Data-Intensive Science and Discovery
CODATA 45 Years On

Summary Report

Organizer:
Committee on Data for Science and Technology
International Council for Science

Hosts:
Center for Earth Observation and Digital Earth
Chinese Academy of Sciences

Chinese National Committee for CODATA

Computer Network Information Center
Chinese Academy of Sciences
This year marks an important milestone in the decades-long history of the International Council for Science (ICSU) Committee on Data for Science and Technology (CODATA). To commemorate its 45th anniversary, CODATA held a high-level scientific meeting in Beijing on October 30th, 2011, to discuss discovery and innovation in the age of data-intensive science.

The meeting brought together CODATA’s members and contributors to look back on the international organization’s history as well as begin formulating the next iteration of CODATA’s strategic plan. A number of internationally renowned scientists attended plenary sessions at the Center for Earth Observation and Digital Earth (CEODE), an institute of the Chinese Academy of Sciences (CAS).

The meeting, CODATA—45 years on, was intended to be a celebration, bringing together the CODATA family after nearly a half-century of service, but this was only part of its purpose. A significant component of the meeting was its integration of different aspects of data science, covering science and policy, national and international organizations, and public and private institutions. The inclusion of a broad range of stakeholders is inherently fundamental to any successful initiative in data science because of the nature of the field. The use of data is not limited to certain disciplines, methodologies, or nations. On the contrary, it is the cornerstone of any endeavor that involves recording, retrieving, or processing information. CODATA, therefore, must take a holistic approach in its task “to strengthen international science for the benefit of society by promoting improved scientific and technical data management and use”, to quote the CODATA mission statement. The success of CODATA—45 years on can be attributed to the exceptional breadth of participation we saw throughout the day.

In the realm of science, the meeting was not short of influential representatives. We were thankful that we could be joined by Yuan Tseh Lee, President of ICSU and a Nobel Prize laureate. As the meeting was held in Beijing, we were fortunate enough to hear from the President of the Chinese Academy of Sciences, BAI Chunli, and our opening remarks were delivered by the Vice-President, DING Zhongli, who is also President of the CODATA National Committee of China. The President of the Indian National Science Academy, Krishan Lal, also participated as a CODATA Past-President. These individuals are notable not just for their scientific leadership, but for the contributions they and their organizations provide in decision-making processes. Coming from a more direct role in government were XU Guanhua, member of the UN-GAID Strategy Council and former Minister of Science and Technology...
in China; and ZHANG Xian’En, Director-General of the Department of Basic Research in the Ministry of Science and Technology (MOST). MOST funds much of the leading research in China, particularly research in high-technology or data-intensive fields. Among other research initiatives, MOST organizes and implements the National Basic Research Program of China (the 973 Program). And coming from the private sector, we had a researcher from Microsoft Research’s eScience division, Jennifer Listgarten. Microsoft Research has put significant effort into advancing data-intensive science. The late Jim Gray, former manager of Microsoft Research eScience, championed data-intensive science as the “fourth paradigm” of science.

We concluded the meeting with the Celebration Ceremony, which culminated in the launch of two new partnerships as part of CODATA’s “Hand in Hand Program”, an initiative to spur international cooperation. The first was with the Integrated Research on Disaster Risk program. CODATA and IRDR signed a memorandum of understanding to lay the groundwork for future collaboration. Disaster risk research is a good example of problem-oriented science spanning across a multitude of fields, all of which need to standardize, share and integrate data. This need highlights the importance of CODATA’s efforts and the CODATA-IRDR partnership. Some of the potential areas of cooperation outlined in the agreement include participation in task and working groups, for example IRDR’s disaster loss data working group, and joint sessions at meetings and conferences. Gordon McBean, then Chair of the IRDR Scientific Committee and President-elect of ICSU, signed for IRDR.

CODATA also signed a memorandum of understanding with the International Society for Digital Earth (ISDE), represented at the meeting by their Vice-President, Milan Konecny. Digital Earth is still a developing concept, but it has tremendous implications for science. The idea is to have a comprehensive, integrated system for storing, sharing, viewing, and analyzing an unprecedented wealth of information about any given geographic location. It is an ambitious goal in which data management plays a key part, and CODATA is well positioned to contribute.

One of the most beneficial outcomes of having a diverse group of participants at CODATA—45 years on was the increased awareness of data issues. Many of our guests and participants would not necessarily consider themselves data science experts, though they rely heavily on data in all their research. While every researcher collecting or using data should not be expected to become a data scientist, they should at least be aware of issues pertaining to data quality, standards, sharing, and preservation. This awareness needs to extend throughout policies and practices, something that CODATA strives to accomplish, and it is in that regard that CODATA—45 years on saw its greatest achievement.

For its success, I wish to thank all of the participants, attendees, and organizers for their invaluable contributions. I would like to recognize in particular the speakers, who gave us food for thought about CODATA and data-intensive science throughout the day. In addition to those already mentioned, special thanks go to Sálvano Briceño, Ron Abler, Belinda Seto, GAO Wei, Takashi Gojobori, Fedor Kuznetsov, John Rumble, ZHANG Zhonghua, Robert Chen, Paul Uhlir, Alik Ismail-Zadeh, Yasuhiro Murayama, Shuichi Iwata, SUN Honglie, LIU Yanhua, Kathleen Cass, Jayakumar Ramasamy, ZHOU Heng, LIAO Xiaohan, LIU Chuang, LI Jianhui, LIU Jie, LIANG Dong, HONG Tianhua, ZHANG Hui, LI Chang, and Luke Driskell.

To capture the thoughts expressed and activities that took place during CODATA—45 years on, I am proud to present this book of the day’s proceedings. It includes not only the material presented in the sessions, but photographs taken throughout the meeting, breaks, and banquet. I hope they adequately serve as a reminder of the relationships—both formal and informal—that were established on that day.

Guo Huadong
President
Committee on Data for Science and Technology
International Council for Science
I. COMPREHENSIVE REPORT

1. Background Information

The Committee on Data for Science and Technology (CODATA) is an interdisciplinary Scientific Committee of ICSU established in 1966 to promote and encourage, on a world-wide basis, the compilation, evaluation and dissemination of reliable numerical data of importance to science and technology. The mission of CODATA is to strengthen international science for the benefit of society by promoting improved scientific and technical data management and use.

Specific objectives include:

- The improvement of the quality and accessibility of data, as well as the methods by which data are acquired, managed, analyzed and evaluated, with a particular emphasis on developing countries
- The facilitation of international cooperation among those collecting, organizing and using data
- The promotion of an increased awareness in the scientific and technical community of the importance of these activities
- The consideration of data access and intellectual property issues

CODATA uses many mechanisms in its on-going effort to reach these objectives. These are: Task Groups, Working Groups, National Member activities, conferences, workshops, publications, co-operation with other organizations on common interests, and a proactive Secretariat and Executive Committee.

It is important to emphasize that data-related activities are not confined to a specific scientific discipline but they relate to data activities in every area of science and technology, from biology to global change, from the physical sciences to engineering. CODATA reaches out to different disciplines through its growing number of members within the CODATA family, each outlining their different data needs, sharing knowledge about their respective data activities and identifying common areas of interest.

The national committees of CODATA often organize data activities on a national level. Much of the most important work of CODATA, however, lies outside its formal activities in its providing a milieu in which data experts from different countries can interact, cooperate directly, develop bilateral collaborations outside of CODATA, and exchange ideas and knowledge.

CODATA is open to new projects, especially in this time of increasing importance of data because of computer-based modeling and the Internet. Today more and more data sources of unknown quality and origin are becoming easily available via the World Wide Web. CODATA provides a home for international data experts needing to address data quality and data access issues to turn the Information Revolution into a positive force for the future. Today's data become the products and processes of tomorrow. CODATA is prepared to help science and technology achieve a better tomorrow through better data today.

At the 20th CODATA international conference, held in 2006 in Beijing, CODATA celebrated its 40th anniversary with a look back at its history, and a look ahead at where CODATA was going. Then in 2011, for CODATA's 45th anniversary, it was decided to return to Beijing for a high-level meeting to mark the occasion and once again examine the past, present, and future of CODATA and data science.
2. Agenda

<table>
<thead>
<tr>
<th>Time</th>
<th>Venue: Digital Earth Demonstration Hall, CEODE, CAS</th>
<th>Date: Sunday, October 30, 2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>8:30 – 9:00</td>
<td>Registration</td>
<td>Ground Floor Lobby</td>
</tr>
<tr>
<td>9:00 – 9:05</td>
<td>Welcome Remarks: DING Zhongli, President, Chinese National Committee for CODATA</td>
<td></td>
</tr>
<tr>
<td>9:05 – 10:00</td>
<td>Keynote Session</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chair: Krishan Lal, Past-President, CODATA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Keynote Speakers:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Enhancing Data Management for Disaster Risk Reduction by Strengthening Collaboration Between CODATA and IRDR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Sálvano Briceño, Scientific Committee Incoming Chair, Integrated Research on Disaster Risk (IRDR) Program</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Data-Intensive Science Discovery Practice in China</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• ZHANG Xian’En, Director-General, Department of Basic Research, Ministry of Science and Technology of the People’s Republic of China</td>
<td></td>
</tr>
<tr>
<td>10:00 – 10:30</td>
<td>Break</td>
<td>Ground Floor Lobby</td>
</tr>
<tr>
<td>10:30 – 12:00</td>
<td>Session A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chair: Takashi Gojobori, Vice-President, CODATA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Machine Learning and Genomic Data, Jennifer Listgarten, eScience, Microsoft</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Molecular Medicine, A Confluence of Data, Belinda Seto, NIH/NIBIB</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Scientific Data and Earth Science, GAO Wei, Colorado State University</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Vision for Data-Intensive Life Science: Genome Information-oriented Society, Takashi Gojobori, National Institute of Genetics</td>
<td></td>
</tr>
<tr>
<td>12:00 – 13:15</td>
<td>Lunch</td>
<td>Cafeteria of CEODE</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13:15 – 14:30</td>
<td>Session B</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chair: Fedor Kuznetsov, Vice-President, CODATA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Highlights of CODATA Success Stories over 45 years</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Overview, Fedor Kuznetsov, Vice-President, CODATA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Fostering the S &amp; T Data Revolution, John Rumble, CODATA Former President</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ongoing Success Stories:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Fundamental Constants, ZHANG Zhonghua, National Institute of Metrology</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• GEO Data Sharing Task Force, Robert Chen, Co-Chair, GEO DSTF</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• CODATA Outreach, Krishan Lal, Past-President, CODATA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Celebration of the Fifth China-U.S. Roundtable Meeting, Paul Uhlir, CODATA-U.S.</td>
<td></td>
</tr>
<tr>
<td>14:30 – 15:30</td>
<td>Session C</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chair: Robert Chen, Secretary General, CODATA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>This panel session focuses on the CODATA Strategic Plan over the next five years</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Input from CODATA Members and Audience</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• CODATA Strategic Plan 2013-2017, Robert Chen, Secretary General, CODATA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• National Member Representative</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Union Member Representative, Alik Ismail-Zadeh, Secretary General, IUGG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Co-opted Member Representative, Yasuhiro Murayama, Acting Director, International Programme Office of ICSU World Data System</td>
<td></td>
</tr>
<tr>
<td>15:30 – 16:00</td>
<td>Break</td>
<td>Ground Floor Lobby</td>
</tr>
<tr>
<td>15:40</td>
<td>Group Photo</td>
<td>Front of CEODE Building</td>
</tr>
</tbody>
</table>
3. Participants

<table>
<thead>
<tr>
<th>Name</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Ronald F. Abler President, International Geographical Union</td>
</tr>
<tr>
<td>B</td>
<td>BAI Chunli President, Chinese Academy of Sciences</td>
</tr>
<tr>
<td></td>
<td>BAO Shuming Senior Research Coordinator, China Data Center, University of Michigan</td>
</tr>
<tr>
<td></td>
<td>Silvano Briceño Scientific Committee Incoming Chair, Integrated Research on Disaster Risk</td>
</tr>
<tr>
<td>C</td>
<td>Kathleen Cass Executive Director, Committee on Data for Science and Technology</td>
</tr>
<tr>
<td></td>
<td>Robert Chen Secretary General, Committee on Data for Science and Technology</td>
</tr>
<tr>
<td></td>
<td>CHUANG Tyng-Ruey Associate Research Fellow, Research Center for Information Technology Innovation, Academia Sinica</td>
</tr>
<tr>
<td>D</td>
<td>DING Zhongli President, Chinese National Committee for CODATA Vice-President, Chinese Academy of Sciences</td>
</tr>
<tr>
<td>F</td>
<td>FU Xiaofeng Ministry of Science and Technology of the People’s Republic of China</td>
</tr>
<tr>
<td>G</td>
<td>Tuvdendorj Galbaatar Secretary General, Mongolian Academy of Sciences</td>
</tr>
<tr>
<td>Name</td>
<td>Position and Affiliation</td>
</tr>
<tr>
<td>---------------------------</td>
<td>-----------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>GAO Wei</td>
<td>Director, USDA UV-B Monitoring and Research Program and the Center of Remote Sensing and Modeling for Agricultural Sustainability, Natural Resource Ecology Laboratory, Colorado State University</td>
</tr>
<tr>
<td>Jean Garnier</td>
<td>International Union for Pure and Applied Biophysics</td>
</tr>
<tr>
<td>GE Tong</td>
<td>Deputy Director, Columbia Global Centers (East Asia, Beijing)</td>
</tr>
<tr>
<td>GE Zhigang</td>
<td>Director, China Nuclear Data Center, China Institute of Atomic Energy</td>
</tr>
<tr>
<td>Takashi Gojobori</td>
<td>Vice-President, Committee on Data for Science and Technology</td>
</tr>
<tr>
<td>T.V. Golashvili</td>
<td>Director, Head Branch Scientific Centre at Atominform</td>
</tr>
<tr>
<td>GU Xingfa</td>
<td>Co-Chair, CODATA Task Group “Preservation of and Access to Scientific and Technical Data in Developing Countries” Director General, Institute of Remote Sensing applications, Chinese Academy of Sciences</td>
</tr>
<tr>
<td>GUO Huadong</td>
<td>President, Committee on Data for Science and Technology</td>
</tr>
<tr>
<td>HAN Chunxu</td>
<td>Deputy Secretary General, Chinese National Committee for the International Union of Biological Sciences</td>
</tr>
<tr>
<td>HO Jan-Ming</td>
<td>Research Fellow, Research Center for Information Technology Innovation, Academia Sinica</td>
</tr>
<tr>
<td>HU Yaruo</td>
<td>Former Secretary General, Chinese National Committee for CODATA</td>
</tr>
<tr>
<td>Alik IsmailZadeh</td>
<td>Secretary General, International Union of Geodesy and Geophysics</td>
</tr>
<tr>
<td>Koki Iwao</td>
<td>Research Scientist, GEO Grid Research Group, Information Technology Research Institute, National Institute of Advanced Industrial Science and Technology</td>
</tr>
<tr>
<td>Shuichi Iwata</td>
<td>Former President, Committee on Data for Science and Technology Professor, University of Tokyo</td>
</tr>
<tr>
<td>JIA Gensuo</td>
<td>Professor, Institute of Atmospheric Physics, Chinese Academy of Sciences</td>
</tr>
<tr>
<td>Milan Konency</td>
<td>Vice-President, International Society for Digital Earth</td>
</tr>
<tr>
<td>Marina Kosinova</td>
<td>Head of Laboratory, Institute of Inorganic Chemistry, Siberian Branch, Russian Academy of Sciences</td>
</tr>
<tr>
<td>Fedor Kuznetsov</td>
<td>Vice-President, Committee on Data for Science and Technology</td>
</tr>
<tr>
<td>Krishan Lal</td>
<td>Past-President, Committee on Data for Science and Technology President, Indian National Science Academy</td>
</tr>
<tr>
<td>Yuan Tseh Lee</td>
<td>President, International Council for Science</td>
</tr>
<tr>
<td>LI Chang</td>
<td>Secretariat, Chinese National Committee for CODATA</td>
</tr>
<tr>
<td>LI Jianhui</td>
<td>Secretary General, Chinese National Committee for CODATA</td>
</tr>
<tr>
<td>LI Wen</td>
<td>Senior Engineer, Computer Network Information Center, Chinese Academy of Sciences</td>
</tr>
<tr>
<td>LIANG Yingnan</td>
<td>Committee Member, Chinese National Committee for CODATA Deputy Director, Department of International Affairs, China Association for Science and Technology</td>
</tr>
<tr>
<td>LIAO Xiaohan</td>
<td>Director, National Remote Sensing Center of China</td>
</tr>
<tr>
<td>LIN Hui</td>
<td>Director, Institute of Space and Earth Information Science, Chinese University of Hong Kong</td>
</tr>
<tr>
<td>LIN Xianwen</td>
<td>President, Jieyang City Art Association</td>
</tr>
<tr>
<td>LIN Yongxin</td>
<td>Jieyang Municipal Government Civil Service Office</td>
</tr>
<tr>
<td>Jennifer Listgarten</td>
<td>Microsoft Research eScience Division</td>
</tr>
<tr>
<td>Name</td>
<td>Position and Affiliation</td>
</tr>
<tr>
<td>-----------------------</td>
<td>------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>LIU Chuang</td>
<td>Professor, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences</td>
</tr>
<tr>
<td>LIU Lei</td>
<td>Professor, Shanghai Center for Bioinformation Technology</td>
</tr>
<tr>
<td>LIU Xu</td>
<td>Committee Member, Chinese National Committee for CODATA Vice-President, Chinese Academy of Agricultural Sciences</td>
</tr>
<tr>
<td>LIU Yanhua</td>
<td>Former Vice-Minister, Ministry of Science and Technology of the People’s Republic of China President, Geographical Society of China</td>
</tr>
<tr>
<td>LÜ Yonglong</td>
<td>President, Scientific Committee on Problems of the Environment Committee Member, Committee on Scientific Planning and Review Director-General, Bureau of International Co-operation, Chinese Academy of Sciences</td>
</tr>
<tr>
<td>James Luo</td>
<td>Program Director, Extramural Science Program, National Institute of Biomedical Imaging and Bioengineering, National Institutes of Health</td>
</tr>
<tr>
<td>MA Juncai</td>
<td>Director, World Data Center for Microorganisms</td>
</tr>
<tr>
<td>Gordon McBean</td>
<td>President-Elect, International Council for Science Scientific Committee Chair, Integrated Research on Disaster Risk</td>
</tr>
<tr>
<td>Yasuhiro Murayama</td>
<td>Acting Director, International Programme Office, World Data System</td>
</tr>
<tr>
<td>Farrokh Nadim</td>
<td>Director, International Centre for Geohazards, Norwegian Geotechnical Institute</td>
</tr>
<tr>
<td>Veronika Nedvědová</td>
<td>Second Secretary, Embassy of the Czech Republic in Beijing</td>
</tr>
<tr>
<td>Baasan Nergui</td>
<td>Scientific Secretary, Institute of Informatics, Mongolian Academy of Sciences</td>
</tr>
<tr>
<td>PENG Jie</td>
<td>Director, Resources Sharing and Promotion Center, Institute of Scientific and Technical Information of China</td>
</tr>
<tr>
<td>Jayakumar Ramasamy</td>
<td>Programme Specialist for Natural Sciences, UNESCO Beijing Office</td>
</tr>
<tr>
<td>Jane Rovins</td>
<td>Executive Director, International Programme Office, Integrated Research on Disaster Risk</td>
</tr>
<tr>
<td>John Rumble</td>
<td>Former President, Committee on Data for Science and Technology</td>
</tr>
<tr>
<td>Belinda Seto</td>
<td>Deputy Director, National Institute of Biomedical Imaging and Bioengineering, National Institutes of Health</td>
</tr>
<tr>
<td>SU Xiaodong</td>
<td>Chair, Commission on Biological Macromolecules, International Union of Crystallography</td>
</tr>
<tr>
<td>SUN Honglie</td>
<td>Former Vice-President, Committee on Data for Science and Technology</td>
</tr>
<tr>
<td></td>
<td>Former Vice-President, International Council for Science</td>
</tr>
<tr>
<td>TANG Yi</td>
<td>Deputy Director, Institute of Earthquake Science, China Earthquake Administration</td>
</tr>
<tr>
<td>Paul F. Uhlir</td>
<td>Director, Board on Research Data and Information, U.S. National Academies</td>
</tr>
<tr>
<td>E. Lynn Usery</td>
<td>Research Geographer and Director, Center of Excellence for Geospatial Information Science, U.S. Geological Survey</td>
</tr>
<tr>
<td>WANG Changlin</td>
<td>Executive Director, International Society for Digital Earth</td>
</tr>
<tr>
<td>WANG Long</td>
<td>Institute of Earthquake Science, China Earthquake Administration</td>
</tr>
<tr>
<td>WANG Xiaqing</td>
<td>Institute of Earthquake Science, China Earthquake Administration</td>
</tr>
<tr>
<td>WANG Zhenyu</td>
<td>Division Director, Bureau of International Co-operation, Chinese Academy of Sciences</td>
</tr>
<tr>
<td>Takashi Watanabe</td>
<td>Scientific Committee Member, World Data System</td>
</tr>
<tr>
<td>Angelika Wirtz</td>
<td>Scientific Committee Member, Integrated Research on Disaster Risk</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>X</td>
<td></td>
</tr>
<tr>
<td>XIAO Yun</td>
<td>Former Secretary General, Chinese National Committee for CODATA</td>
</tr>
<tr>
<td>XIONG Yun</td>
<td>Deputy Director, Research Center for Dataology and Data Science, Fudan University</td>
</tr>
<tr>
<td>XU Guanhua</td>
<td>Former Minister, Ministry of Science and Technology of the People’s Republic of China; Strategy Council Member, United Nations Global Alliance for ICT and Development</td>
</tr>
<tr>
<td>XU Zheping</td>
<td>Institute of Botany, Chinese Academy of Sciences</td>
</tr>
<tr>
<td>XU Zhihong</td>
<td>Former National Delegate, Chinese National Committee for CODATA</td>
</tr>
<tr>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>YAN Baoping</td>
<td>Former Vice-President, Chinese National Committee for CODATA</td>
</tr>
<tr>
<td>YAN Jingfu</td>
<td>Technology Analyst, Asian Technology Information Program</td>
</tr>
<tr>
<td>YANG Shengtian</td>
<td>Director, School of Geography, Beijing Normal University</td>
</tr>
<tr>
<td>Z</td>
<td></td>
</tr>
<tr>
<td>ZHANG Hui</td>
<td>Secretariat, Chinese National Committee for CODATA</td>
</tr>
<tr>
<td>ZHANG Kan</td>
<td>President, Chinese Psychological Society, International Union of Psychological Science</td>
</tr>
<tr>
<td>ZHANG Xian’En</td>
<td>Member, Chinese National Committee for CODATA; Director-General, Department of Basic Research, Ministry of Science and Technology of the People’s Republic of China</td>
</tr>
<tr>
<td>ZHANG Yili</td>
<td>Professor, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences</td>
</tr>
<tr>
<td>ZHANG Zhonghua</td>
<td>Member, CODATA Task Group on Fundamental Physical Constants; Professor, National Institute of Metrology</td>
</tr>
<tr>
<td>ZHOU Heng</td>
<td>China Meteorological Administration</td>
</tr>
<tr>
<td>ZHOU Xiang</td>
<td>Institute of Remote Sensing Applications, Chinese Academy of Sciences</td>
</tr>
<tr>
<td>ZHU Yangyong</td>
<td>Director, Research Center for Dataology and Data Science, Fudan University</td>
</tr>
<tr>
<td>ZOU Xiaolei</td>
<td>Professor, China Meteorological Administration</td>
</tr>
</tbody>
</table>
II. ADDRESSES AND REMARKS

