U.S. – China Roundtable on Scientific Data Cooperation: October 2006 Meeting Summary

US National Committee for CODATA[1]
National Academy of Sciences

Chinese National Committee for CODATA
Chinese Academy of Sciences

Background Information

Justification for the Bilateral Roundtable
Rapidly changing technological capabilities for creating, manipulating, disseminating, and using digital scientific data are producing many new opportunities and challenges. The opportunities arise primarily in data-intensive research and applications, in the integration of diverse data for new results, and in making vast amounts of factual information available for a broad spectrum of users. The inherent challenges are in effectively managing these data resources for optimal access and use, and for developing rational rules and structures for such processes.

Both the United States and China have vigorous ongoing and planned scientific data collection and related research activities. Indeed, there are many good reasons for increased cooperation between the United States and China, both generally and specifically in scientific data activities. In the larger context, there is expanding trade and other forms of cooperation in different areas of research. Within scientific research, various mechanisms— contacts, exchanges, and joint research activities—already exist between numerous government and academic institutions in the two countries, some of which were discussed earlier in the week of the roundtable, in the bilateral governmental S&T Joint Commission. However, none of these projects focuses expressly on cooperation in scientific data activities, policies, and practices. Increased cooperation in scientific data activities can help address significant research problems in our respective countries and promote improved capabilities for cooperation in joint research.

In particular, both countries have extensive and important publicly-funded scientific database activities in all unclassified and non-proprietary discipline and applications areas. Researchers in the United States and China have much to gain by fostering greater access to and exchange of such factual information, and by engaging in more joint research activities, either in developing or improving scientific databases or in using those databases in cooperative research. The international CODATA and our respective national CODATA committees provide a well-established core mechanism for planning and facilitating such cooperative activities. The scientific database communities of both countries, as represented by their CODATA committees, have a mutual interest in promoting greater cooperation.

History of Bilateral Cooperation between the US and Chinese CODATA Committees
Such cooperation has already been initiated through a series of six ad hoc meetings between the two national CODATA committees and expert representatives of other governmental and non-governmental organizations on both sides over the past six years. It began with two bilateral symposiums, the first at the U.S. National Academy of Sciences in Washington DC in March 2000, and the second at the Chinese Academy of Sciences (CAS) in Beijing in December 2000, to discuss scientific data policy and activities. In July 2002, the US CODATA hosted a Chinese delegation from CAS and the Chinese Ministry of Science and Technology (MOST) to continue these discussions of policy and practice with regard to scientific

[1] We gratefully acknowledge the support of the National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health under Contract No. N01-OD-4-2139.

databases in the United States, in preparation for the launch of the China Scientific Data Sharing Program. In October 2003, the CAS and MOST organized an international Conference on Scientific Data Sharing, with participation by five U.S. experts through the US CODATA. In June 2004, the Chinese and U.S. CODATA Committees, in collaboration with the CAS, MOST, and the international CODATA Task Group on Preservation of and Access to Scientific Data in Developing Countries, organized a major international workshop on Strategies for Preservation of and Open Access to Scientific Data. A report summarizing those proceedings has been made available to all of you. This was followed by another Chinese delegation from CAS and MOST to visit the United States in December 2004, hosted by the USNC/CODATA, to discuss further U.S. scientific data policies and practices. The background information on this series of meetings (agendas, participants, summaries) is available at: http://www7.nationalacademies.org/usnc-codata/China_US_Data_Seminars.html/.

Despite this extensive and productive series of meetings—and especially because of it—a more formally established forum for continuing these discussions is considered to be both important and useful to the research communities in both countries. Although these meetings have been considered successful by both sides, they have been too ad hoc and have not been able to maintain a continued focus and agenda--with the exception of data policy issues--with a broader set of potentially interested participants. At the same time, the meetings have demonstrated the desirability and feasibility of holding such discussions through a sustained forum.

Purpose of the Roundtable

The U.S. and Chinese national CODATA committees therefore believe that the time is ripe to establish a U.S – China Roundtable on Scientific Data Cooperation, which will operate for an initial period of four years. Such a forum can serve as a focal point for governmental and non-governmental experts who are either engaged in bilateral cooperative research activities with significant data considerations, or experts interested more specifically in certain bilateral scientific data policy and management issues. The proposed task statement for the Roundtable is as follows:

1. 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.
2. 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.

