Fundamental Constants is the periodic generation of a least squares adjustment of the values of the fundamental physical constants based on all of the relevant data available at a given point in time. The major scientific data challenges facing the TGFC are inconsistent data and limited redundancy among key data available for any given CODATA adjustment. These challenges have received increased international attention since it now seems universally accepted by all concerned groups that the TGFC recommended values will be used in the proposed redefinition of the International System of Units (SI).

CODATA International can help address this challenge by following the recommendations of Resolution 12

David Newell, Chair

of the 23rd meeting of the General Conference on Weights and Measures (2007), in particular:

 i) Help initiate awareness and alert user communities of the possible redefinition of the SI and the technical and legislative implications of such a redefinition (see http://www.bipm.org/en/si/new_si/ on the possible future revision of the SI);

 ii) Encourage and support research efforts in national metrology institutes, the International Bureau and Weights and Measures (BIPM) and academic institutions pursuing relevant experiments that will help provide consistent and redundant data for the redefinition of the kilogram, the ampere, the kelvin, and the mole.

David Newell

GLOBAL INFORMATION COMMONS FOR SCIENCE INITIATIVE
(http://www.codata.org/taskgroups/TGgicsiEU/index.html)

The overall goal of the Global Information Commons for Science Initiative is to accelerate the development and scaling up of open scientific data and information resources on a global basis, with particular focus on "common use" licensing approaches. In particular this involves work on alternative legal and technical regimes built by the scientific communities such as copyleft licensing, open access publishing and dynamic distribution of metadata. The Task Group focuses on public research
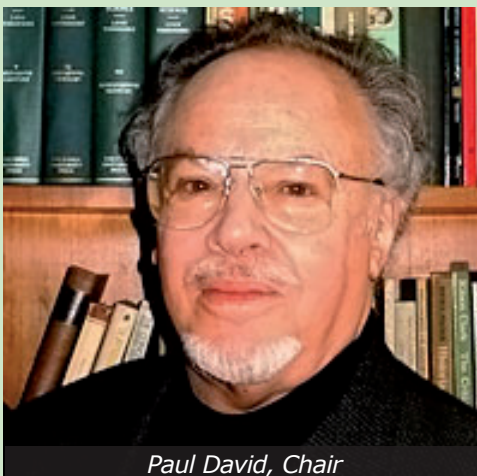
Paul David, Chair

and academic institutions' policies and practices for deposit of data. CODATA International plays a crucial role in addressing this challenge, both as a forum that is well recognized with science and science policy leaders and as a nucleus to gather knowledge on best practices in this field.

Tom Dedeurwaerdere

GLOBAL ROADS DATA DEVELOPMENT
(http://www.codata.org/taskgroups/TGglobalroads/index.html)

There is a major demand for open access road and other transportation infrastructure data on the part of the environmental conservation and economic development communities, as outlined in the Global Roads Strategy paper. The highest quality roads data are generally available only from the private sector (e.g. mapping and navigation firms). However, there is a growing volume of public data that are better than previously available public domain roads data and there are new methods for roads data extraction (e.g. from satellite imagery). The role of the Global Roads

Alex de Sherbinin, Chair

Data Development Task Group is to compile improved roads data and to foster roads data extraction.

 CODATA International has served as an extremely useful platform for organizing this work and publicizing the need. CODATA can continue to highlight the need for improved global data in a range of fields that are important to sustainable development and might consider fostering data development in other areas, such as electricity grids or rail transport infrastructure, where data are needed.

Alex de Sherbinin

Linked Open Data for Global Disaster Risk Research (LODGD)

Guoqing Li, Co-Chair

Angelika Wirtz, Co-Chair

Susan Cutter, Co-Chair

The challenges for scientific data may be found in every stage of the data life-cycle for every field. These challenges include problems related to data observation, transfer, preservation, processing, publishing, visualization and understanding along with that of linking data with other data as a whole. The challenge for LODGDRR is how to link different disaster risk datasets to serve as a global data resource view of a particular disaster event. The scientific value of research data is guaranteed by their completeness and diversity. Linked data as a whole provide more scientific usefulness to disaster researchers.

First of all, CODATA International provides an umbrella to bring different but related data providers together, even though some may not identify themselves as disaster data providers. CODATA can protect data diversity while maintaining their multidisciplinary nature. Secondly, CODATA's data sharing principles provide a baseline so that all data providers linked in LODGDRR activities may understand the basic data policy. Lastly, but not least, CODATA has a long history as a forum in which disaster data may be discussed. It provides a great community to discuss and share experience on disaster-related topics. The LODGDRR activity is jointly supported by CODATA and the IRDR (Institute for Risk and Disaster Reduction) in the form of financial and in-kind support. The LODGDRR is optimistic that its activity can enhance the joint relationship between these two ICSU bodies.