On the occasion of the 45th Anniversary of the Committee on Data for Science and Technology, on behalf of the Chinese Academy of Sciences, I would like to congratulate the opening of this Celebration Ceremony and extend my warm welcome to all experts, scholars and delegates from the International Council for Science, CODATA, the Integrated Research on Disaster Risk program, and the impressive variety of other organizations represented here today.

Since their establishment in 1966, CODATA has been devoted to promoting and encouraging the compilation, evaluation and dissemination of reliable scientific and technical data around the world. Considering the important role CODATA has played and will play in data science and technology, China joined CODATA to contribute to its efforts. In 1984, the Chinese Academy of Sciences took the initiative in launching the Chinese National Committee of CODATA. And I am honored to be informed that in October 2010, Professor GUO Huadong was elected as the new president of CODATA. I believe that with his efforts and contributions CODATA will make further achievements in data science.

We have witnessed that with its 45-year effort, CODATA has successfully stepped up to its goal of effective, efficient management of scientific data. Decades ago, their pioneering efforts in computerized data management helped scientists keep up with the observations and information pouring in to their possession. The scientific community would thank ICSU and CODATA for the level and scale of cooperation they have brought to modern science. Sciences are more and more connected, globally and integrated across disciplines. Data sharing, and everything that entails, underpins all of this. It allows truly collaborative research, where scientists can build on the findings of others, and make new discoveries in unexplored data. Moreover, they can uncover information that had previously gone unnoticed in an old dataset.

The scientists of tomorrow should not be concerned with varying practices and data formats preventing them from using data from a colleague in a different field. They should not have to worry about complex barriers limiting their access to data in another country. And they should not have to question whether their projects will have the resources and expertise to process and store all the valuable data they produce. Instead, standards, reasonable access policies, and general awareness of the importance of data science should be prevalent in all forms of scientific practices.

Clearly this places a lot of responsibility on organizations like CODATA, but they have proven that they are up to the challenge. So please join me in congratulating them on their 45 years of service to science and society, and wishing them success in the years to come.

It’s really my pleasure to be here. On behalf of the Chinese Academy of Sciences and the Chinese National Committee for CODATA, I would like to extend my warm welcome to all of the participants of this meeting, particularly to the international experts, and I express my sincere congratulations for the 45th anniversary of CODATA.

In the last decades, with recent advancement of research methods and rapid development of information technology, we have seen a rapid rise of scientific data acquisition through observation, experiments, and simulation. The problem of long-term preservation, integrated management and efficient use of mass data has become an important issue for science and engineering research and development. The establishment and development of scientific data platforms have effectively promoted multi-disciplinary, cross-regional, multi-institutional exchange and cooperation in large-scale data sharing. And the formulation of large-scale data sets through the efforts of various research bodies has provided critical support for data-intensive science and engineering.
Thanks to their establishment 45 years ago and their development, CODATA has become the world's largest scientific data organization, playing a leading part in the improvement of scientific data quality and management, and the promotion of data sharing. The various academic activities organized by CODATA has helped establish a world-wide platform for data storage and sharing. That is especially important for scientists in the less developed countries.

China joined CODATA in 1984. Over the years, with the guidance of the Chinese National Committee for CODATA, CODATA-China has established 10 inter-departmental, multi-disciplinary scientific data working groups, aiming to improve data quality, standardization, management, dissemination, and sharing, and has gradually formed and improved a data evaluation system.

Through the project of “scientific database and information system” that was launched in 1986, the Chinese Academy of Sciences has built a distributed scientific database and service system, consisting of over 500 disciplinary sub-systems with the participation of 45 research institutes. This is the most comprehensive database system in China now, supported by advanced technology and management systems. We have also built what we call the “CODATA Chinese Physics-Chemistry Database” that is providing efficient data-sharing services. China has really enjoyed a long-standing collaborative relationship with CODATA. The Chinese National Committee for CODATA has participated in all kinds of meetings organized by CODATA and we have an integrated relation with the colleagues in a number of data producing countries. We also successfully hosted CODATA international conferences in 1992 and 2006. In brief, the Chinese National Committee has been helping in supporting and taking part in all types of CODATA activities and has actively promoted communication and collaboration with other CODATA members, and will continue to do so in the days to come.

To conclude, I wish the conference great success and wish all of you a pleasant time in Beijing.

XU Guanhua
Former Minister, Ministry of Science and Technology of the People’s Republic of China
Strategy Council Member, UN Global Alliance for ICT and Development

I am greatly honored to be here today to join the celebration ceremony of the 45th Anniversary of the Committee on Data for Science and Technology. I would like to take this opportunity to express my sincere congratulations to the success of CODATA over the past 45 years.

Since the World Summit on Information Society in 2003 Geneva and 2005 Tunis, Building the Information Society has become the common desire and commitment of the United Nations. Digital data, knowledge, information and communication technology are at the core of human progress, development and well-being. For meeting the targets of the Millennium Development Goals, the United Nations established the Global Alliance for ICT and Development after the WSIS in 2006. Making digital scientific data and knowledge openly available to society is a task for the scientific communities. CODATA is one of the pioneers and leaders of such communities.

Five years ago I participated in the 20th international CODATA conference held in Beijing. As a keynote speaker, I talked about the importance of data sharing. Today I realize that CODATA has played a leading role in this field. International scientific organizations have an important role in the production of knowledge. Because of the immeasurable value of this process, such organizations have a great deal of work. This is especially true for CODATA. If knowledge is a product of science, then data are the raw materials. The product can only be as good as the materials from which it is made; thus proper data collection, management, and use are the cornerstone of sound science, reliable information, and applicable knowledge. Furthermore, our data collections can only be used to their full potential if they are openly accessible. On this aspect, technical limitations are one of the least significant barriers to the free flow of information, calling for a much more comprehensive approach to the promotion of data sharing.

Fortunately, for 45 years now, CODATA has made great progress in this task. The impact that CODATA has on the effectiveness of scientific research is encouraging. I do appreciate CODATA’s contribution to helping scientists deal with and make use of the massive amounts of observation data. I am also very pleased to see that CODATA has been continuing its mission not just for science, but for the greater good of society. As both citizens and scientists, it is important for us not to lose sight of why we do what we do.

Now we are in an age of data-intensive science. I believe that it is necessary to discuss the discovery and innovation of scientific data and information in this current context, and this high-level scientific meeting has definitely provided such an opportunity for the exchange of ideas and to further promote work on the management and accessibility of data. I hope this meeting will be seen as a milestone on the road to the development of data science and technology.

Once again, Happy Birthday, CODATA!
I am honored to be here in the celebration of CODATA’s 45th birthday. During the last 45 years, CODATA has certainly made some important contributions which we all feel proud of. It is almost impossible to overstate the importance of data and information. Data is not only fundamental to the scientific enterprises, its successful management is also essential for any research organization. For instance, the ICSU interdisciplinary bodies both generate and use an enormous amount of data and their success depends heavily on the effective management of this aspect.

Clearly, as one of the main representatives of the global scientific community, ICSU must assert a unique leadership role on this front. And as the ICSU strategic plan 2012-2017 makes clear, our firm commitment is to ensure the long term stewardship, quality, availability, and usefulness of scientific data and information. However, we will be trying to achieve these objectives in an environment of unprecedented challenges. For instance, rapid technological change has led to an explosion in the amount of data across virtually all fields. An average citizen can now receive data on air pollution on a simple cell phone. In fact, experts project that in the coming decades, the amount of data we generate will be much, much greater than anyone ever thought possible. Yet, at the same time, technology is also a driving factor of our unsustainable lifestyle. Therefore, it is up to us to do things with data and information that we never thought possible. Yet, at the same time, technology is also a driving factor of our unsustainable lifestyle. Therefore, it is up to us to figure out how to use these tools wisely to save human civilization. After all, data is not only an enormous amount of data and their success depends heavily on the effective management of this aspect.

But data is not only becoming more plentiful, it is also getting more complex with vastly different modes, scales, and dimensions. The movement towards inter- and trans-disciplinary research will elevate the challenge even more because different disciplines produce and use different kinds of data in different ways. The other challenge is the assurance of the quality of the data, because the quality of the data heavily affects the quality of the science. Erroneous data not only does not help the user, it may bring negative impacts. Ensuring quality data will necessarily entail learning how to eliminate data that is erroneous and obsolete. And of course, while many of us worry about too much data and overly complex data, the reality for our colleagues in the less economically developed countries is the opposite. They have the passion and intellect to do great science, but too many of them simply lack access to data information and technology. As long as this is true, the universality of science will remain out of reach.

For ICSU, the International Council for Science, the principle challenge for the next six years, as stated in ICSU’s Strategic Plan 2012-2017 will be about elevating the importance of data and information management throughout the ICSU family. Even as we push forward new major initiatives, the fact that these initiatives bring together very different disciplines and do not have data management as their chief concern further adds to the challenge of achieving a coordinated global approach. And CODATA will lead the charge along with the World Data System and other bodies, such as the International Network for the Availability of Scientific Publications (INASP). With the World Data System, which was generously hosted recently by Japan, our goal is to link up the hundreds of data centers and services across the world and across different scientific domains. We hope to integrate them into a truly global data system that provides worldwide access to high quality, interoperable datasets and products. This will be a global data infrastructure like none that ever existed.

CODATA, on the other hand, tackles the management and quality challenges, especially of an inter-disciplinary nature. Over the next three years, it will have a critical role to play in taking on the recommendations coming out of the ad hoc Strategic Coordinating Committee on Information and Data (SCCID), and the earlier Priority Area Assessment in 2004. The upcoming CODATA review will also tell us a lot about how it can continue to fulfill its critical mandate. Currently, CODATA has 20 national members, contrasted with the 121 national members of ICSU. For CODATA to help push our wider global dialog on data, we may want to think about broadening its membership, as suggested by President GUO of CODATA earlier. Also, business and technology actors increasingly drive scientific and data issues. It might be a good idea for CODATA and ICSU to engage with these actors more, such as ISO and the World Wide Web Consortium.

Distinguished guests, ladies and gentlemen, the rapid advancement of technology allows us to do things with data and information that we never thought possible. Yet, at the same time, technology is also a driving factor of our unsustainable lifestyle. Therefore, it is up to us to figure out how to use these tools wisely to save human civilization. After all, data is not all that useful unless we learn to transform it into knowledge and then wisdom. This is something we must take into account in the generation of new data. ICSU must be a leader in this endeavor, and this requires that CODATA, WDS, INASP, and all other bodies work together with coherence and a common purpose. Nothing shorter than that will be enough. Well, as somebody who is visiting China, I would like to complement the fact that China is taking a more and more important role in making contributions in international science. This is something that is really wonderful that all of us like to see.

Yuan Tseh Lee
President, International Council for Science
Thank you very much, I’m very pleased to be speaking on behalf of the Integrated Research on Disaster Risk program; our cooperation with CODATA is fundamental to us being able to meet our objectives. As we’ve been discussing in our scientific committee meeting in the last two days, including the Disaster Loss Data working group, the issue of information is critical to both preventing disasters from happening and warning people when they are about to happen if we can’t prevent them. So let’s really work together; this is fundamental to both of us and meets the ICSU goal of strengthening international science for the benefit of society—all societies.

SUN Honglie
Former Vice-President, International Council for Science
Former Vice-President, CODATA

It is with a profound feeling of pleasure that I meet all new and old friends in this golden autumn in Beijing, attending the CODATA 45th anniversary. First of all, please allow me to express my warm welcome and gracious greetings to our distinguished guests and my heartfelt congratulations on the 45th anniversary of the founding of CODATA.

In the past 45 years, CODATA has committed itself to promoting scientific data preservation, management, sharing and applications all over the world, and has made outstanding contributions to communication and cooperation across different regions and subjects.

From 1996 to 2004, I joined the CODATA family as a member. It was a wonderful opportunity to meet the world’s leading experts in the field of scientific data, and build up cordial and constructive working relationships with them. Even now I still maintain close attention on scientific data development, hold a great interest in the rapid development of data science, and would like to express appreciation for my international colleagues’ prominent achievements.

With the advancement of human society as well as the rapid growth and wide application of information technology, CODATA may meet new chances and challenges. I believe CODATA will surely continue to maintain its unique position and play an increasingly important role with its improvement and development, making even greater contributions to the scientific data enterprises of all mankind.

Finally, on behalf of Chinese scientists, I would like to express my heartfelt gratitude to CODATA. Thanks for your consistent support for Chinese scientific data activity and Chinese scientists. I hope that CODATA will play a more important role in the future.

John Rumble
Former President, CODATA (1998-2002)

Thank you so much for the opportunity to address this ceremony as the Past President of CODATA. One of the proudest scientific achievements I’ve had in my career has been to work with the outstanding scientists and data specialists who have formed the CODATA community. I have been particularly happy to make many friendships and partnerships with the data scientists and physical scientists within China over the last 20 years. It has been an extremely rewarding situation to work with friends here in Beijing and other places within China. I think today has shown that CODATA has had a tremendous impact in the scientific community, and I look forward to another 45 years and being invited back to celebrate CODATA’s 90th anniversary.
Thank you very much for giving me a chance to speak a little bit. Recently I have been thinking about natural disasters and human made disasters, driven by the various experiences of the Tohoku earthquake, tsunami and nuclear disasters that happened after March 11. Through that experience I have felt the very warm network of CODATA, and also the very good and very important background on how to deal with data. The data that were announced through the media and copied and pasted were sometimes exaggerated, sometimes ignored. There was a real deluge of data with different qualities. As a consequence of that, there were fear and stress put on the people, especially on the general public who could not control and improve their situation. We are not only facing these kinds of risks, but also there are many conflicts and problems in the world due to many reasons. We need to fight with climate change with a long-term view about risks for other human beings. The main reason for such kind of risk and also our fear and stress might be uncertainty. That kind of risk can be described by data and models, but in the case of models, modeling is very context dependent and also depends on authoritative data. To reduce the fear of the people and to start to make the right decisions for the future, we need to increase the quality of data, and we need to prepare that kind of data for society. It might be our goal and it is the source of our next action to work together for big, global problems.