The following types of possible cooperation initiatives regarding scientific database policies, practices, and activities are suggested for discussion and potential implementation:

1) Exchanging of information and identification of issues concerning scientific data activities, policies and developments in intellectual property law and public information policies, including barriers to data exchange at the national and international levels, which may have implications for database development, access, sharing, and use.
2) Identification of scientific data and information resources that might be translated and made more widely available in our respective countries.
3) Identification of mutual high-priority databases in both countries that should have either mirror sites, or subsets of the contents established in each other's country, and determine how to implement that.
4) Promotion of opportunities for both senior and junior scientists and engineers to visit each other's countries for various periods of time to learn about each other's scientific database activities and to engage in cooperative research in select areas.
5) Promotion of opportunities for university students to visit and study in each other's universities and research centers in areas within CODATA's scope of activities.
6) Exploration of the possibility for joint projects in scientific database development, studies, or training, including topics such as common standards and interoperable systems and techniques; metadata management practices; clearinghouses and portals for data resources; and other topics by mutual agreement.

Structure of the Roundtable
The Roundtable is being organized under the auspices of the U.S. and Chinese national committees for CODATA, respectively. A co-Chair and approximately 6-8 expert members were (??) appointed on each side to provide expertise in the topical areas identified above, including data management, information technology, database law and policy, and science policy. The Roundtable includes representatives from the government and other sponsoring organizations in each country as well. The group's expertise also was augmented by *ex officio* members from other related U.S.-China research activities and programs, both within and outside the respective Academies. This will enable the Roundtable to serve as an integrating forum for the existing cooperative bodies that have an interest in scientific data issues and cooperation between the two countries.

A small steering committee consisting of the U.S. and Chinese co-chairs of the Roundtable and of the U.S. and Chinese CODATA committee Chairs and Directors will coordinate the work of the Roundtable within their respective Academies and with other national and bilateral data-related activities. It should be emphasized that as is the case with other National Academies roundtables, the U.S.-China Roundtable on Scientific Data Cooperation will not be subject to the bias and conflict-of-interest procedures of the U.S. National Academies, and therefore will not issue reports with consensus conclusions and recommendations.

Roundtable Activities
The Roundtable plans are to hold one two-day meeting approximately every nine-to-twelve months for an initial four-year period, alternating between the United States and China. These meetings will be focused on developing a detailed understanding of the issues and interests of the committee's sponsoring organizations in cooperative scientific data activities, and on establishing an agenda of future issues and activities. The core two-day meeting will be supplemented by additional site visits and side meetings on a mutually agreed upon basis. The two national CODATA committee Web sites will be used to publicize the activities of the Roundtable, subject to the U.S. National Academies restrictions on the public dissemination of information related to Roundtable discussions. Information will be exchanged among the members of the Roundtable steering committee throughout the year.

In addition to the Roundtable meetings, the two national CODATA committees may agree to organize separately authorized and funded workshops, symposia, or conferences consistent with the interest of the roundtable participants and sponsors. Requests for consensus studies and other advisory products will be referred to the Roundtable's parent bodies, or to other appropriate groups within the respective Academies. After the first three years of operation, the parties involved will evaluate the Roundtable's effectiveness and the desirability of continuing it.

The topical areas that have been identified jointly for initial discussion include scientific data policy, cyberinfrastructure data applications, health and biomedical data, and environmental and geospatial data. All four of these subject areas will be on the agenda of the CODATA Conference as well.

Background on US and China CODATA Committees

The U.S. National Committee for CODATA

The U.S. National Committee (USNC) for the Committee on Data for Science and Technology (CODATA) functions as a bridge between the scientific and technical community in the United States and the international CODATA on data issues addressed in the natural and social sciences. The USNC operates within the National Research Council's Board on International Scientific Organizations and is the principal organizational entity of BISO's Office of International Scientific and Technical Information Programs. The committee pursues these objectives through consensus studies, conferences, workshops, discussions forums, Web sites, and semi-annual committee meetings.