Guoqing Li

Octopus: Mining Space and Terrestrial Data for Improved Weather, Climate and Agricultural Predictions

Kassim Mwitondi, Co-Chair

Kehe Wang, Co-Chair

Extracting knowledge from data has always been the ultimate goal of any research activity. OCTOPUS seeks to develop, enhance and apply data mining techniques for analysing the interactions of space and terrestrial weather and how they impact life on earth. Developing global or regional policies, strategies and indices relating to the supply of food, water or other resources largely depends on the accuracy and reliability of such analyses. However, recent developments in methods of collecting, archiving and analysing data have continued to pose both challenges and opportunities to the data science community. For instance, with more than 300 years of sunspot number records, more than 60 years of ionospheric data and natural disasters and famines data spanning many years, some of the data challenges OCTOPUS faces are embedded in the very issues the groups seek to address. Thus, the major scientific challenge OCTOPUS faces is to develop a general multidisciplinary framework for dealing with spatio-temporal data across disciplines – a challenge that has subtly characterised the scientific community for generations.

OCTOPUS' major scientific challenge hinges on, inter-alia, global weather/climate parameters, space and terrestrial phenomena and related activities such as agricultural production, water supply and natural hazards. Therefore, from the Task Group's view, it is imperative that CODATA International supports and promotes a balanced influence not only across regions but also across disciplines. OCTOPUS initiatives include bridging the current coverage gap in the Middle East and Africa – where OCTOPUS believes CODATA International needs to extend its activities in the near future.

Kassim Mwitondi

Preservation of and Access to Scientific and Technical Data in/for/with Developing Countries (PASTD)

(http://www.codata.org/taskgroups/TGpreservation/index.html)
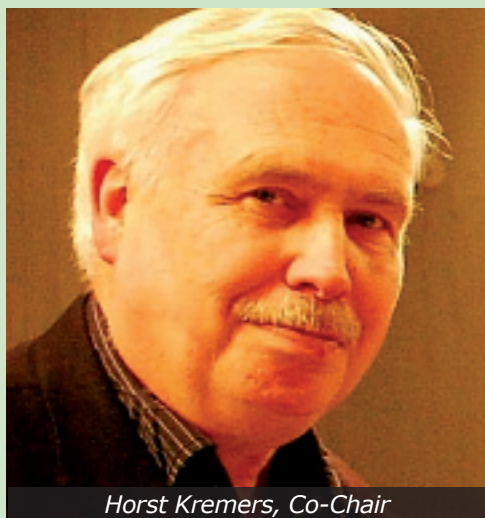
Horst Kremers, Co-Chair

The most critical issues relating to scientific data today concern their volume, their velocity and their veracity. We are generating very huge volumes of data on a daily basis. These data zoom past rapidly leaving no time for adequate comprehension. Finally, it is increasingly difficult to make sense of the meaning and context of the massive volume of texts, photos, conversations and figures and even more difficult to write programs that can enable scientists to process the data.

To tackle these challenges, CODATA International has to:

1. Foster mutually beneficial relationships with key players in the big data environment, such as IBM and Huawei among others, who have invested and innovated in huge data storage and processing infrastructure (hardware and networks) as well as robust data processing applications.

2. Foster versatile multidisciplinary teams in the science and technology arena to constantly and consistently

Mabel Imbuga, Co-Chair

work together in unravelling the huge data volumes, keeping pace with the rate of its creation and life as well as making sense of diverse data.

Mabel Imbuga

58th Executive Committee Meeting, Paris, April 2013

The newly elected Executive Committee will meet in Paris in April 2013 to discuss and oversee the implementation of the decisions that were approved by the General Assembly in Taipei. Among some of the scientific issues that will be discussed by the Committee are:

- Implementation of the **CODATA Strategic Plan** 2013-2017
- CODATA Task Group Work Plans 2012-2014
- Data challenges within **Future Earth**, the new 10-year international research initiative being led by ICSU and an Alliance of Partners, that will develop the knowledge for responding effectively to the risks and opportunities of global environmental change.
- Collaboration with the **Research Data Alliance (RDA)** and the CODATA/RDA Working Group proposal on legal interoperability.
- Its work on **Nanomaterials** under the CODATA/

VAMAS Working Group including interaction with the European Commission and International Scientific Unions in this area.

- Outreach strategy to the **Social Scientific Community and Members** in general.

The importance of interaction between CODATA and the **ICSU World Data System** will also be addressed. Representatives of both organizations will meet in Paris to discuss the collaboration that is currently underway and how this can be built upon and strengthened as both communities move forward leading to the International Conference in New Delhi in 2014.

This is an exciting time for CODATA with great opportunities for collaboration amongst many relevant stakeholders in the global data community.

CODATA is delighted to announce that
the **24th International CODATA Conference**
and the **29th CODATA General Assembly**
will take place in New Delhi, India.

More details will follow shortly.