Thank you Professor GUO for inviting all of us and organizing this wonderful event. I’d like to introspect, look at the future, and think about what we can do in the light of ever changing science, technology, and continuous generation of a very immense volume of data in the many different streams of science and technology for the benefit of society. We have a lot of success stories, but we also have several areas that need our consideration, particularly, the issues of the quality of data and data standards, for which the world will more and more look towards CODATA. The issue is that we get an enormous amount of information and data from the Internet, but most of it is not peer reviewed. So it’s high time we start some activity more serious than before to solve this problem. There are other issues of integrating the least developed and developed countries, and the middle-order countries like China and India and so on. So this also will remain a very, very important challenge. For me it has been a great pleasure being associated with CODATA. Having a physics background, I learned something about life sciences, something about engineering, something about others, trying to just comprehend what they were talking about. The harmony and working together of people with very different backgrounds has been a very interesting human experience. I recall often while working with CODATA, a very small incident in the epic Mahabharata from India. This relates to a dialog between Lord Shiva and Uma Parvati. As you may know Lord Shiva is a manifestation of God who in India is like the goddess supreme. She asked, “How do you define feminism? How do you tell the true sense of feminism?” While replying to Lord Shiva, Uma Parvati feels a little introspective and consults the holy rivers, and she states “O Ocean bound (holy) rivers! I do not see in this world or in Heavens any knowledge, which any one has created on his own (without collaboration of others), therefore, I humbly seek your counsel.” That is the spirit when you sit in CODATA. Each one of us has done something in our own field and risen to some height, but we have not achieved anything on our own. It is through cooperation and collaboration with others. That is, I think, the lesson we learn by working in CODATA, and I hope we keep this up in the future.
I would like to take this opportunity to follow up on the theme I raised in my award lecture last year at the CODATA conference in South Africa, which focused on data for policy and policy for data. The theme of today’s meeting is “data-intensive science and discovery”, but I submit to you that all spheres of human endeavor are now data intensive, including law.

In this context, I would like to draw your attention to a truly remarkable white paper published by the State Council of China this Thursday on “the socialist system of laws with Chinese characteristics”. The English version was printed in the China Daily newspaper on Friday. I encourage everyone to read it. This white paper summarizes the historical evolution of laws since 1978 and the gradual implementation of the rule of law in this great country. It cites the Chinese constitution as the fundamental document from which all legal instruments must flow, but it also notes the development of some 240 major laws in all spheres of life.

The white paper notes that these laws are essential for the country’s development and progress, including the progress of science and technology. It also recognizes that more needs to be done to implement the laws that are on the books.

The white paper lists many of the laws in areas in which S&T play a major role, including the health, environment, defense, and education sectors. In science itself, the State Council notes the law on scientific and technological progress and the law on popularization of science and technology. In information law and policy, there is the government Freedom of Information Act, intellectual property laws, such as for copyright and patent protection, the protection of computer software, and the right to network dissemination of information. And even more specifically to scientific data, there are the administrative guidelines of MOST on the China scientific data sharing program that ZHANG Xian’En discussed this morning.

The state council’s white paper reminds us that scientific research, generally, and data activities, specifically, are governed by the rule of law. Different countries have different approaches to the development and implementation of their national laws, and in the scope of legal regulations for scientific data and information. Lawmakers typically are not scientists or engineers themselves, and thus have limited understanding of the research process and the role of scientific data in it. They are concerned with all of society in their nation, not just with the progress of science. And while most lawmakers in both the legislative and administrative branches have a general appreciation of the importance of science for the economic and social well-being of their constituents, they need the directed advice and inputs of experts—people like you—to form the laws and regulations appropriately. Therefore, scientists and engineers, as citizens of their countries and of the world, have a duty to be engaged in the process of continual improvement of those laws, and not just attend to their own narrow areas of research and data activities.

As is the case with science and technology, the law is constantly evolving and requires increasing engagement of scientists and engineers, as well as the use of more factual data on which to base legal decisions and instruments. As the State Council’s white paper points out, “China will attach more importance to legislation in the scientific and technical fields. In order to meet the requirements of promoting reform…and advancing science and technology, we will improve the legal institutions which encourage scientific innovation, and protect intellectual property rights so as to build an innovation-oriented country.” Lawmakers in other nations, of course, have similar concerns, so it is our responsibility, both individually and collectively, to engage in the legal process, which affects us and our fellow citizens, according to our abilities.

And that concludes my sermon for this Sunday afternoon.

Jayakumar Ramasamy
Programme Specialist, UNESCO Beijing Office

It gives me great pleasure to take part in the 45th Anniversary celebration of CODATA, with whom UNESCO enjoys great partnership in exchange of scientific data and information as well a key partner in World Summit on the Information Society. Today we have many representatives who represent scientific associations, government bodies and international and national scientific organizations. We all have one thing in common, though—we all have a keen professional interest in the creation and distribution of knowledge—and are links in the scientific information chain.
UNESCO and ICSU enjoy their long standing cooperative partnership, a partnership that has always strived to derive full benefit from the fruitful symbiosis between an intergovernmental and non-governmental organization concerned with science. Hand in hand, both partners had devoted their efforts to promoting science and international co-operation and solidarity in the adverse atmosphere of the Cold War. UNESCO appreciates ICSU’s effort to catalyze the action of its scientific unions, interdisciplinary bodies and national members in response to the recommendations of the World Conference on Science. The holding of the World Conference on Science was a tangible expression of the close relationship that exists between UNESCO and the ICSU family. One such member, CODATA, has worked with ICSU and UNESCO in promoting the role of scientific information in the emerging information society, as input to the World Summit on the Information Society. CODATA also contributed in partnership with ICSU and UNESCO on “Science and the Information Society”, where scientific experts, managers, and representatives from several intergovernmental agencies identified the major issues for science in relation to the WSIS.

Halfway between the 2005 meeting of the World Summit held in Tunis, Tunisia and the 2015 target date for achieving the WSIS goals, we can reflect on the achievements made to date and determine how we can use the knowledge gained to enhance our impact.

A critical challenge is to promote an enabling environment for harnessing the potential of information and communication technologies (ICTs)—from radio to mobile media and the Internet—for the benefit of each and every individual, including the marginalized and most vulnerable. ICTs can enhance learning opportunities, facilitate the exchange of scientific information and increase access to content that is linguistically and culturally diverse.

Through its continued support to countries towards the achievement of the WSIS goals, UNESCO means to ensure that our vision of inclusive and pluralistic knowledge societies based on quality education for all, universal access to information and knowledge, and respect for cultural and linguistic diversity, becomes a reality across the world.

ZHENG Guoguang
Executive Council, World Meteorological Organization Administrator, China Meteorological Administration
(Represented by ZHOU Heng)

On behalf of Dr. ZHENG Guoguang, who is a member of the Executive Council of the World Meteorological Organization and the Administrator of the China Meteorological Administration, I have the honor to be here with colleagues from so many other international scientific organizations to celebrate the 45th anniversary of CODATA. Such a diverse gathering itself has showcased the identity of CODATA, an organization that has transcended disciplinary boundaries.

I would here extend appreciation very much to CODATA for the efforts made to encourage the development of data policy, data exchange and related management technology, which are crucial to ensure that data is not confined to any single field of study, but a source that can be shared by all.

In the case of meteorology, it is undoubtedly true that our research and operation are remarkably interdisciplinary, including, among others, the various subfields of meteorology, climatology, physics, oceanography, and chemistry. Actually, some of the tasks of the WMO are very similar to those of CODATA, such as integrating, standardizing, and sharing data collections from around the world and from different disciplines.

It is a routine work for meteorologists to deal with tremendous volume of data streaming in from the ground, the oceans, the sky and space, which has made us traditionally accustomed to cooperation and collaboration with colleagues from other fields. In the framework of WMO, data policy is also of great importance, aiming at encouraging data exchange. For example:

Essential data and products which are necessary for the provision of services in support of the protection of life and property and the well-being of all nations shall be provided on a free and unrestricted basis;

Members should provide to the research and education communities, for their non-commercial activities, free and unrestricted access to all data and products exchanged under the auspices of WMO, etc.

From the theme of the ceremony, everybody may have recognized importance of data to science. In the view of atmospheric science, the coverage, representativeness, accuracy, standardization, management and overall usability of data are fundamental to our research and services. It is quite safe to predict that, in the not too distant future, all science will be data-intensive. If that is to be the case, we had better be sure, from now on, of our methods for collecting, analyzing, sharing, storing, and presenting data. Under this scenario of sciences dominated by data, we may clearly see the importance of coordinated international cooperation between Organizations like the WMO and CODATA. For this reason, I am very grateful to be part of such an effort during this extraordinarily exciting time and I am also very thankful to have been given the opportunity to be here today.
CAO Jianlin
Vice-Minister, Ministry of Science and Technology of the People's Republic of China
Co-chair, Executive Committee, Group on Earth Observations
(Represented by LIAO Xiaohan)

It is my great pleasure to attend the 45th anniversary of CODATA. I want to send my congratulations as a representative from the China GEO office, and in particular I want to say that Mr. Cao Jianlin, the Vice-Minister of Science and Technology and also joint chair of GEO, is now in North America and my presence here is on behalf of him.

The mission of CODATA is to strengthen international science for the benefit of society by promoting improved scientific and technical data management and use. As we have seen, CODATA is providing a platform to international organizations and encouraging input from a variety of scientific data sources today. The success of international endeavors depends heavily on inclusiveness and cooperation, and the extent of that seen today is a good indicator of the future of our global data initiatives.

The ministry played an important role in international activity associated with the GEO mission. We try to pull together the numerous Earth observation systems of China domestically and those available from international sources to monitor a vast range of natural phenomena, including disasters, health, energy, climate, agriculture, ecosystems, biodiversity, water and weather. In this process, we inevitably face issues with coordination and compatibility, both in technical and political terms. This means that in the framework of GEO we have to communicate extensively with various operators and administrators even across cultures and disciplines, trying to fit them into a comprehensive Earth observation system, which will help a wide variety of users to share the Earth observation system data, giving a better understanding of the Earth system and benefitting human society.

To be truly inclusive and cooperative, rather than acting unilaterally, we must be willing to change our standards, re-align strategies, and formulate new goals. Reaching the kinds of international, interdisciplinary objectives CODATA and GEO both want to achieve is only possible through extensive, persistent cooperation. This can come in the form of formal agreements, joint research, sharing data, making policies or organizing conferences and meetings. Therefore I applaud CODATA's willingness to gather its partners together and not only allow them to voice their needs and concerns, but also use their input in their strategic planning.

A point I think has already been clearly made today is that good data underpins scientific development, and the availability of such data depends on perfect data policies and practices. I hope that CODATA can continue their tradition of excellence in helping to create effective data policies, especially in new initiatives like GEOSS. For this I give my best wishes to CODATA for its 45th anniversary, and for many more to come.

Thank you.
III. REPORTS AND PRESENTATIONS

CODATA — 45 Years On

GUO Huadong
President, CODATA

Abstract
In the years and decades after the Second World War, science and technology underwent significant development. A new kind of science was emerging to support space exploration, military applications, and geophysical sciences. High-tech instruments were pushing science into new frontiers with computations and simulations done electronically on computers. At the centre of all this was data—persistent, scientific data of ever-increasing volume. Along with this data came a need to manage it and put it to good use. This is the context in which CODATA was born. It is clear that the need for well-developed, globally-coordinated data strategies is no less pronounced today than it was decades ago at CODATA’s inception. Data and information sciences have matured into formal disciplines of their own with many sub-disciplines and areas of specialization. CODATA itself has matured with its evolution into a more strategic role supporting ICSU and international science. As CODATA continues forward, it will be looking to ICSU’s needs and strategies to help inform decisions on how to shape CODATA’s future. CODATA will need to help coordinate the data-related activities of organizations outside of and within ICSU. CODATA’s outreach and collaborative activities fall under what has been dubbed the Hand in Hand Program, whereby CODATA links together various international organizations through their common interest in and reliance on scientific and technical data. Researchers today need to be aware of the potential for discovery within these data. For scientists collecting extremely large datasets, they need to accept that they cannot always use that data to its full potential, and only by releasing it to the world and allowing others to build on it can it be thoroughly explored and exploited. This is CODATA’s place in science and society: doing all that is necessary to unlock the power of data.
CODATA in Context

Post-WW2

New technology

New science

Data processing and storage

Modeling

Programmatic analysis

Fortran

More technology, more data

IBM 350 RAMAC disk storage unit (over 4 MB)

REPORTS AND PRESENTATIONS

1966: CODATA Founded

ICSU

International Council for Science

Canada

West Germany

France

Japan

Soviet Union

United Kingdom

United States

CODATA National Members, 1968

Frederick Rossini, First CODATA President

Union Members, 1968

International Union of Biological Sciences
International Union of Pure and Applied Biophysics
International Union of Geodesy and Geophysics
International Geographical Union
International Union of Crystallography
International Union of Pure and Applied Chemistry
International Union of Theoretical and Applied Mechanics
International Union of Geological Sciences
International Astronomical Union
International Union of Pure and Applied Physics
**Task Groups before 1980**

- Accessibility and Dissemination of Data
- Computer Use
- Fundamental Constants
- Gas Phase Chemical Kinetics
- Internationalization & Systematization of Thermodynamic Tables
- Key Values for Thermodynamics
- Methodologies for Handling Space- and Time-Dependent Data
- Physical Property Data for the Chemical Industry
- Presentation of Data in the Primary Literature
- Training Courses on Handling of Experimental Data
- Transport Properties

**New technology/science/demands**

- Personal computers and the Internet...
- ...brought data-intensive science to new people and new places.

**Pioneering data science**

**Education and training**

- Task Group on Computer Use
- Freiburg Symposium on Man Machine Communication for Scientific Data Handling
- International Training Course in Data Dissemination

**Into the information age**

As CODATA sailed into the information age, at the helm for 20 years was Phyllis Glaeser.

Executive Director, 1979-1999
CODATA’s Mission Statement

“The mission of CODATA is to strengthen international science for the benefit of society by promoting improved scientific and technical data management and use.”

Focus on Science and Society

- Scientific Data and Sustainable Development
- Scientific Information for Society - From Today to the Future
- Scientific Data and Knowledge within the Information Society
CODATA Conference 2012: Taipei

“Open data and information for a Changing Planet”

CODATA 23
Taipei 2012

1. “Open” — Nurturing an open environment for data sharing and collaboration
2. “Data and Information” — Data-intensive science is needed to transform raw observations into information for applicable, intelligible results and discoveries
3. “for a Changing Planet” — Science is the key to providing reasoned solutions to climate change, population pressure, natural hazards, and other global issues

Problem-oriented Science: a spectrum of fields

Problem-oriented Science: a spectrum of fields

Data across borders and disciplines

Barriers to data sharing

Disaster Risk Reduction

Hydrology
Psychology
Geography
Data science
Civil Engineering
Geology

Technology
Organization
Politics
Education
Data across borders and disciplines

Barriers to data sharing
(main challenges)

Technology Organization Politics Education

CODATA Strategic Plan

3 Major Initiatives
- The Global Information Commons for Science Initiative (GICSI)
- The Scientific Data across the Digital Divide (SD3) Program
- Advanced Data Methods and Information technologies for Research and Education (ADMIRE)

Recent Activities
- Symposium on the Data Sharing Action Plan for GEOSS and the Benefits of Data Sharing
  November 2010, Beijing
- International Symposium on the Case for International Scientific Data Sharing: A Focus on Developing Countries
  April 2011, Washington
- International Association for Social Science Information Service and Technology
  May 2011, Vancouver
- WDS Conference
  September 2011, Kyoto
Recent Activities

- First volume, 2002
- Open access
- Electronic
- Publishes data or data compilations, online simulations, database, and other experiments

Outline

1. CODATA in Context
2. Contemporary CODATA
3. CODATA Continued...

Driving CODATA onward

- 11 task groups
- 2 working groups
- 20 national members
- 16 scientific unions
- 3 co-opted organizations
- 19 supporting organizations
- Executive Committee and Secretariat
- Biennial CODATA Conference
Prepare for data overload…

Promote:
- The importance of high-quality data
- Standards
- Distribution:
  - For Validation,
  - Collaborative analysis, and
  - Innovation

Stand on the shoulders of giants

- Release your data’s potential
- Others produce new knowledge from your data
- Data as an engine of discovery and innovation

CODATA: Unlocking the power of data

Thank you.
www.codata.org

Credits:
- Graphic of “The fanciful Automatic Coding System for the IBM 704”
  http://www.bhcausa.de/HTML/History/Languages/3474172305961074819d5110.html
- Edison Valley South blog
- http://esrbnews.com/3777/3975182/gallery1.0
- Photo by G. E. Kornblum (G. E. Kornblum, 1982)
- Map of internet: colored by IP address by William K. Cheswick
- The Economist, Feb 2011, 2010
I am very pleased to be with you today. I am grateful to the CODATA colleagues for having invited me to make this statement, for which I feel very honored.

This is a very valuable opportunity to meet and share with you some ideas on the collaboration between CODATA and IRDR. IRDR is the Integrated Research on Disaster Risk program of ICSU, also sponsored by the International Social Sciences Council (ISSC) and the United Nations International Strategy for Disaster Reduction (UNISDR).

I checked CODATA's website, consulted with my colleagues of the IRDR Scientific Committee and realize how important it is to develop a closer collaboration between IRDR and CODATA.

I see this collaboration as an essential step in the development of the IRDR program, in particular on the Disaster Loss Data project.

The Disaster Loss Data is one of the four main project activities of IRDR. Its implementation would greatly benefit from the experience of CODATA and the support of its members working on the same or related subjects. I will provide more information on it later.

Facilitating the identification, collection, analysis and dissemination of relevant data is a basic step in research for the development of new knowledge that serves to reduce risk related with natural hazards.

For IRDR, counting on reliable and relevant data is essential for guiding and supporting policy development and awareness raising efforts aiming at reducing the risk of disasters and building resilience of communities and nations towards natural hazards.

**Relevant Data For Relevant Purposes**

We know a myriad of technologies and processes, of which the development of cloud systems is a recent example, can make data collection, interpretation and application more demanding.

Similar to Marshall McLuhan that in 1964 stated that the ‘media is the message’, we can also say, with regards to natural hazards, that the ‘data is the message’. A single focus on natural hazards and disasters has allowed data to be biased towards natural events and the ultimate disastrous consequences instead of a more balanced approach when including data on social and human vulnerabilities that are the main cause of the disasters.

New technologies such as the e-cloud system can be extremely helpful in connecting experts and institutions working on common issues, in our case, linking research centers focusing on disaster risk reduction. We have a good example of connecting experts for a common cause and putting data together for a collective use; the Global Assessment Report on Disaster Risk Reduction of the ISDR system has been for the last 5 years such an exercise, bringing together databases from the World Bank, UNDP, UNEP, several universities to produce valuable biennial reports.

This requires however, the willingness of working together as a team, which is sometimes missing in international efforts that remain scattered, fragmented or not fully utilized. Working as a team requires pulling resources (not only data) together to work in a systematic manner with a long-term vision.

We also need to keep in mind the ultimate purpose of our work, the international goals of eradicating poverty and ensuring a sustainable development. For this we need to produce and disseminate knowledge that helps enhance the quality of life of populations living in hazard-prone areas. Accurate data to support this knowledge is therefore very important.

**DRR Depends On Accurate And Credible Models For Which Relevant Data Is Essential**

Disaster risk reduction also depends on accurate and credible models of risk. Modeling risk at different scales requires integrating vast quantities of data at different scales, relating to physical hazards of different kinds, the exposure of people and economic assets and vulnerability. For example, the global risk model that underpins the UN Global Assessment Report on Disaster Risk Reduction (UNISDR, 2009 and 2011) combines data on the intensity of floods, tropical cyclones, earthquakes and landslides, with data on the exposure of people and economic assets to these hazards and a set of over forty social and economic vulnerability proxies.

The validity of the results of this and other models is directly related to the quality and compatibility of the data used. The development of international data standards and protocols...
that facilitate data sharing and transfer, as well as techniques for dealing with uncertainty and data gaps, is therefore critical to the accuracy of risk models. Given that decisions are taken based on such models, the pricing of insurance for example, we are increasingly living in a modeled world. And thus issues of data quality and compatibility become ever more important.

Such models are often calibrated using data on reported disaster losses and impacts. The absence of internationally agreed standards for collecting and documenting disaster loss and impact data can have a serious impact on the quality of risk models. Trends and patterns of disaster loss, imputed from reported loss data without recognizing its inherent limitations, heterogeneity and uncertainty, can lead to false conclusions.

The IRDR Disaster Loss Data Working Group

The Disaster Loss Data Working Group is meant to study issues related to the collection, storage, and dissemination of disaster loss data. When human, monetary, or environmental losses occur as a result of a disaster, extensive loss data are collected and stored, but the thoroughness and accuracy of the data varies from country to country and even among local entities. Government agencies, private companies, and other organizations may collect and manage data related to their own areas of interest using their own standards and procedures, without significant collaboration with other groups. This results in gaps, inconsistent overlaps, and biases that ultimately affect the quality of research conducted and policies made based on the data.

To meet these needs, the working group intends to establish an overall framework for disaster loss data for all providers, establish nodes and networks for databases, conduct sensitivity testing among existing databases, and create mechanisms for archiving loss data.

To this effect, the Data Working Group has identified the following specific project areas:

• Identify what data and quality are needed to improve integrated disaster risk management.
• Bring together loss data stakeholders and develop and utilize synergies.
• Develop recognized standards, minimize uncertainty.
• Education of users regarding data interpretation and data biases.
• Ensure increased downscaling of loss data to sub-national geographical levels for policy-makers.
• Definition of “losses” and creation of a methodology for assessing it.

Finally, I would like to suggest that we explore the possibility for the IRDR Disaster Loss Data project to become also a CODATA project, to be carried out together so that both can benefit from each other’s capacities, networks and expertise. The co-location of IRDR with CEODE and having GUO Huadong as President of CODATA, provides an excellent opportunity to facilitate and nurture such collaboration.

I do look forward to an increased collaboration between our program for a more effective contribution to reducing the risk of disasters and building resilience to hazards that is required for eradicating poverty and achieving a sustainable development.

---

**Data-Intensive Science Discovery Practice in China**

ZHANG Xian’En
Director-General, Department of Basic Research, Ministry of Science and Technology of the People’s Republic of China

**Abstract**

Data have taken a prominent role in research, such as with the human genome project, GenBank, the Large Hadron Collider, and magnetic spectrometer data. Open access to scientific data is now an important part of the research process, and in China, progress is being made in this respect, especially through the Scientific Data Sharing Program. Open access to information requires sufficient technological support and scientific data infrastructure, but also supportive policies. Technology has seen many advances, but two other areas, law and regulation, have not been as successful. Balanced progress in all aspects will facilitate data-intensive science. Highlights of some of China’s scientific progress include work in disaster emergency monitoring, bioinformatics and systems biology, ecosystem monitoring, carbon cycle research, and Earth and global change observation. Scientific endeavors such as these need strong bases in theoretical research, which data-intensive science can provide. Enhancing theoretical research should be helpful to organize data in a way that allows scientists to look at the big picture and start answering broader questions of major concern to the world. In conclusion, data-intensive science, including the examples given and areas that have yet to be fully explored, requires infrastructure, policy frameworks, and technology to be fully realized.
Data-Intensive Science Discovery Practice in China

Xian-En Zhang 张先恩
China Ministry of Science & Technology
Chinese Academy of Sciences

How much information is there in the world?


The researchers calculate that humankind is able to store at least 295 exabytes of information. (That’s a number with 20 zeroes in it.)

- 2002 could be considered the beginning of the digital age, the first year worldwide digital storage capacity overtook total analog capacity. As of 2007, almost 94 percent of our memory is in digital form.
- In 2007, humankind successfully sent 1.9 zettabytes of information through broadcast technology such as televisions and GPS.
- In 2007, humankind shared 85 exabytes of information through telecom.
- In 2007, all the general-purpose computers in the world computed 6.4 x 10^18 instructions per second. Doing these instructions by hand would take 2,200 times the period since the Big Bang.
- From 1986 to 2007, the period of time examined in the study, worldwide computing capacity grew 68 percent a year, ten times faster than the United States’ GDP.
- Telecommunications grew 28 percent annually, and storage capacity grew 23 percent a year. The world’s technological information processing capacities are growing at exponential rates.

http://www.sciencedaily.com/releases/2011/02/110210141219.htm

The human genome Feb.15, 2001, entire issue of Nature with insights from the completion of the HGP working draft, led by Francis Collins

The human genome Feb.16, 2001, entire issue of Science with insights from the completion of the HGP and Celera working draft, Led by Craig Venter
Recent improvements in Sequencing technology have sharply reduced the cost of sequencing.

It is the first project to sequence the genomes of a large number of people, to provide a comprehensive resource on human genetic variation.

The samples mostly are anonymous and have no associated medical or phenotype data. The data will be used to study many diseases in sets of disease and control samples that have been carefully phenotyped.

Total samples are 2500, including 100 Chinese in Beijing, 100 Han Chinese South, 100 Chinese Dai in Xishuangbanna and 100 Chinese in Denver, Colorado.

The sequence and alignment data generated by the 1000 genomes project is made available as quickly as possible via mirrored ftp sites.

Data access links: EBI/NCBI

http://www.1000genomes.org/

Growth of GenBank

Oct 15 2011, NCBI-GenBank Flat File Release 166.0

- 144458648 loci
- 132067413372 bases
- from 144458648 reported sequences

Transcriptome shotgun data Increased 900% in the last two years

http://en.wikipedia.org/wiki/GenBank
LHC - The precise circumference of the LHC accelerator is 26,659 m, with a total of 9,330 magnets inside. The working temperature is down to -271.3° C (1.9 K), with an ultra-high vacuum (the internal pressure of the LHC is 10^-13 atm, less than one ten millionth the pressure on the Moon). Two beams of protons will each travel at a momentum of 7 TeV (tera-electronvolts), corresponding to head-on collisions of 14 TeV. Altogether some 600 million collisions will take place every second, generating temperatures more than 100,000 times hotter than the heart of the Sun.

Data recorded by LHC and its simulation was estimated at approximately 15 petabytes (15 PB) per year. The ATLAS Computing Grid was designed from scratch to handle this massive amount of data and hosted in the Open Science Grid.

Magnetic spectrometer data are daily returns in the 1–10 million
Analysis of the data on the magnetic spectrometer has been started. The main institutions involved in the analysis of the data include MIT, University of Aachen, Germany, Italy and France as well as research institutions in China, Southeast University, Zhongshan University and Shandong University.

Professor Samuel C.C. Ting

The launch of Endeavour was named after May’s report, after long-waiting LHC data and HEP.
- Development plan for SDSP
- Standard specification system, 263 standards and criteria
- 3000 databases for basic research and public welfare, 200 institutions, 140 TB data
- SDSP system, 1 homepage, 10 data centers and service network, 100 branches and nodes
- 2009, 170,000 registered users, 62 M visits, 430 TB download
- Data Collection and submission under 973 program
- International collaborations
- Law and regulations

http://www.sciencedata.cn/
A few highlights

Argo project

- **Argo** is an observation system for the Earth’s oceans that provides real-time data for use in climate, weather, oceanographic and fisheries research.
- Argo consists of a large collection of small, drifting oceanic robotic probes deployed worldwide. The probes float as deep as 2 km. Once every 10 days, the probes surface, measuring conductivity and temperature profiles to the surface. From these, salinity and density can be calculated.
- The data are transmitted to scientists on shore via satellite. The data collected are freely available to everyone, without restrictions.

The focus of Argo 973 program is on the northwest Pacific and the tropical Indo-Pacific Oceans, which are key areas that have strong influences on global and regional climate systems.
2.2 Geology disaster after earthquake

2.3 House collapses monitoring

Yingxiu house collapses
Beichuan house collapses

For man interpretation
For auto extraction with texture and spectrum feature

Jianjiang river has 96 collapses, area: 363271 m², 106 landslides, area: 903809 m², 236 debris, area: 15096227 m².

Minjiang river has 253 collapses, area: 632079 m², 534 landslides, area: 2739086 m², 525 debris, area: 23673497 m².
-omics, bioinformatics and systems biology

- **Systems biology** is “biology” that focuses on complex systems in life.
- **Omics** focuses on large scale and holistic data/information to understand life in encapsulated omes, such as genomics, proteomics, transcriptomics, epigenomics, meta-genomics, etc. The main focus is on 1) mapping information objects such as genes, proteins, and ligands, 2) finding interaction relationships among the objects, 3) engineering the networks and objects to understand and manipulate the regulatory mechanisms, and 4) integrating various omes and omics subfields.
- **Bioinformatics** is an information science that analyzes life processes using computational tools for solving biological problems and give direction/overview in biology.

Biological data reach up to Pbytes!!!

China’s Genome sequencing

- Human Genome 1%, 2001, Nature
- Rice Genomes, 2002, Science, Nature
- Bacterial genomes, 2003, Nature
- Chicken genome, 2004, Nature
- C. elegans genome, 2007, Nature
- Cucumber genome, 2009, Nature Genetics
- Silk worm genome, 2009, Science
- Human pan-genome, 2009, Nature Biotech
- Panda genome, 2009, Nature
- Potato genome, 2011, Nature
- Enterotypes of human gut microbiome, 2011, Nature
- Genome of tree of life,........
- In 2011, BGI publish the results in CSN once a month
China Establishes National Genmoe Database to House Output from BGI
June 24, 2011 By Udum Grace Thomas

China’s National Development and Reform Committee is creating the
database in collaboration with BGI. The partners expect the resource to be
"one of the world’s largest gene banks," Yang Huanming, BGI’s president,
said in a statement.

Qi Cheng Yuan, Chief of the NDRC, stated that the resource is based on data
and facilities belonging to the BGI, but is expected to grow with the help of
"extensive cooperation with other biological organizations both at home and
abroad.

BGI is currently the world’s largest genome center and houses more than 130
Illumina HiSeqs, around 30 Life Tech SOLIDs, and more than 100 Sanger
sequencers. The institute generated about 500 terabytes of next-generation
sequence data last year. Its current output is around 5.6 TB of data a day, a
spokesperson said.

BGI will continue to submit data to NCBI, and other databases for use in
research publications.

NONCODE: decoding the non-coding

Non-coding RNAs (ncRNAs) are functional RNAs that are not translated into proteins.

NONCODE is a database of all kinds of noncoding RNAs. It is
distinguished from other ncRNA databases by:
1) big data amount; 2) 80% data are from experiments; 3) new
classification based on the cellular functions; 4) efficient search option.

In NONCODE v3.0, now it contains 423,976 public sequences
from 1,239 organisms covering all kingdoms of life, including
134 traditional classes, distributed in 26 cellular process. For
each entry in NONCODE one can: Get text information such as
class, name, location, related publication, mechanism through
which it exerts its function. NONCODE links to other databases.

Massive ecological monitoring data on Chinese Ecosystem Research Network (CERN)

- Field observation
  - Long-Term
  - +100 stations
- Monitoring data
  - Real-time Data Stream
  - Ecology, Environment, Space, etc.
  - Data record rate:
    Over 100 million/Yea
Data intensive science practice

• e-Carbon
  – e-Science Environment for Carbon Budget Evaluation Research of Chinese Terrestrial Ecosystem

• Earth Observation and Global change

• HEP Grid in China
  – Data Intensive Grid Computing for High Energy Physics

• Data Intensive Life Science

E-Carbon: carbon cycle research

- More Data
- More Computation
  - Data processing
  - Modeling
  - Data assimilation
- Faster Assessment
- More interplay and collaboration

E-Carbon-Data acquisition

- Many data acquisitions
  - Atmospheric Inversion
  - Remote sensing
  - Field experiment
  - Flux observation
  - Carbon modeling
  - ...
E-Carbon-Data characters

- multiple sources, different formats
- Including Real-time data stream
- Data record rate: **Over 100 million/Year**

Earth Observation and Global change

- Earth Observation data is in the situation of explosive growth
  - Satellite Observation
  - Aviation Observation
  - Ground-based Observation
  - In-situ data and so on

Scientific simulation

The scale of data has reached PB and tends to extrascale (1000PB)

Requirements of spatial data are increasing quickly

- Global Change and disaster monitoring need the support of spatial data with multi-resolution, multi-mode and multi-temporal
- Distributed spatial data needs to be integrated for applications
Earth Observation and Global change

The Digital Earth Development

data and products
remote sensing and GNSS
sensor webs
users and consumers
social networks
value adders

technology
broadband and internet
computer memory
processing speed
processing cost
software

development
servers
software
virtual globes

reference framework

users
professional
research
lay

Future driven for Digital Earth

Global Change Project

Projects on Global Change study

• In 2009, the first 973 project focus on global change has been launched: Mechanisms and Methods for Space-based Observations of Global Change-sensitive Factors

Conclusion

• Technology in future: Data intensive science
  – Need reliable and easily accessible scientific data infrastructure
  – Need data policy framework to promote data management and broad use
  – Need new technologies to support scientists processing data across scales and disciplines

• Theory science in future: Data Science
  – Enhancing theory research of data intensive science should be helpful to organize data for broader questions not for a few questions
  – Computer Science - acquire and parse data
  – Mathematics, Statistics, & Data Mining - filter and mine
  – Graphic Design - represent and refine
  – Infovis and Human-Computer Interaction (HCI) - interaction

• In 2010, 31 national projects are started, focus on the EO for global change
• Different study area on Global change
  – Data acquisition, simulation, application
  – 300 million Y support for global change plan in China per year
Abstract

Understanding the genetic underpinnings of disease is important for screening, treatment, drug development, and basic biological insight. Genome-wide associations, wherein genetic markers are systematically scanned for association with disease are one window into disease processes. Additionally, finding out which parts of our DNA affect particular intermediary processes such as gene expression (eQTL) can shed light on important pathways. Naively, these associations can be identified using a simple statistical test on each hypothesized association. However, a wide variety of confounders lie hidden in the data, leading to both spurious associations and missed associations if not properly addressed. My talk will focus on a class of statistical models, linear mixed models (LMMs), which correct for these confounders.
Analysis of Large-Scale Genomic Data

Jennifer Listgarten
Microsoft Research
Los Angeles

Talk Outline
1. Importance of Genetics Studies
2. Basic Terminology and Types of Studies
3. Advances in Analysis of Genome-Wide Association Studies

Importance of Genetic Studies
1. Personalized Medicine Has Arrived
   Adapting treatments to a person’s genetic make-up.
   - Targeting patients who can benefit (e.g., 10% of people cannot respond to codeine), and not develop toxicities (e.g., Abacavir bad for HIV patients with HLA-B57).
   - Appropriate dosage of a drug by using genetic variants to understand drug metabolism (e.g., anti-depressants, beta blockers, opioid analgesics).
   - More drug approvals because can target the right sub-group based on genetics.

Importance of Genetic Studies
2. Disease Sub-Classification aka “Molecular Classification”
   Previously unknown sub-classes of a disease means treatments can be refined resulting in improved outcomes (e.g., pre-leukemia/MDS).
3. Risk Identification
   If patient is at risk of developing disease (e.g., colorectal cancer), can be more aggressive in monitoring that patient to achieve early detection and better outcome.
Importance of Genetic Studies

4. **Basic Biological Insight**
Discovery of biological pathways, how different parts of the system are interacting, what is causal vs. what is downstream effect.

*E.g.*, GWAS uncovered pathways underlying Age-Related Macular Degeneration and Crohn’s disease; Merck obtained obesity and heart disease leads.

### Basic Terminology

Genetic variants such as *SNPs*—single nucleotide polymorphism (*discrete: A/C/T/G*).

- **Amount of gene expression** (*real-valued number: -23.1, +42.6*).
- **Phenotype of interest**: *e.g.*, *hasAsthma vs. doesNotHaveAsthma* (*discrete*), *e.g.*, blood pressure (*continuous*).

### e.g. Types of Studies

1. **“GWAS” Genome-Wide Association Study**
   - Association between DNA markers and target phenotype.
   - *E.g.*, ask whether each of 1 million SNPs are correlated with asthma.

2. **“Differential Expression”**
   - Association between gene expression and target phenotype.
   - *E.g.*, ask whether each of 200 gene expression levels are associated with asthma.
**e.g. Types of Studies**

3. **"eQTL"** (genetics of gene expression)
   - Association between gene expression and target phenotype.
   - *e.g.*, ask which of 1 million SNPs are correlated with which of 20K gene expression levels (1 million x 20K hypotheses!)

**Statistical Challenges**

1. **Large scale multiple hypothesis testing (aka “Winner’s Curse”)**: Throw out traditional measures of statistical significance instead using more appropriate measures, *e.g.*, False Discovery Rate.

2. **Data heterogeneity (aka hidden confounding structure)**: Some people are more similar to each other than other people because of:
   - Relatedness (*e.g.*, families)
   - Ethnic diversity (*e.g.*, African, Caucasian, etc.)
   - Causes spurious results and lack of power

**Statistical Wins**

Large scale of data means:

A. We can *see* problems in the analysis (quantile-quantile plots of p-values).
   - Not possible before large-scale hypothesis testing

B. We have enough data to *infer* hidden confounding structure such as relatedness and ethnicity and more.
   - Not possible before large-scale data sets.

**Genome-Wide Association Studies (GWAS)**

**Input:**
- A set of people with/without a disease
- Measure a large set of genetic markers for each person (*e.g.*, SNPs).

**Desired output:**
- A list of genetic markers underlying the disease.
Recent Explosion of GWAS Successes

GWAS Pitfalls: Hidden Structure

Fundamental assumption in most models is that the subjects are sampled identically and independently from the same distribution.

**Then if** subjects:
- Spurious correlations induced giving spurious hits.
- Cross over different ethnicities.
- True signal swamped out reducing power to detect true effects.
- Search for idiosyncratic effects (processed slightly differently, and not at random).
- Contain minimum we don't yet know about.

Intuition of Hidden Ethnicity Structure

- Suppose the set of cases has a different proportion of ethnicity X from control.
- Then genetic markers that differ between X and other ethnicities in the study Y, will appear artificially to be associated with disease.
- These spurious associations can swamp out the true signal of interest, or induce spurious associations.

- The larger the study (# people), the worse the problem, since the power to detect spurious’ signal increases.
- But large studies are needed to detect markers with weak effect.
**Statistical-Challenges → Statistical Win**

- Evidence of problematic analysis.
  - Expect very few markers to truly be associated with disease.
  - Distribution of p-values should be close to a uniform (diagonal) on this plot.

- Statistical-Challenges → Statistical Win
  - *Evidence* of improved analysis.

**Statistical-Challenges → Statistical Win**

- Use the large scale of the genetic data to see how ‘similar’ every two people are, and incorporate this into the analysis.

*genetic similarity matrix*

- three ethnicities
- families
- no structure

Best current approaches are:
1. Principle Component Analysis
2. Linear Mixed Models

**Naïve Approach → Linear Regression**

- Regress target against genetic marker.
- e.g., regress blood pressure level on SNP (and do for each SNP).
- Evaluate SNP by comparing this model to one without SNP, e.g. use LRT statistical test.

$$y = X\beta + \varepsilon$$

- blood pressure level
- SNP
- learned regression weight (importance of SNP to blood pressure)
**Principle Components Analysis Approach**

- Project each person’s markers into the low dimensional space capturing confounders.
- Add projections as covariates in a standard regression analysis that looks for associations between marker and phenotype.

\[ y = X\beta_1 + P\beta_2 + \epsilon \]

*Projection in low-dim space*  
*Learned regression weight*

*Price et al. Nature Genetics 2006*

---

**Rich Information Contained in Genetic Similarity Matrix**

- Do *not* reduce space to a set of directions. Use it in its entirety.
- Regress on hidden regression coefficients within a standard regression analysis.

\[ \mathbf{u} \sim \text{Normal}(\mathbf{0}, \mathbf{I}) \]

\[ \mathbf{y} = X\beta_1 + \int u\beta_2 du + \epsilon \]

*o.g. Yu et al Nature Genetics 2006; Kang et al Genetics 2008*

---

**Mixed Model Approach**

- Works well to capture broad structure.
- Sensitive to outliers (bad!).
- Cannot capture fine-grained structure (bad!).
- Fast computations (good!)

*Works well to capture broad structure.*

*Sensitive to outliers (bad!).*

*Cannot capture fine-grained structure (bad!).*

*Fast computations (good!)*
Compare Performance on Real Data

Mixed model works better than PCA approach here.

~7500 SNPs, ~1300 people: variety of ethnicities (~500 caucasian, ~200 hispanic + ~200 african american) and families within these

Summary: Large-scale genetic studies

• Large scale of the data inherently leads to problems:
  – “Winner’s Curse”
  – Data heterogeneity which causes spurious results.

• But… the scale of the data also let’s us:
  – Diagnose analyses.
  – Infer and correct for confounding signal.
  – We’ve also done work in this space on the genetics of gene expression [PNAS 2010].

Mixed Model Approach

➢Captures multiple levels of similarity:
  broad- and fine-grained (good)

➢Not sensitive to outliers (good)

➢Computationally expensive
  •Cannot analyze data sets > 13K individuals with current methods.
  •We developed a new approach that is dramatically faster (have analyzed 120K).
    [Nature Methods, 2011]

Mixed Models are Expensive

• Scale badly with number of individuals (cubically).
• Large data is here now:
  – e.g., 23andMe, Kaiser Permanente, DeCode (~100K individuals)
Mixed Models are Expensive

Focus in the genetics community on speeding up these models

- Technical so won’t go into details
  - still unable to analyze large data sets (>13K)

- Our recent work broke the barrier of analysis of extremely large data sets [*Nature Methods*, Sept 2011]
  - analyzed 120K individuals
  (Joint work with Microsoft colleagues: Christoph Lippert, David Heckerman, Bob Davidson, Carl Kadie, Ying Liu).

Suite of Related Problems

   (SNPs $\rightarrow$ phenotype)

   (gene expression $\rightarrow$ phenotype)

   (SNPs $\rightarrow$ gene expression)

Experimental running time and memory

---

**eQTL Analysis**

Can think of eQTL analysis in two ways:
1. One **GWAS** analysis for each gene probe.
2. One **Differential Expression** analysis for each SNP.

Need to properly deal with:

i. **Confounding by Population Structure and Family Relatedness** (as required in GWAS)
ii. **Confounding by expression heterogeneity**: technical, environmental, demographic, or genetic factors [Leek & Storey *PloS Genetics* 2007; Kang Genetics 2008].
**Double Correction Model: PS+EH**

1. **SNP-space** confounder coefficients
   
   \[ p(\tilde{u}) = N(\bar{0}, \sigma^2) \]

2. **Gene-expression-space** confounder coefficients
   
   \[ p(\tilde{v}) = N(\bar{0}, \sigma^2) \]

\[ p(y | \tilde{u}, \tilde{v}) = N(X\beta + \eta_{PS}\tilde{u} + \eta_{EH}\tilde{v}, 1\sigma^2) \]

\[ p(y) = N(X\beta, \eta_{PS} + \eta_{EH} + 1\sigma^2) \]

**Results on Mouse Data**

- **Linear Regression**
- **LMM-PS**
- **LMM-ICE-PS**

**How to fix the ICE+PS model?**

Instead of computing ahead of time as a pre-processing step, learn it jointly with the other parameters of the model so that we have a **consistent** estimate of it.

\[ p(y) = N(X\beta, \eta_{PS} + \eta_{EH} + 1\sigma^2) \]

**How to fix the ICE model?**

From the graphical model one can see that \( K \) should be equal to the covariance of the hidden variables \( u \) (not to the covariance of the gene probe data).
In the post-genomic era, the opportunities and challenges of medicine will be to acquire a deep understanding of disease mechanisms and pathophysiology at the molecular level by integrating genomic, proteomic, and other “omic” data. The confluence of these disparate data presents opportunities to generate information on the complex biological processes that occur in individual cells. The challenge is to develop the capability to collect large volumes of data in parallel and the informatics tools to analyze the data.

One strategy to investigate disease mechanisms is to understand a “healthy state” versus a “disease state” of cells. An active area of research focuses on the identification of the signatures or biomarkers for a healthy state so as to recognize dysfunctions associated with disease. Biomarkers express messages from cellular functions and processes and serve as early warning signals for disease before clinical symptoms appear. For example, numerous proteins are over-expressed on the surface of cancer cells, including epithelial cell adhesion molecules, epidermal growth factor receptor, etc. By taking advantage of such over-expression, one can devise a high-throughput microfluidic chip to detect one lung cancer cell among a billion normal blood cells. This is a powerful tool that can be viewed as a liquid biopsy. These cells can be visualized with specific fluorescent labels for cell surface, nuclear stain, and proliferating markers. Imagine what we might learn if we could capture in images the potential interactions of these circulating tumor cells with neighboring cells. What might we learn about metastatic processes? What innovative treatment modalities might we develop for cancers that reflect the deep understanding of metastasis?

With this paradigm of molecular medicine, we hope to predict risks, identify patient subtypes, and personalize treatments based on the molecular characterization of disease, as well as to predict treatment outcomes.
Molecular Medicine: A Confluence of Data

Belinda Seto, Ph.D.
Deputy Director
National Institute of Biomedical Imaging
and Bioengineering

Molecular Medicine: Understanding the Complexity of Biological Systems

- Multiple cross-talks between molecular pathways
- Many molecular targets in intricate signaling networks
- Novel experimental and theoretical concepts needed

Molecular Medicine: Grand Challenges

- Single cell: characterize normal cells so as to identify diseased cells
- Single cell genomic, proteomic, metabolomic, spatial and temporal dynamic
- Parallel data collection using advanced mathematical and computing approaches

Molecular Medicine: Present and Future

- Biomarkers for disease states
- Molecular theranostics: combined diagnostics and therapy
- Integration of patient test results, genomic data, medical histories, disease mechanisms and predictive models for health and disease to determine risks and improve treatment outcomes
Two Stories

- Influence of genetic variations, at the level of single nucleotide polymorphisms (SNP) on whole brain structure and function
- Circulating tumor cells: detecting 1 lung cancer cell in a billion blood cells

Genome-wide Association Studies (GWAS): Purpose, Goals

- To identify common genetic factors that influence health and disease
- To study genetic variations, across the entire human genome, that are associated with observable traits
- To combine genomic information with clinical and phenotypic data to understand disease mechanism and prediction of disease
- To develop the knowledge base for personalized medicine

Combined Neuroimaging and Genome-wide Association Study of Alzheimer’s Disease

Genetic Variants: Influence on Brain Structure and Function

Why do we need to know genetic variants?

- Genetic variants influence disease risks: Alzheimer’s disease, other neurodegenerative diseases, obesity, diabetes, etc
- Improve early detection
- Target treatment: individualized medicine
- Influence brain structure and cognition
Correlation between Genetic Variants and Brain Structure

- Search the entire human genome
- Examples: glutamate receptor gene was over-represented in Alzheimer’s disease and associated with 1.5% lower temporal lobe volume
- Extend this genome-wide search and apply to each location (voxel) in brain image

Confluence of Data

- Genotypes
- Brain images
- Vast amount of different types of data, huge computational load
- Statistical methods for multiple comparisons: linear regression of each voxel (4x4x4 mm³) and each genetic variant (single nucleotide polymorphism, SNP) and healthy group as an average template image

Correlation Between Genetic Variation and Regional Brain Volume Changes in Alzheimer’s Disease

Paul Thompson, Art Toga, et al., UCLA (EB002173, EB002221, EB006437)

- 450,000 SNP’s were correlated with regional brain volume changes in this “High-Throughput Screening” study of 740 subjects from the Alzheimer’s Disease Neuroimaging Initiative (ADNI)
- Examples of correlations between specific SNP’s and volume changes in different brain regions are given below.
- These early results are promising for identifying genetic factors in the diagnosis and treatment of Alzheimer’s Disease.

Circulating Tumor Cells: Key to Metastasis?

Cell Capture Technology
Optical Imaging of Single Cell
Molecular Characterization
**Automated Image Processing**

Program Output | High Magnification Images
--- | ---
DNA | PSA | Merged

**Taking Imaging Beyond Enumeration**

Novel classification schema for CTCs using cross-correlation image processing algorithms

- DNA
- PSA
- KI67
- M30

**Questions to be addressed:**

1. Can CTCs be detected in the blood of patients with EGFR- and ALK-mutant lung cancer?
2. Do CTC numbers change with treatment?
3. Can tumor-specific mutations be detected in CTCs?
4. Can signaling be measured in CTCs (and does it change with targeted therapy)?
5. Do the signaling effects in CTCs match those of the primary tumor?
6. Does this predict clinical outcome?
**CTC Numbers Track with Disease Course**

![Graphs showing CTC numbers tracking disease course](image)

Nigrath et al., *Nature* 2007

---

**Iressa Sensitivity in Lung Tumors**

- Mutations in tyrosine kinase are commonly found in lung tumors sensitive to Iressa
- Mutations adjacent to lysine 745, critical ATP-binding site
- A substitution of Arginine for Leucine adjacent to the highly conserved DFG motif (aa 855-857) in the activation loop
- Mutations are commonly found in lung tumors from “never smokers”

---

**Mutations in EGFR Leads to Greater Drug Sensitivity**

- Detection of primary sensitizing EGFR mutation (del, L858R, others)
- Acquisition of secondary drug resistance (T790M-EGFR)

![Diagram showing EGFR mutations and drug resistance](image)

Yuan et al., *Science* 2003
Park et al., *PloS A* 2003
Wells et al., *PNAS* 2000

---

**CTCs in a patient with EGFR-mutant lung cancer decrease in number within 2 days of starting erlotinib (and remain suppressed to 3 months)**

![Graph showing CTC decline](image)
Earth science is an all-embracing term related to planet Earth, which is recognized as a complex system consisting of several spheres or subsystems. Earth sciences cover a broad range of fields studying the atmosphere, biosphere, lithosphere, hydrosphere, etc., all of which are heavily dependent on data. Scientific data are crucial to understanding our changing Earth, and the need to collect and manage data has given rise to data science. Scientists must work with various kinds of data from field observations, experiments and simulations, and remote sensing, each with its own advantages, disadvantages, and varying degrees of error. Such error may be introduced into data through a number of ways. Meteorological observations, for example, may contain errors or inconsistencies due to changes in equipment, sites, measurement time, or measurement method, influencing the homogeneity of long-term observations. Therefore, intercalibration and validation of multi-source data are necessary steps in the data collection process. One example of a multi-source data integration strategy is that of the USDA UV-B Monitoring and Research Program, which incorporates data from a ground observation network and satellite remote sensing. Such systems have potential for producing more reliable results in applications like crop yield estimation or air quality modeling. Applications that aim to address large-scale challenges, such as global change, require an integration of data and models from multiple spheres in consideration of the complex interactions taking place in the Earth system.
Scientific Data and Earth Science
— With application examples

Gao, Wei (高炜)
Colorado State University
East China Normal University

Background

✧ Earth Science

is an all-embracing term for the sciences related to the planet Earth, which is recognized as a complicated system consisting of several spheres (subsystems).

✧ Earth’s spheres and major earth science topics

✧ Atmosphere
  - Atmospheric chemistry
  - Climatology
  - Meteorology
  - Paleoclimatology

✧ Biosphere
  - Biogeography
  - Paleontology
  - Geomicrobiology
  - Geoarchaeology

✧ Lithosphere (Geosphere)
  - Geology
  - Geophysics
  - Geography
  - Geochemistry
  - Geomorphology
  - Glaciology
  - Hydrogeology
  - Mineralogy
  - Petrology
  - ...

Outline

1. Background
2. Data
3. Applications
Data Intensive Science and Discovery — CODATA 45 Years On
Summary Report

REPORTS AND PRESENTATIONS

**Background**

- Earth’s spheres and major earth science topics
  - Hydrosphere
    - Hydrology
    - Hydrogeology
    - Oceanography
    - Marine geochemistry
  - Pedosphere
    - Soil science
  - Atmosphere
    - Reductionist approaches
    - Holistic approaches
  - Geography

**Outline**

1. Background
2. Data
3. Applications

**Data**

- Data Science
  - bits and bytes
  - meaning and impact of information
  - extended complexity science

... not only the stylized patterns and dynamics resulting from the non-linear interaction of simple elements. It also needs to understand the result of interactions between individuals with cognitive complexity or system elements with a complex response to the surrounding world.

- data revolution

... Such systems with various levels of complexity are probably not analytically tractable and, therefore, require the use of future supercomputers.

(The FuturICT Consortium)
What kind of earth science data do we have?
- field/in-lab experiment data
- long-term observation data
- remote sensing data

Long-term observation data
- earthquake
- hydrological records
- meteorological stations
- terrain deformation
- sea level
- ...

Field/in-lab experiment data
- present status (e.g., soil texture, soil moisture ...)
- past status (e.g., ice core, geological profile ...)
- high accuracy
- limited research area
- limited dynamic measurement

Remote sensing data
Satellite (Spaceborne platform)
- atmosphere / ocean / land
- limited retrieval accuracy
- large spatial coverage
- dynamic observation
**RS data processing chain**

- **Data to Product**
  - L0 to L1
  - L1 to L2

- **Product to Application**
  - L2 to L3
  - L3 to L4

- **Preprocessing**
  - Classification (qualitative info)
  - Retrieval (quantitative info)

- **Data assimilation**
  - Modeling (coupling)
  - Simulation

**RS data processing II - Retrieval**

*Retrieval* is a basic technique to extract quantitative information of our Earth from remotely sensed data.

**RS data processing III - Data assimilation**

*DA* is the combining of diverse data, possibly sampled at different times and intervals and different locations, into a unified and consistent description of a physical system, such as the state of the atmosphere, ecosystem, etc.
Example: Regional Data Assimilation using Direct Broadcast (DB) Satellite Remote Sensing Data

- Data

72h NWP with a regional DA system based on direct broadcast MODIS data

- Data

Errors must be contained in observations and data.
- Meteorological observations
  Change of equipment / site place / measurement time / measurement method … may influence the homogeneity of long-term meteorological observations.
- Satellite observations
  Sensor calibration, channel frequency shift, device noise, sensor update and data processing (eg. inaccuracy atmospheric correction, retrieval algorithms) may cause errors.

Inter calibration and validation of multi-source data are necessary.
**Multi-source Data Integration Strategy**

- Example: Ground monitoring network (bottom-up)
- Satellite remote sensing (top-down)

**Measured irradiance**
- UV (Langley calibrated)
- UV (Lamp calibrated)
- Visible (Langley calibrated)
- Erythemal
- PAR

**Derived products**
- Daily sums
- UV index
- Synthetic spectra
- Column ozone
- Instantaneous optical depths
- Average optical depths
- UV irradiance estimator

**Feature: Separate Direct & Diffuse irradiance**

Automated rotating shadowband

Direct irradiance + Diffuse irradiance = Total irradiance

Data

USDA UV-B Monitoring and Research Program
37 Climatologic Sites and 5 Research Sites

USDA CSU

Data

Typical one day graph of irradiance

Column ozone
Data-Intensive Science and Discovery — CODATA 45 Years On

Summary Report

REPORTS AND PRESENTATIONS

Data

- From UV band to IR band...

- Integration of bottom-up station observations with top-down RS data

Data

- Measurements and retrievals

- Optical depth by AERI
1. Background

2. Data

3. Applications

Data → Model → Applications

Applications

Case 1: Crop yield estimation

Integrated Agricultural Impact Assessment System:
- couple the Earth’s climate and comprehensive crop growth models
- assimilate satellite and in situ observations
- Study climate-crop interactions and related yield and economic impacts stemming from crop response to a wide range of global change stressors.
Applications

- Improved coupling of climate and cotton models using realistic climate conditions,
- Good agreement is obtained between the observed (top) and simulated (bottom) cotton yields,
- The differences are within ±12% at most of the sites in the US cotton belt.

Reduction in yield for cotton due to ±25% precipitation change

- Upper images are change as bilacre
- Lower images are change as %

From left to right, the color bar shows percent REDUCTION.
Case 2: An air quality research model at ECNU with direct broadcast RS data

Features:
- Near real-time
- Easy to update
- Physically based
- Predictability

48h simulation of an air pollution event in Shanghai
Case 3: Satellite/ground observations of CO₂

Satellite remote sensing

Ground observation

Different satellite CO₂ products

- Different coverage
- Different values
  because of
  - Different weighting functions
  - Different retrieval algorithms

Correlation of AIRS CO₂ and ground observations

0 - 0.5
0.5 - 0.6
0.6 - 0.7
0.7 - 0.8
0.8 - 0.9
0.9 - 1

AIRS CO₂ and ground observations at Wuliquan, China from 2002 to 2010

Integration of satellite and ground CO₂ data to support global change researches...

More...

Climate-Hydrology interaction

Assessing Climate Change and Extreme Weather in the Urban Environment: Helping Cities Worldwide Prepare for the Future
Large-scale data production has become essential in conducting scientific research in a variety of disciplines. This is mainly due to technological advancements of useful methodologies. Large data-driven science has changed the scientific paradigm, which eventually affects the future of our society. For example, the so-called Next-Generation Sequencers (NGS) have an enormous capability of determining nucleotide sequences of genomic DNA with amazingly high cost- and time-efficiencies. In fact, the NGS has dramatically changed the paradigm of genome-related science but also the perspective of our society. Personalized medicine on the basis of genomic information has almost become a reality. Moreover, the so-called meta-genomics will provide us with marine environmental monitoring, revolutionary approaches to intelligent agriculture, and prompt detection of emerging and reemerging viruses and bacteria in the air. In this situation, we should be able to make a proposal of the vision of our future society focusing upon the genome and its related information in particular. Thus, I would like to call it the "Genome Information-oriented Society (g-Society)". This can be achieved by innovations based upon large data-driven science. It is no doubt that databases have an essential role, in many ways, in the formation of the g-Society.
Vision from Data-Intensive Life Science: GENOME INFORMATION-ORIENTED SOCIETY

Digital Earth Demonstration Hall, CEODE, CAS
Beijing, China
30 October, 2011
Takashi Gojobori
Center for Information Biology and DDBJ
National Institute of Genetics (NIG), Mishima, Japan
\textbf{Genome} = \textbf{Gene} + \textbf{Chromosome}

- Coined by Hans Winkler (1920)
- Modified in functional context by Hitoshi Kihara (1930)
Commitment to International Collaboration

GenBank
NCBI (USA)
Since 1982

EMBL
EBI (Europe)
Since 1980

DDBJ
DDBJ (Japan)
Since 1986

~Information~

About 2.4 billion nucleotides have been stored in the DDBJ/EMBL/GenBank for 25 years.

~Revolution~
GnuBiG Enters Next-Generation Sequencing Sweepstakes
June 2, 2010, By Kevin Davies

BOSTON — The latest entry in the next-generation sequencing sweepstakes is a public-domain sequencing genetics platform. GnuBiG is a company based on the technology of David Wertz, software professor at MIT. Its partnership with GnuBiG did not avoid an ostensible or even slightly detailed sequencing data. The platform is widely used and modestly but not definitive a microfluidic platform that creates hundreds of millions of reads in a single run for some very affordable DNA sequencing.

The Wertz group says its technology is a very quick and efficient DNA sequencing platform. The maps are sent to next generation sequencing methods that yield a 10-fold increase in efficiency can contain 10™giga maps.

The process occurs in a 1 millisecond per base, and using the DNA sequencing platform, the reads are sorted and sorted in parallel.

For the sequencing applications, the DNA sequencing platform, and each lane can run into larger drops (nanoliter volumes). In parallel, those techs and chemists could do it in 300 bases. The data was paired and sent to 15 of 1900 channels.

Wertz predicted that the instrument will cost $20,000 and the technology will cost about 10,000.

The instrument would cost a mere $20,000 and take about 10 hours to complete 10,000 base pairs. This means that the orders of magnitude are not that far off. That’s what Wertz says: “The technology is not just in the postdoc and grad students.” Wertz has co-founded a start-up biotech called GnuBiG.

A small group of people have delta systems ready to the end of 2010. There are already commitments to purchase instruments from the company.


~Omics~

Post-genome epoch: Future perspectives

23,148 genes

~40,000 genes

Organism → Health clinics

Phenotype → Phenomics

Molecule → Omics

Cell → Genomics

Tissue → Transcriptomics

Gene → Genomics

Genome → Regulome

Interact → Interomics

Localizome → Integrative Genomics

Metabolomics → Genotyping

Phenome → Genotyping

Gene → Genotyping

Genome → Genotyping

[Space] Apollo/17からみた地球(NASA)
~Problem~

- Problem -

*Sydney Brenner says, “Low input, High throughput, and No output”!

Issues on retrieving the necessary information

(The Goal) Lack of the standard format without unified information often hinders research and development seriously

Where is the information needed?

How about the relations between each information?

Too many databases......

Be Unified and Easily Retrievable Format from Sporadic Information!

Which one to use?

"Submission only" Database

Accelerating R&D

"Unified and Easy Retrievable" Database New Model

Databases with various formats

This is a right one to use!
Data-Intensive Science and Discovery — CODATA 45 Years On

Summary Report

REPORTS AND PRESENTATIONS


~Curation/Annotation for Database~

41,118 FLcDNA clones
21,037 Loci (possible genes)
5,155 New loci vs RefSeq.

Nature (2002) 419: 3-4
H-InvDB Statistics (release 5.0)

- Total number of transcripts (HTL): 167,156
- Total number of clusters (HIX): 36,037
- Representative alternative splicing variants (RASV): 41,287

<table>
<thead>
<tr>
<th>Functional Annotation Category</th>
<th>Number*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein coding</td>
<td>35,338</td>
</tr>
<tr>
<td>I. Identical to known human protein (experimentally validated)</td>
<td>13,433</td>
</tr>
<tr>
<td>II. Similar to known protein</td>
<td>3,759</td>
</tr>
<tr>
<td>III. InterPro domain-containing protein</td>
<td>2,306</td>
</tr>
<tr>
<td>IV. Conserved hypothetical protein</td>
<td>3,842</td>
</tr>
<tr>
<td>V. Hypothetical protein</td>
<td>5,480</td>
</tr>
<tr>
<td>VI. Hypothetical short protein (20-80 aa)</td>
<td>5,845</td>
</tr>
<tr>
<td>VII. Pseudogene candidate (transcribed)</td>
<td>874</td>
</tr>
<tr>
<td>Non-protein-coding (with strict criteria)</td>
<td>696</td>
</tr>
</tbody>
</table>

*representative transcripts

Current answer: ~ 35,000.
Beyond the 4th Paradigm proposed by Jim Gray

**1st Paradigm:** Data explosion

**2nd Paradigm:** Theory

**3rd Paradigm:** Simulation

**4th Paradigm:** Data-driven Scientific Discovery

**5th Paradigm:** Data-driven Scientific Innovation

~Challenge: Paradigm shift~

~Vision~
Proposal of a view of the new society by innovation~

Genome Information-oriented Society!
Omics on Cell / Tissue Level

- Human consisted of 60 trillion of cells
- The cells biochemically clustered into 250 categories
- Constructs the database with validated annotation

Omics-based “Medicine”

- Medical Revolution in the g-Society
- Spiral cycle in high graded research by the trinity

Personalized Health Care System

Point of Care
Summary

"Genome Information-oriented Society"
~ g-Society ~
Reason and Conditions for Establishment of CODATA

Fedor Kuznetsov
Vice-President, CODATA

Abstract

The establishment of CODATA was an answer to the development of devices and technologies of global significance in the middle of the 20th century. These developments became possible due to the development of science, which is a result of contributions from multiple people from various parts of the world over a long period. A natural way to ensure proper use of the technical achievements was to make the collected knowledge a common property, available to all people and countries. This became a significant goal of society, including many existing and newly formed organizations like the UN, UNESCO, and ICSU. As a part of this global strategy, CODATA was established. The founders of CODATA represented the main countries contributing in the development of the aforementioned technologies. Today we cannot imagine many essential aspects of everyday life without the use of space technology, e.g., communications (TV, Internet, weather forecasts). Hopefully international cooperation in this field will prevent the terrible consequences of misuse of this power. By now CODATA has made significant contributions to cooperation on an international scale in the coordination of scientific and technical progress, and the management of data collected by researchers in many fields of human activity. An example is data activity related to natural gas hydrates. It is clear now that compounds of methane with water (and some other hydrocarbons) exist on the Earth in huge quantities. CODATA, with the help of ICSU, was able to organize a widespread international network connecting specialists on gas hydrates in many countries. CODATA should continue to be active in organizing and spreading knowledge for the benefit of all people on our small Earth. I am sure that with such powerful technologies as the Internet, CODATA will be able to make contributions to improving the lives of people.
Objective necessity
Middle of 20s century is marked by appearance of technical problems of global significance (global scale).
These problems are:

- Atomic energy
- Space exploration
- Green revolution
- Solid state electronics

Availability of proper people
Founders of CODATA

» Frederick Rossini (U.S.A)
» Boris Vodar (France)
» Sir Gordon Sutherland (U.K.)
» Wilhelm Klemm (West Germany)
» Masao Kotani (Japan)
» Mikhail Styrikovich (Soviet Union)

Nuclear energy

Nuclear energy in wrong hands
Data-Intensive Science and Discovery — CODATA 45 Years On

Summary Report

REPORTS AND PRESENTATIONS

---

**Space research and exploration**

- Yuri Gagarin - USSR
- USSR - USA team
- Lu Bo Min - China
- Cupper and Mitchell USA
- S. Savitskaya - V. Tereshkova - USSR, A. Anders - Iran - USA
- Space ladies

**Green revolution**

**World Wide Web**

---

The Opta Project (Barret Lyon)
http://www.opta.org/map1/
NIST Gas Hydrate Research Database and Web Dissemination Channel

K. Krooisteln, C. D. Muzny, A. Kazakov, V. V. Dikty, R. D. Charico,
M. Frenkel, Thermophysical Properties Division, National Institute
of Standards and Technology, Boulder, CO 80303
and
E. D. Stolz, Department of Chemical Engineering,
Colorado School of Mines,

Center for Gas Hydrate Research
Chinese Academy of Sciences

The CODATA Task Group on Natural Gas Hydrates
Beijing 27-28 April 2004

Indo-Russian ILTP workshop on Gas Hydrates, March 2000, N Delhi
Data-Intensive Science and Discovery — CODATA 45 Years On
Summary Report

REPORTS AND PRESENTATIONS

IRCGH

CODATA TASK GROUP
Data on Natural Gas Hydrates

CODATA in World Wide Web
CODATA at 45: Fostering the S&T Data Revolution

John Rumble
Former President, CODATA

Abstract
CODATA, the ICSU Committee on Data for Science and Technology (S&T), is celebrating its 45th anniversary as a prime mover in the S&T data revolution. From its beginnings in 1966, CODATA has addressed the major issues in harnessing the information revolution for S&T. CODATA’s successes have come about because its work has always been (1) relevant, (2) effective, and (3) action oriented. In addition, CODATA has aggressively communicated the knowledge it has generated through workshops, reports, newsletters, position papers, its International Data Conference series, and the CODATA Data Science Journal. From the beginning, CODATA has provided a home for data scientists on a multi-disciplinary basis, facilitating the sharing of new ideas and innovation. Over the years, the data issues addressed by CODATA have been very broad and included providing a neutral, nationality-free mechanism for international cooperation, assessing data quality, establishing internationally-accepted definitive data sets, advancing the use of computers, the Internet, and the Web to build, manage, and disseminate large data collections, addressing data ownership and open access issues, and involving developing countries in data work. Much of CODATA’s success is due to its task groups, which are democratically selected by the membership. The fact that CODATA’s leaders have been drawn from many different countries and have represented many scientific disciplines has also been a key factor. As science has changed and the information revolution progressed, CODATA has changed in step. Its successes, from the fundamental physical constants to the International Conferences, have greatly impacted S&T data positively. As science becomes more multi-disciplinary, as for example in ICSU major new initiatives, CODATA has an opportunity to continue being a major force in S&T data.
Relevance to Science

By the early 1960s, visionary scientists realized an important impact of computers in science would be on scientific and technical data. The potential was to:

- Generate and collect amounts of data far larger than possible before.
- Analyze data more fully.
- Harmonize data and generate better quality property values.
- Use computers and telecommunications to provide easy access to data.

The Data Revolution was about to hit science, and science needed to be ready.

First Steps Towards Relevance

- In 1964, ICSU set up a Working Group that proposed a new international organization under ICSU to improve the management and preservation of scientific data and to facilitate coordination on an international basis.

Prime Movers:
- Frederick Rossini (USA) [1st President]
- Harrison Brown (USA) [Chair of ICSU WG]
- Boris Vodar (France) [One of two 1st VPs]
- Gordon Listerland (UK) [One of two 1st VPs]
- Wilhelm Klein (Germany) [Secretary-Treasurer]
- Masao Katoji (Japan) [Bureau member]
- Mikhail Shlykov (USSR) [Bureau member]
Data-Intensive Science and Discovery — CODATA 45 Years On
Summary Report

More Steps Towards Relevance

- In 1986, CODATA became the first International organization to provide a forum for data scientists, regardless of discipline, to gather together, learn from each other, and make progress on scientific data management.

CODATA now provided the needed home for data scientists!

- Throughout its 45-year history, CODATA has continued to be at the forefront of S&T data and has addressed the major challenges as data capabilities have progressed.

Data Revolution Challenges for CODATA

- Providing a neutral, nationality-free mechanism for international data cooperation
- Assessing data quality
- Establishing definitive data sets
- Using computers to build and manage large-scale data collections
- Improving data access
- Taking advantage of the Internet and the Web
- Balancing public and commercial interests in S&T data
- Involving developing countries in S&T data activities
- Developing science-friendly data access, ownership, preservation, and sharing policies

Addressing these challenges successfully has ensured CODATA is and remains relevant to science.

CODATA in Action

CODATA has been active from its beginning

- Task Groups and Working Groups were started immediately
- CODATA activities are proposed by and voted on by its members, unlike many international unions
- When a group has finished its work, the group ends, and the resources are applied to a new activity
- Many times CODATA provided a catalytic effect, offering a neutral home for an initial organizing activity, leveraging limited funds
  - The International Virtual Observatory
  - Sequence data
  - Atmospheric kinetics
  - Biodiversity
  - Material properties
CODATA Task Groups in Action

1980-1982
Chemistry, Physics, and Chemistry-related
- Chemical Kinetics
- Data for the Chemical Industry
- Chemical and Bio-thermodynamics
- High Temperature Phases
- NMR and Photoelectron Spectroscopy
- Thermophysical Properties of Solids
- Fundamental Constants

The Beginnings of Computerized Data Handling
- Accessibility and Dissemination of Data
- Computerized Data Handling

Expanding The Scope
- Space and Time Dependent Data
- Bioscience Committees
- Industry Advisory Committees
- Geosciences Advisory Committees

30 Oct 2011 CODATA at 45

1987
The Scope expanded to new areas
- Materials Database Standards
- Computer Graphics
- Directory of Far-East Data Sources
- Teaching and Training
- Biological Databases Working Group
- CODATA Workshops on Materials Databases (Schuchardt, Germany)
- Nucleic Acid and Protein Sequencing Data (Gaithersburg, MD)

CODATA in Action

CODATA has provided numerous services and specifically addressed major issues
- Data referrals services before the Web/Google came along
- Training courses, especially in developing countries
- Data reporting guides, especially before more formal data exchange standards were developed
- Data preservation and data access policy guidelines
- S&T data practice books
- Database directories on a discipline basis

CODATA has been able to respond to new data news quickly, getting project going, and have them taken over by more specialized and better funded organizations (Unions, etc.)

30 Oct 2011 CODATA at 45
Knowledge and Communication Meetings

- From the first, CODATA emphasized developing, sharing, and communicating new knowledge generated from S&T data.
- The CODATA International Conference series was started in 1968 and continues today to bring scientists and others interested in S&T data together on a multi-disciplinary basis.
- No other international data conferences have spread as much knowledge about S&T data as CODATA.

Other Meetings

- CODATA has also sponsored hundreds of workshops, symposia, and expert meetings to address specific topics and issues.
**Knowledge and Communication**

From the beginning CODATA realized the power of bringing small groups of experts to collect and evaluate data in specific areas.

- The two earliest examples are:
  - CODATA Adjustment of the Fundamental Constants
  - CODATA Key Values of Thermodynamics

30 Oct 2011  CODATA@45

**Task Groups (cont.)**

Task groups set the stage for international, non-commercial cooperation on S&T data collections and data evaluations, a concept that has been adopted by virtually every discipline.

30 Oct 2011  CODATA@45

**The Impact of CODATA Knowledge Sharing**

- As computer science developed new tools and technology that scientists could use, CODATA conferences and meetings often provided the first exposure to:
  - Database management systems
  - Data quality assessment techniques and software
  - Data visualization
  - Personal computers in data work
  - Data networking and data integration standards
  - Data mining
  - Discovery in databases

30 Oct 2011  CODATA@45
The Impact of CODATA Knowledge Sharing

- These fora also addressed many data policy and practice issues, including:
  - Data center operations
  - Long term data preservation
  - Data exchange standards development
  - Economics of data operations
  - Open access policies
  - Copyright

- For many of these topics, CODATA had Task Groups as well as meetings.

30 Oct 2011  CODATA at 45

Knowledge Sharing with Scientists in Many Disciplines

- Publication Series
- Directories
- Guides to data recording
- Workshops proceedings
- International Conference proceedings
- Training course materials
- Data compilations
- CODATA Referral database developed by F. Jack Smith (Google before Google!) with 2002
- CODATA Data Science Journal (all electronic)

30 Oct 2011  CODATA at 45

Source Books

CODATA Data Science Journal

- Started in 2002
- Online at no charge
- Papers from scientists around the world

30 Oct 2011  CODATA at 45
Addressing Current Data Issues

Recent CODATA Activities
- Strong support of the International Polar Year
  - Establishing of IPY data management group
  - Developing concept of Polar Information Commons
- Global Earth Observations
  - Expert meetings at CODATA 2005
  - While Paper of Geo data sharing
  - Draft implementation guidelines for GEOSS data sharing principles
  - Many other supporting activities

CODATA Effectiveness
Key parameters that help define the effectiveness of CODATA
- Truly international: Before 1990, CODATA was a major mechanism for East-West cooperation on S&T data
- All scientists can and do participate, regardless of whether their country or union is a member
- CODATA priorities change as priorities in science change
- Democratic approval of Task Groups
- Strategic planning: Major planning meetings in 1983 and 1999
- Officers from many disciplines: Brings a diversity of viewpoints
- Small and skilled secretariat: Proactive and involved
- Good working relations with ICSU and other international groups
- Leveraging limited funding by partnerships
CODATA Effectiveness

We will hear several success stories next, but there are many successes for CODATA
- CODATA Chemical Kinetics Task Group (1970s) and ozone depletion, leading to the data efforts of the Global Change community
- Standardization of thermodynamics data (1970s), eventually leading to ThermoUTL (2000s)
- Highlighting bio-data and bringing together sequencing efforts (1980s)
- Alerting the community to data quality issues in the Web era (1990s)
- Publicizing and challenging data access restrictions (2000s)
- Providing focus on the importance of data in the WSIS Summit and the UNESCO Symposium on Data Access (2003)
- Support for major multi-disciplinary data programs: IPY and GEO

30 Oct 2011 CODATA at 45

CODATA Effectiveness

Ultimately CODATA’s effectiveness depends on the people involved
- A challenge for the future is how to attract highly skilled S&T data experts when every discipline has adopted data science as a key foundational tool for success
- Real world problems are multi-disciplinary and require data from many disciplines
- CODATA should aggressively work to get involved in ICSU’s new major multi-disciplinary programs, such as
  - Urban Health Initiative
  - Earth System Sustainability Initiative (ESSI)
  - Intergovernmental Platform on Biodiversity and Ecosystem Services (IPBES)

30 Oct 2011 CODATA at 45

CODATA at 45

The key question is “Has CODATA made a difference?”
CODATA has been relevant, effective, and active in developing and sharing the new knowledge that underpins the Data Revolution in science and in communicating that knowledge on a multi-disciplinary and international basis.
CODATA is still relevant and effective and should continue to play a major role in its next 45 years!

30 Oct 2011 CODATA at 45
CODATA Task Group on Fundamental Constants

ZHANG Zhonghua
Member, Task Group on Fundamental Constants

Abstract

The Task Group on Fundamental Constants (TGFC) was formed in 1967, almost at the same time as CODATA. Therefore, it is one of the earliest CODATA task groups. At present, TGFC has 14 members from 11 national laboratories. The main work of TGFC is the adjustment of published fundamental constant values using the least squares method to give an optimum value set of fundamental constants. Such adjustments were performed in 1965, 1973, 1998, 2002 and 2006. The latest one was in 2010. For this tune of adjustment, the Data cut-off date was Dec. 31, 2010. 150 input or observational data, 135 distinct types of observational equations, 79 adjusted constants or unknowns were used for this adjustment. The adjustment result was published on the web in Jan. 2011 and will be in the Review of Modern Physics late in 2011. These values will play an important role in the activity of redefinition of SI basic units by fundamental physical constants. As pointed on the CIPM Recommendation 1 (CI-2005), “Preparative steps towards new definitions of the kilogram, the ampere, the kelvin and the mole in terms of fundamental constants” will occur in the near future.
The Adjustment is a Generalized Multivariable Analysis

- Measured quantities (input data) expressed as theoretical functions (observational equations) through a particular independent subset of the constants and variables (adjusted constants).
- 150 input or observational data, 135 distinct types or observational equations, 79 adjusted constants or unknowns.
- Variance weighted, generalized, multivariate least squares adjustment with accounting of covariance.
- Equations from classical physics, quantum mechanics, relativity, QED, QCD, ...

Least Squares and Fundamental Constants

- Method of Least Squares – an old tradition
  - The First Least-Squares Adjustments of the Constants
Past CODATA Least Squares Adjustments

The primary outputs of the Task Group are the publications in:

- 2006
- 2002
- 1998
- 1986
- 1973

2010 Least-Squares Adjustment (LSA)

2010 LSA Schedule

- Data cut-off date: Dec. 31, 2010
- On the web: Jan. 2011
- Review of Modern Physics: late in 2011

>150 input or observational data, 135 distinct types of observational equations, 79 adjusted constants or unknowns.

Variance weighted, generalized, multivariate least squares adjustment with accounting of covariance.

Impact

Values used and referenced world-wide
- LSA schedule used for project management planning
- Numerical basis for the proposed changes to the SI

CIPM Recommendation 1 (CI-2005)

Preparative steps towards new definitions of the kilogram, the ampere, the kelvin and the mole in terms of fundamental constants
(adopted by CIPM at 94th meeting, 4-7 October 2005).

One of several Responses

CODATA and Its Role in Supporting Implementation of the GEO Data Sharing Principles

Robert Chen
Secretary General, CODATA

Abstract

The Group on Earth Observations (GEO) is a voluntary partnership of governments and international organizations which is coordinating efforts to build a Global Earth Observation System of Systems (GEOSS). Since 2006, CODATA, as an interdisciplinary committee of ICSU, has played a lead role in supporting the implementation of the GEO Data Sharing Principles, which call for the “full and open exchange of data, metadata, and products shared within GEOSS, recognizing relevant international instruments and national policies”. This has included preparation of a data sharing implementation white paper, drafting of recommended data sharing guidelines, and organization of a number of supporting activities at the GEO plenary meetings and summits and at other international conferences. In 2009, CODATA became one of the co-chairs of the GEO Data Sharing Task Force and contributed to the development of a Data Sharing Action Plan, which was accepted at the GEO-VII Plenary and Ministerial Summit in Beijing. More recently, CODATA has been an active contributor to two Data Sharing Task Force working groups, one on legal interoperability and a second on legal liability. CODATA is participating in the November 2011 GEO plenary in Istanbul and plans to continue its efforts in support of data sharing implementation in the 2012-15 GEO Work Plan.
CODATA and its Role in Supporting Implementation of the GEO Data Sharing Principles

Dr. Robert S. Chen
CODATA Secretary General
Manager, NASA Socioeconomic Data and Applications Center
Director, CIESIN, The Earth Institute, Columbia University

October 2011

Group on Earth Observations (GEO)

GEOSS Data Sharing Principles

- Full and open exchange of data, metadata, and products shared within GEOSS, recognizing relevant international instruments and national policies
- Shared data, metadata & products available at minimum time delay and minimum cost
- Shared data, metadata & products free of charge, or no more than cost of reproduction, are encouraged for research & education

CODATA and the GEO Data Sharing Task

- GEO Task DA-06-01 led by CODATA
  - Data Sharing Implementation White Paper
    - Review of existing guidelines and data sharing barriers
    - Peer-review version published in fall 2009 by the Journal of Space Law and CODATA Data Science Journal
  - Data Sharing Guidelines
    - Drafted recommendations for data sharing implementation in GEOSS
    - Extensively circulated and revised based on community comments
  - Supporting activities
    - Side events at Ministerial Summit in Cape Town in 2007 and GEO Plenaries in Romania and Washington DC in 2008 and 2009
    - Special session on data sharing at 11th Global Spatial Data Infrastructure Conference in Rotterdam in June 2009
GEO Data Sharing Task Force - 1

- GEO Data Sharing Task Force formed in early 2009 based on open invitation to GEO members and participating organizations
  - More than 60 telecons to date!
  - Six co-chairs:
    - China – Geofu Wang
    - India – V. S. Hegde
    - European Commission – Alan Edwards
    - Japan – Chiyoshi Kawamoto
    - USA – Helen Wood (previously Linda Moodie)
    - DA-06-01 Task Team – Paul Uhlir/Robert Chen/Joanne Gabrynowicz
  - Other members: Brazil, Cameroon, Canada, Czech Republic, France, Italy, The Netherlands, Slovenia, UK, COSPAR, ISDI, ICIMOD, OGC, UNEP, UNOOSA...

2010 GEO Ministerial Declaration

Committed GEO Members to:

(i) maximize the number of documented datasets made available on the basis of full & open access;
(ii) create the GEOSS Data Collection of Open Resources for Everyone (GEOSS Data-CORE), a distributed pool of documented datasets with full, open and unrestricted access at no more than the cost of reproduction and distribution; and
(iii) develop flexible national and international policy frameworks to ensure that a more open data environment is implemented, thus putting into practice actions for the implementation of the GEOSS Data Sharing Principles.

GEO Data Sharing Task Force - 2

- Action Plan accepted at GEO-VII plenary and 2010 Ministerial Summit in Beijing
  - Established new “GEOSS Data-CORE”
  - Plan seen as a major accomplishment of Ministerial Summit
- CODATA participated in GEO-VII Plenary and 2010 Ministerial Summit
  - Side event organized by CODATA on 2 November (>50 participants)
  - CODATA part of ICSU delegation and booth

GEO Data Sharing Task Force - 3

- DSTF extended for one year to follow up (ending next month)
- Two working groups formed:
  - Legal Interoperability: addressing use of open access licenses and related intellectual property issues (led by Paul Uhlir)
  - Legal liability: addressing need for appropriate disclaimers, terms of use, and protections for contributors to and users of GEOSS (led by Harlan Onsrud)
  - Reports to be presented at the GEO-VIII Plenary in Istanbul in November
- Extensive coordination with ADC, Sprint to Plenary on Data-CORE implementation
- K. Cass to lead ICSU Delegation at the November GEO plenary in Istanbul
CODATA has been involved in numerous international initiatives during its 45 year history, and this commitment has continued in recent years. This presentation describes some of CODATA's key involvements, particularly those from 2006 to 2010, when Professor Krishan Lal was President of CODATA. At the outset of this period, CODATA helped finalize GICSI, the Global Information Commons for Science Initiative, which had been in preparation since the World Summit on the Information Society in Geneva in 2003. CODATA also helped move forward COMMUNIA to tackle issues related to the public domain in the digital environment. Representing ICSU on the international stage, CODATA took the lead in formulating data sharing guidelines for GEOSS. Similarly, for the International Polar Year, CODATA started the Task Group on Polar Year Data Policy and Management. To address the digital divide, CODATA approved task groups on data sources for sustainable development in Southern African Development Community countries, and on preservation of and access to science and technology data in developing countries. Furthermore, CODATA held its 2010 conference in South Africa and the 2008 conference in Ukraine in an effort to involve places that have traditionally lacked adequate access to scientific and technological resources. Another important area in which CODATA has progressed is in encouraging young scientists and providing a platform for them to interact with international experts and share their research. The past two CODATA conferences have featured special sessions for young scientists and CODATA now has a Young Scientists Working Group.
Association with CODATA: A Rewarding Experience

Krishan Lal
Past President, CODATA

Outline: CODATA activities 2006-2010

- Taking forward the major initiatives: GICSI; COMMUNIA
- Representing ICSU at Global level
- IPY data
- Special efforts on tackling digital divide
  
  * Kiev 2008
  * Cape Town 2010
- Initiating a Young Scientists Programme
- Initiating a discussion on an International Data Science Academy

A Significant Achievement

Helped in firming up the new initiative on Global Information Commons for Science after discussions from legal, economic and other perspectives
COMMUNIA

The European Thematic Network on the Public Domain in the Digital Age

- European point of reference for both theoretical and practical analysis as well as strategic policy discussion of existing and emerging issues related to the public domain in the digital environment - as well as related topics including, but not limited, alternative forms of licensing for creative material; open access to scientific publications and research results; management of works whose authors are unknown (i.e. orphan works)

COMMUNIA Contd...

Coverage
- Geographical territory of the European Union as well as neighbouring and accessing countries
- Plans to build strategic relationships with third countries such as the United States and Brazil

COMMUNIA Contd...

Activities and Deliverables during three years
- Thematic workshops (3/yr) and conferences (1/yr) to maintain a strong link between all the participants and use face-to-face meetings as a source of material for the analytical and practical work of the project
- Production of a book
- Establishment of an academic journal
- A "best practices" guide for European research and reference centers on the topics covered by COMMUNIA
- A final strategic report containing policy guidelines that will help all the stakeholders tackling the issues that the existence of a digital public domain have raised and will undoubtedly continue to raise.

COMMUNIA Contd...

Stakeholders
- EU Member States
- Three candidate countries and Switzerland
- USA and Brazil
- CODATA
- Creative Commons
- EBLIDA
CODATA and *Task DA-06-01* of GEOSS

**Lead Author:** Paul F. Uhlir  
**Core Authors:** Robert Chen, Charles Barton, Joanne Irene Gabrynowicz, Kathleen Janssen

*A white paper on guidelines for data sharing for GEOSS*

---

**Timely Important Activity**  
*Offshoot of ICSU General Assembly, Mozambique*

Task Group on  
Polar Year Data Policy and Management

---

**Final Steps**

**2008:** *Bucharest Plenary (in place of Beijing)*

**2009:** *Presentation to GEO Plenary in Washington*

**2010:** *Ministériel Meeting*

---

**Tackling the Digital Divide**

- *Task Groups addressing Problems of the developing world*  
  *Data Sources for Sustainable Development in SADC Countries*
  
  *Preservation of and Access to S&T Data in Developing Countries*
Data-Intensive Science and Discovery — CODATA 45 Years On
Summary Report

REPORTS AND PRESENTATIONS

Tackling the Digital Divide

• Deliberate efforts to organize CODATA International Conferences in the Developing World

CODATA 2008 at Kyiv
CODATA 2010 at Cape Town

Initiated Discussion on International Data Science Academy

GA of CODATA 2010 had along discussion on the International Data science Academy, a part of CODATA Strategic Plan

Encouraging Young Scientists

• Young Scientists Sessions at CODATA 2008
• Working Group of Young Scientists
• Participation of 50 young African scientists in CODATA 2010, Cape Town, courtesy of NRF South Africa

Summary

• GICS
• EU Project: COMMUNIA
• Participation in GEOSS Task Group
• Tackling the Digital Divide
• Young Scientists Programme
Abstract

The China-U.S. Roundtable on Scientific Data initiative began its first phase in 2000 with China-U.S. cooperation on scientific data policy. From 2000-2005, there were meetings in Beijing and Washington resulting in major support of the China Scientific Data Sharing Program, the Chinese Ministry of Science and Technology’s Publication of Strategies for Open Access and Archiving of Scientific Data in China (National Academies Press, 2006), and the formation of the China-U.S. Roundtable on Scientific Data. Bilateral activities focus on providing a unique forum for government, academic, and private-sector stakeholders in the U.S. and China to discuss and address scientific data practices and policies, and catalysing and coordinating cooperation on scientific data practices and policies at the Academy and national level in each country, with appropriate recognition and representation of other thematically related bilateral and international activities. The second phase involves implementation of the phase-one cooperation agreement, covering four main areas: health and biomedical data, environmental and geospatial data, cyber-infrastructure data applications, and scientific data policy. Five roundtable meetings have been held, with a number of successes, including an increase in exchange of real-time seismological data, identification of data sets available on full and open basis through GEOSS, providing greater access to disaster data on a real-time basis, the establishment of the International Biomedical Data Sharing Platform in China, and support of the “Healthy China 2020” Program. Phase three of the initiative will see a China-U.S. Scientific Data Cooperation Project Series with proposals for bilateral China-U.S. agreement on biomedical data cooperation, two new CODATA Task Groups, collaboration on data attribution and citation practices, and joint meetings on improved management of data-intensive eScience.
China – U.S. Roundtable on Scientific Data: Summary of Success Stories

Paul F. Uhlir
U.S. CODATA, National Academy of Sciences
And
Li Jianhua
Chinese CODATA, Chinese Academy of Sciences

Success Stories of China – U.S. Scientific Data Cooperation

Phase 1: China-U.S. cooperation on scientific data policy, 2000-2005

Meetings in:
2000 (2) in Washington and Beijing
2002 in Washington
2003 in Beijing
2004 in Beijing
2005 in Beijing

Results:
- Major support of China Scientific Data Sharing Program, MOST
- Publication of Strategies for Open Access and Archiving of Scientific Data in China (National Academies Press, 2006)
- Formation of China – U.S. Roundtable on Scientific Data Cooperation

Success Stories of China – U.S. Scientific Data Cooperation

Phase 2: Implementation of China – U.S. Roundtable on Scientific Data Cooperation

Four subject areas: health and biomedical data; environmental and geospatial data; cyber-infrastructure data applications, and scientific data policy. Bilateral activities focused on the following tasks:

- Provide a unique bilateral forum for government, academic, and private-sector stakeholders in the United States and China to discuss and address scientific data practices and policies, pursuant to a mutually agreed agenda.
- Serve as a catalyst and coordinating body for bilateral cooperation on scientific data practices and policies at the Academy and national level in each country, with appropriate recognition and representation of other thematically related bilateral and international activities.
Success Stories of China – U.S. Scientific Data Cooperation

Successes of the bilateral Roundtable:

- Increase in exchange of real-time seismological data
- Identification of data sets available on full and open basis through the Global Earth Observation System of Systems (GEOSS)
- Providing greater access to disaster data on a real-time basis
- Establishment of the International Biomedical Data Sharing Platform in China
- Support of the “Healthy China 2020” Program

Success Stories of China – U.S. Scientific Data Cooperation

Successes of the bilateral Roundtable (continued):

- Memorandum of Understanding between the CAS Computer Network and Information Center (CNIC) and the U.S. National Center for Atmospheric Research (NCAR)
- Memorandum of Understanding between CNIC and the United States Geological Survey (USGS)
CODATA Strategic Plan: 2013-18 and Beyond

Robert Chen
Secretary General, CODATA

Abstract
CODATA’s mission is to “strengthen international science for the benefit of society by promoting improved scientific and technical data management and use”. Both CODATA and ICSU have recognized the growing importance of scientific data in international research and also in sustainable development and other societal applications. CODATA’s Strategic Plan for 2006-12 recognized the need for CODATA to take a leadership role in addressing new data sharing approaches, overcoming the digital divide between developed and developing countries, and applying new data methods and information technologies in research and education. Looking ahead to 2018, CODATA needs to build on its strengths to address critical data issues, especially those identified by ICSU in its 2012-17 strategic plan and by other members of the ICSU family concerned with data, e.g., the ICSU World Data System (WDS) and the new program on Integrated Research on Disaster Risk (IRDR). It will also be important for CODATA to build on the grassroots interests of CODATA’s community, e.g., in tackling issues such as data at risk of loss and data citation standards and practices. CODATA has the opportunity to continue playing important roles on key issues such as data access and policy and standards, to address emerging areas of science such as nanotechnology, and to contribute to major international events such as the Rio+20 conference, the Eye on Earth Summit, and the ICSU Planet Under Pressure conference. In thinking strategically about these initiatives and partnerships, CODATA needs to recognize challenges and opportunities posed by evolving information technologies, population growth, sustainable development, environmental variability and change, and disaster risks. CODATA very much welcomes input from the community on its new Strategic Plan for 2013-18 and how it can best help international science provide real benefits for society.
**Brief Timeline of CODATA’s Strategic Plan**

2005-6  
ICSU recommends development of Plan in its Priority Area  
Assessment on Data and Information and the ICSU Strategic Plan

2006  
25th CODATA GA in Beijing reviews draft Plan and approves new  
misson statement and key initiatives and actions

2007  
CODATA completes full version of plan and sends for review;  
ICSU GA in Maputo endorses Plan

2008  
26th CODATA GA in Kiev endorses Plan

2006-10  
Key initiatives are launched

2009  
ICSU establishes Ad hoc Strategic Coordination Committee on  
Information and Data (SCCID) for 3 years

2010  
27th CODATA GA in Cape Town calls for review of current  
Plan and development of new Plan for 2013 and beyond

2011  
CODATA Executive Committee appoints Strategic Plan Committee;  
ICSU approves its Strategic Plan for 2013-17 including CODATA review

---

**Cross-Cutting Initiatives, 2006-12**

1. Global Information Commons for Science Initiative (GICSI)
2. Scientific Data across the Digital Divide Program (SD3)
3. Advanced Data Methods and Information technologies for Research and Education (ADMIRE)

---

**ICSU Strategic Plan, 2012-17**

- ICSU Foresight Analysis, scenarios for the development of international science over the next 20 years
- Earth System Science Initiative
- Belmont Challenge on delivering knowledge to support human action and adaptation to regional environmental change
- Strategic review on science education
- Review of ICSU Regional Offices
- New ICSU programs, e.g., hazards, urban health & wellbeing
- Review of CODATA to be undertaken by ICSU
**Session organized by CODATA: Integrated Disaster Data in Support of Integrated Disaster Research**
- Importance of High Quality Data for Calibration and Verification of Hazard and Risk Models (Nadim)
- OpenISDM: An Open Framework for Disaster Information Systems (Ho)
- GEO Grid Disaster Response Task Force Activity for the 2011 Tohoku, Japan Earthquake (Iwao)
- Progress and Challenges in Integrating Disaster Data and Information (Chen)

**Session organized by ICSU, CODATA, WDS-SC: Data Challenges for Global Sustainability**
- Data intensive science and our changing planet (Guo)
- Multidisciplinary, Integrated Data on Sustainability (Chen)
- Long time series of data for sustainability (Minster)
- Open science and open data: funders’ role in ensuring open access to research data (Thorley)
- Linking research data to articles: the role of publishers in supporting open and transparent research (Wilson)
- The need of disaster loss data for global sustainability (Wirtz)
New CODATA Partnerships: Hand in Hand

- WDS-SC
- IRDR
- ISDE
- IGU and GSC

Other Strategic Opportunities

- RIO+20
- Data Access and Policy (CFRS, SCCID recommendation)
- Data Standards
- Nanotechnology
- Eye on Earth Summit
- Planet Under Pressure

Population and Development Challenges

Borgman, 2008, NSF Cyberlearning Report
IUGG, the International Union of Geodesy and Geophysics is one of ICSU’s 31 international scientific unions. IUGG consists of eight scientific associations representing a wide array of Earth sciences including geomagnetism, oceanography, hydrology, seismology, volcanology and physics and chemistry of the solid Earth. Their sphere of study spans from the deep interior of the Earth to the Sun. Several other geo-unions are members of ICSU as well, and for them, the main influence of ICSU is its coordinating role in multidisciplinary science, carried out largely by ICSU’s interdisciplinary bodies, including CODATA. From the point of view of IUGG, CODATA’s role in ICSU is to promote the management and use of scientific and technical data in contribution to ICSU’s goal of strengthening international science for the benefit of society.

Regarding data management in ICSU, CODATA is not alone. For many decades, there were two other bodies in ICSU dealing with data. The World Data Centers and the Federation of Astronomical and Geophysical Data Services—the latter of which was established by IUGG, the International Astronomical Union, and Union Radio-Scientifique Internationale—were established before CODATA, but today have been transformed into a new body. The World Data System has been established through the combination of those two interdisciplinary bodies to continue and enhance their data provision services. WDS is geared more toward coordinating data sources, something in which CODATA is not directly involved, thus cooperation between CODATA and WDS is essential. From IUGG’s perspective, these two bodies should work towards building a sufficient level of coordination among the data initiatives within ICSU, without the need for a third-party committee such as SCCID.

For data policy, CODATA could make a significant contribution by focusing on issues related to scientific analysis of the data collected by the established data centers and services.
This will mean close cooperation with the unions, which develop and manage these data sources, and CODATA, which defines the proper science of data. IUGG, for example, oversees the Global Geodetic Observing System and other large, international systems that collect a huge amount of data. The work between CODATA and these unions should involve two-way interaction.

The CODATA Strategic Plan 2006-2012 contained three major initiatives—the Global Information Commons for Science Initiative (GICSI), Scientific Data Across the Digital Divide (SD3), and Advanced Data Methods and Information technologies for Research and Education (ADMIRE). Focusing on the first two, IUGG supported the Polar Information Commons, put forth resolutions addressing the digital divide at the 2008 ICSU General Assembly in Maputo, and helped organize the Electronic Geophysical Year.

And finally, as CODATA now looks toward formulating its next strategic plan, the recommendation of IUGG is that CODATA should strengthen coordinating activities in data analysis and make the promotion of the science of data their highest priority.

Abstract

A new international body, the World Data System (WDS), was established at the ICSU General Assembly in 2008. WDS was formed incorporating the legacy of WDC (World Data Center) and FAGS (Federation of Astronomical and Geophysical data-analysis Services). It aims at long-term stewardship of scientific data, qualification of data/databases, as well as promotion of larger capacity of multidisciplinary databases/centers/services with up-to-date information and communications technology. As recommended in the 2008 SCID report and the 2011 SCCID report, it is expected that ICSU’s international bodies will work in cooperation between WDS, CODATA, ICSTI, INASP, and other related stakeholders.

For example, WDS would be seeking possible collaboration and closer relations to various activities of those bodies from now on, especially after the recruitment of the official Executive Director of its International Programme Office (IPO). For example, the Polar Information Commons (PIC), which is a shared virtual commons of datasets made through the International Polar Year (IPY) 2007-2008 activities, is an example of successful international sharing of multidisciplinary datasets under CODATA’s initiative, and fits the WDS’s concept of a multidisciplinary, interoperable data system.

WDS-SC (Scientific Committee) is now proceeding with WDS membership applications; so far 30 WDS members’ applications have been approved, while about 130 facilities are showing interest to join WDS. The first International ICSU-WDS Conference was held 4-6 September 2011 in Kyoto. The WDS program implementation, coordination and management will be conducted by the International Programme Office (IPO). The host institution of WDS-IPO was decided by the ICSU Executive Board to be Japan’s National Institute of Information and Communications Technology (NICT; former Radio Research Laboratories), November 2010, through ICSU’s international call and review.
ICSU World Data System

- Yasuhiro Murayama
  (WDS-IPO/
  National Inst. of Info. & Communications Tech.
  Japan)

ICSU-SCID vision

The International Council for Science envisions a Global World Data System, in order to:

- emphasize the critical importance of data in global science activities
- further ICSU strategic scientific outcomes by addressing pressing societal needs (e.g. sustainable development, digital divide)
- highlight the very positive impact of universal and equitable access to data and information
- support services for D&I long-term stewardship
- promote and support data publication and citation
Where we are... and where we are going

[ICSU, 2010]

The ICSU WDS Data Policy

WDS Data Policy Final Statement
The International Council for Science World Data System (ICSU WDS), recognizing the benefits and importance of contributing to the growing international efforts of data sharing, has adopted the same principles from GEO/CODATA data sharing principles as follow:

- There will be bulk and open exchange of data, metadata and products shared within WDS, recognizing relevant international instruments and national policies and regulations.
- All shared data, metadata and products will be made available with minimum time delay and at minimum cost.
- All shared data, metadata and products being free of charge or no more than cost of reproduction will be emphasized for research and education.

ICSU World Data System and IPO

ICSU (the International Council for Science) 1931
- National members: national scientific academy or council
- International academic bodies/programs

ICSU Bodies related to data information

WDS (2009~)
- SC Science Committee
- IPO Implementation
- CODATA Committee of Scientific Data (CODATA)

http://wds-kyoto-2011.org/
IPY-IPO

International Council for Science and World Meteorological Organization:

- PROCLAIM: International Polar Year 2007-2008
- CONVENE: Joint Science Committee
- ESTABLISH: International Programme Office
- INVITE: International Proposals for IPY Projects

[D. Carlson, 2011]

$1.2B science funds
63 countries
50,000 participants

IPY-IPO: Polar Info. Commons

D. Carlson, 2011
IV. HAND IN HAND PROGRAM

CODATA is currently expanding its international cooperation through the Hand in Hand Program. This means reaching out to national scientific institutions, ICSU interdisciplinary bodies, and international scientific organizations dealing with data. Because CODATA is concerned with data and information across all scientific disciplines, its role is necessarily that of a coordinator, requiring CODATA to work closely with a broad range of organizations. Doing so allows CODATA to ensure that data issues, such as sharing, quality control, and standardization, are given due attention in research projects and scientific agendas. More than that, these matters can be ingrained in the research process at an early stage, rather than as an afterthought. CODATA seeks involvement in its partners’ research initiatives, conferences, working groups, or similar activities to provide guidance on how their data and information can best be managed. CODATA also incorporates its partners into CODATA activities as a way of connecting them to other organizations with relevant research and data to form new, mutually beneficial alliances.

During the CODATA 45 years on meeting, three initiatives were launched under the Hand in Hand Program. The first was the signing of a memorandum of understanding between CODATA and the Integrated Research on Disaster Risk program. The science behind understanding, predicting and preventing disasters relies on large amounts of data in numerous forms and requires a well-organized research plan, including a strategy for managing data throughout its lifecycle—from collection, through analysis, and into publication and dissemination. Second was the signing of another memorandum of understanding, this one formally uniting CODATA with the International Society for Digital Earth. The society has an ambitious goal of establishing a platform for storing, analyzing, and visualizing the world’s geographic information on an unprecedented scale, including numerous data types and formats. The concept, known as Digital Earth, will obviously be a data-intensive task requiring significant expertise in data management. Lastly, CODATA jointly launched the LIN CHAO Digital Geomuseum with the International Geographical Union and the Geographical Society of China. The geomuseum allows the digital transfer of cultural knowledge and is named for the 20th-century Chinese geographer LIN Chao, who helped advance geographic thought during a tumultuous time in China, and also a revolutionary time in the history of the field of geography.

These and future Hand in Hand partnerships will be crucial to CODATA’s ability to serve the International Council for Science and achieve CODATA’s mission of strengthening international science for the benefit of society by promoting improved scientific and technical data management and use.
V. HIGH-LEVEL MEETINGS

1. CAS President BAI Chunli meets with ICSU President Yuan Tseh Lee

On 30 October, 2011, Professor BAI Chunli, President of the Chinese Academy of Sciences (CAS), met with Professor Yuan Tseh Lee, President of the International Council for Science (ICSU), at the Center for Earth Observation and Digital Earth (CEODE), CAS.

First, BAI congratulated Lee on being elected President of ICSU. He said that the IRDR International Programme Office (IPO) was the first of its kind in Asia, and that CAS would support scientists from all over the country to join in the academic activities of ICSU. The two presidents subsequently communicated with each other concerning the greenhouse effect and energy consumption. Lee pointed out that the increasing greenhouse gas emissions and temperature might bring flood, drought and other extreme weather events. Therefore, it was important to control the emission of carbon dioxide and green gas to avert the appearances of disasters. BAI affirmed Lee’s views and stressed the importance of scientific surveys on a wide range of research. Afterwards, Lee emphasized the importance of Earth observation, to which BAI agreed and introduced the efforts China has made recently in reducing energy consumption.

2. CAS President BAI Chunli meets with Indian National Science Academy President Krishan Lal

On 30 October, 2011, Professor BAI Chunli, President of the Chinese Academy of Sciences (CAS), met with Professor Krishan Lal, President of the Indian National Science Academy, at the Center for Earth Observation and Digital Earth (CEODE), CAS.

First, BAI warmly welcomed Lal to the CODATA 45th anniversary celebration and the IRDR conference in Beijing. Lal, who is also Past-President of CODATA, exchanged views with BAI and CEODE Director-General Professor GUO Huadong on international cooperation in scientific research, aiming to strengthen cooperation in the future. GUO also introduced the development and current status of CEODE and CODATA. Lal gave a gift to BAI while BAI presented books compiled by CEODE in return.
3. CAS President BAI Chunli meets with IRDR Delegation

On 30 October, 2011, Professor BAI Chunli, President of the Chinese Academy of Sciences (CAS), and a delegation from the Integrated Research on Disaster Risk (IRDR) program met at the Center for Earth Observation and Digital Earth (CEODE), CAS. IRDR is a decade-long integrated research program co-sponsored by the International Council for Science (ICSU), the International Social Science Council (ISSC) and the United Nations International Strategy for Disaster Reduction (UN-ISDR).

The meeting began with welcome remarks by BAI. The IRDR delegation expressed thanks for being invited to the CODATA 45th anniversary celebration, also being held at CEODE that day. BAI said that the Chinese Academy Sciences and CEODE had a close relationship with IRDR, which was hosted by CEODE. BAI expressed that CAS would not only invite more related speakers and scientists for the IRDR conference, but also welcome IRDR to arrange more activities in China. They expressed that the preparation of the IRDR conference was run well and IRDR will strengthen cooperation with CAS and CEODE in the future.

4. CAS President BAI Chunli meets with CODATA Delegation

On 30 October, 2011, Professor BAI Chunli, President of the Chinese Academy of Sciences (CAS), met at the Center for Earth Observation and Digital Earth (CEODE), CAS, with a delegation from the Committee on Data for Science and Technology (CODATA). CODATA was founded in 1966 as an interdisciplinary body of the International Council for Science (ICSU). It is responsible for coordinating data policies and standards across research fields in support of scientific activities.

The members of CODATA said that CODATA is like a scientific bank that facilitates data sharing in various scientific fields. They congratulated the delegates beforehand on the CODATA 45th anniversary celebration that day, wishing them a successful, productive meeting. The delegation thanked the CODATA National Committee of China for hosting the international scientific meeting. BAI said that CAS would pay great attention to the CODATA conference next year and other CODATA activities. The members coming from different countries welcomed BAI to visit their countries. They also said that they were impressed by the talents of the Chinese students studying in their universities and institutes, and that the students would be great contributors to future research for China.
VI. CODATA 45 YEARS ON VIDEO

Transcript

Over the last five centuries it has to be acknowledged that the quantitative measurement of natural phenomena, materials and processes has played a central part in the evolution of scientific understanding.

The data from Tycho Brahe’s astronomical observations led to Kepler’s solar system model.

Galileo’s measurements of pendulum swings were important in the development of Newtonian mechanics.

Science could not have advanced to its present state if data from each generation had not been preserved and made available to subsequent generations.

Following the end of World War II, science and technology progress led to an exponential expansion in the amount of data published in the scientific literature and compiled in handbooks and repositories.

By the early 1960s, a number of scientific leaders began to realize that the scientific and technical community needs an international effort to address issues of data quality, reliability, management and accessibility. Thus the spirit of CODATA was created.

A distinguished scientist called Frederick Rossini, a member of the National Academy of Sciences, initiated a program to compile and evaluate thermochemical data in the 1930s.

Rossini discussed these data issues with colleagues and found a nucleus of scientists willing to take action.

They proposed a new organization under the umbrella of ICSU, the International Council for Science, and in January 1966, the ICSU General Assembly in Bombay approved the creation of CODATA, the Committee on Data for Science and Technology. CODATA was born!

The first governing body, which was called at that time the Bureau, consisted of six people.

Since Rossini, CODATA has had 11 eminent scientists lead the organization into the 21st century, and our current President, GUO Huadong, was elected in South Africa in 2010.

The first central office of CODATA was established in Washington DC and Dr. Guy Waddington became the first Executive Director.

In 1970, CODATA moved to Frankfurt in Germany, and in 1973, the Secretariat moved to Paris.

Under Boris Vodar’s endeavor, the Secretariat office was graciously offered by the French Ministry of Education in Paris.

One of the many functions of the central office is the organization of the biennial International CODATA Conferences.

These conferences are notable for the various locations of different National Committees with distinctive individual memories.

The first Conference in 1968 in Germany was held in a small private Academy in rural upland country.

The second conference in 1970 was held at the University of St. Andrews.

In 2006, the 20th International Conference took place in this very city. It was one of the best attended events with over 600 participants.

One of the successes of CODATA 2008 in Kyiv was that it was the first time to introduce a strong young scientist agenda into the scientific framework of the conference.

CODATA’s strongest mechanism in implementing the scientific agenda of the organization over the years has been through its Task Groups. CODATA has led over 60 task Groups over this period. The Fundamental Constant TG is the oldest one still running, which you heard about today.

The first recipient of the prestigious CODATA Prize award was given to Barry Taylor in 2000 for his outstanding work in this area.

Our Task Group on Data Sources in Asian and Oceanic Countries ran over 10 years, as did the Task Group on Physical Property Data for the Chemical Industry.

The most recent task Group meeting on Data Citation Standards and Practices took place in Berkeley in August 2011.

The World Summit on the Information Society 2005, where its then President Professor Iwata addressed delegates on the central role of Science in the information Society.

Its work on polar research and the launch of the Polar Information Commons was in June 2010. His Royal Highness the Crown Prince of Norway and Serene Highness Prince Albert II of Monaco visited the PIC Booth in Oslo to help celebrate the launch.

CODATA’s history would not be complete without highlighting one person, the previous Executive Director, Phyllis Glaeser. In November 1991 CODATA celebrated its 25th anniversary.

This is very appropriate in light of the theme for our next CODATA conference “Open Data and Information for the Changing Planet” in Taipei in October 2012.

CODATA Mission Statement was approved at the 25th CODATA General Assembly:
“The mission of CODATA is to strengthen international science for the benefit of society by promoting improved scientific and technical data management and use.”

To the needs of the future, CODATA identified its leading position in the information era through its Strategic Plan 2006-2012.

In the aspect of initiating and coordinating international programs and data intensive research worldwide, President GUO Huadong raised the Hand-in-Hand Program and established a friendly relationship with IRDR and GEOSS.

CODATA is only part of a much larger family and yes, it does strive to make the planet a better world for our children.

Happy Birthday CODATA!

---

CODATA PRESIDENTS

1st Conference

2nd Conference

Frederick Rossini, USA, 1966-1970
Boris Vodar, France, 1970-1974
Paul Melchior, Belgium, 1974-1978
Masao Kotani, Japan, 1978-1982

William Hutchison, Canada, 1982-1986
David Lide, USA, 1986-1990
David Abir, Israel, 1990-1994
Jacques-Émile Dubois, France, 1994-1998

John Rumble, USA, 1998-2002
Shuichi Iwata, Japan, 2002-2006
Krishan Lal, India, 2006-2010
Huadong Guo, China, 2010-2014
VII. THE DAY IN PICTURES
THE DAY IN PICTURES
ACKNOWLEDGMENTS

CODATA—45 years on was made possible by the dedication and planning of the numerous individuals comprising the Organizing Committee. It was through their behind-the-scenes efforts that the meeting took shape and ultimately materialized as a successful day of speeches, presentations, and celebration. Here, each contributor is formally recognized for his or her hard work in the months, days, and hours leading up to the opening of the event.

Organizing Committee

<table>
<thead>
<tr>
<th>Name</th>
<th>Position/Office</th>
</tr>
</thead>
<tbody>
<tr>
<td>Robert Chen</td>
<td>Secretary General, CODATA</td>
</tr>
<tr>
<td>Kathleen Cass</td>
<td>Executive Director, CODATA</td>
</tr>
<tr>
<td>Cecile Carbonell</td>
<td>CODATA Secretariat</td>
</tr>
<tr>
<td>LI Jianhui</td>
<td>Secretary General, Chinese National Committee for CODATA</td>
</tr>
<tr>
<td>LIU Jie</td>
<td>Director, International Cooperation Office, CEODE</td>
</tr>
<tr>
<td>LIANG Dong</td>
<td>International Cooperation Office, CEODE</td>
</tr>
<tr>
<td>LIU Chuang</td>
<td>Institute of Geographical Sciences and Natural Resources Research</td>
</tr>
<tr>
<td>ZHANG Hui</td>
<td>Chinese National Committee for CODATA</td>
</tr>
<tr>
<td>HONG Tianhua</td>
<td>Deputy Director, HIST</td>
</tr>
<tr>
<td>WANG Changlin</td>
<td>Executive Director, ISDE</td>
</tr>
<tr>
<td>Luke Driskell</td>
<td>International Cooperation Office, CEODE</td>
</tr>
<tr>
<td>LI Chang</td>
<td>Chinese National Committee for CODATA</td>
</tr>
</tbody>
</table>