USNC/CODATA Committee Members as of October 2006

Roberta Balstad, Chair, Senior Research Scientist, Earth Institute**,** Columbia University
Hal Abelson, Professor of Electrical Engineering and Computer Science**,** MIT
Peter W. Arzberger, Director, Life Sciences**,** University of California, San Diego
Helen M. Berman, Director, Protein Data Bank. Rutgers University
Christine Borgman, University of California, Los Angeles, Department of Information Studies
Bonnie Carroll, CENDI Executive Director, and President of Information International Associates
Sara Graves, Director, Information Technology and Systems Center. University of Alabama in Huntsville
James N. Gray, Distinguished Engineer**,** Microsoft Bay Area Research Center
Myron Gutmann, Inter-University Consortium for Political and Social Research, University of Michigan
Krishna Rajan, Department of Materials Science and Engineering. Iowa State University
David Scott, Noah Harding Professor of Statistics**,** Rice University
Mary Waltham, Publishing Consultant
STAFF :
Paul F. Uhlir, Director, USNC/CODATA
Raed Sharif, Research Associate, USNC/CODATA
Wendy White, Director, Board on International Scientific Organizations. The National Academies

Recent USNC-CODATA Activities at the National Level

- Symposium on the Role of S&T Data and Information in the Public Domain, *September 2002*
- Symposium on Electronic Scientific, Technical, and Medical Journal Publishing and Its Implications, *May 2003*
- Monitor and assess changes to the domestic and international intellectual property and policy regime relating to S&T data activities (ongoing)
- USNC staff provides institutional coordination on IP-related activities at the Academies
- Maintain *IP @ The National Academies* Web site and distribute a Newsletter (see http://ip.nationalacademies.org)
- Organize a series of discussion forums, most recently on the National Science Board's report on Long-Lived Digital Data Collections

Planned Activities at the National Level

- Symposium and Workshop on the Future of Scientific Knowledge Discovery in the Virtual Environment
- Continuation of biannual USNC/CODATA meetings, discussion forums, and Web-based activities

Recent International Activities

- International Symposium on Open Access and the Public Domain in Digital Data and Information for Science, *Paris, March 2003*
- International Archiving Workshop on the Selection, Appraisal, and Retention of Digital Scientific Data, *Lisbon, Portugal, December 2003*
- Inter-American Workshop on Access to Environmental Data, *Campinas, Brazil, March 2004*
- Workshop on Developing Strategies for Open Access and Preservation of Digital Scientific Data Sources in China, *Beijing, June 2004*, and series of data policy meetings since 2000
- World Summit on the Information Society, coordinating U.S. public science community involvement in WSIS, 2003-2005

- International Workshop on Creating the Information Commons for e-Science: Toward Institutional Policies and Guidelines for Action, *Paris, September 2005*
- International Workshop on Strategies for Permanent Access to Scientific Information in Southern Africa *Pretoria, South Africa September 2005*
- Bilateral U.S.-China Roundtable on Scientific Data Cooperation, *Beijing, October 2006*
- Support of an InterAcademy Panel Initiative on Access to Scientific Information in Developing Countries, workshop *Beijing, October 2006*
- Workshop on Global Information Commons for Science, *Beijing, October 2006*
- Support of international CODATA Conference and General Assembly, *Beijing, October 2006*

Planned International Activities
- Workshop on Permanent Access to Scientific Information in Latin American region in *Atibaia, Brazil, May 2007*
- Workshop on the Socioeconomic Effects of Online Public Sector Information (with the OECD).

Publications: Selected National Academies reports by USNC/CODATA and staff regarding policy and management issues in scientific data and information:
- Strategies for Preservation of and Open Access to Digital Scientific Data in China (2006)
- Open Access and the Public Domain in Digital Data and Information for Science (2004)
- The Role of Scientific and Technical Data and Information in the Public Domain (2003)
- Scientific Data for Decision Making Toward Sustainable Development:: The Senegal River Basin Case Study (2002)
- A Question of Balance: Private Rights and the Public Interest in Scientific and Technical Databases (1999)
- Bits of Power: Issues in Global Access to Scientific Data (1997)
- Preserving Scientific Data on Our Physical Universe: A New Strategy for Archiving the Nation's Scientific Information Resources (1995)
- Finding the Forest in the Trees: The Challenge of Combining Diverse Environmental Data (1995)

Reports available from National Academies Press at http://www.nap.edu

Chinese National Committee (CNC) for CODATA

China joined CODATA in 1984 represented by the Chinese Academy of Sciences as the National Member. The Chinese National Committee for CODATA ever since has been making efforts on promoting China's data activities of science and technology through its executive body—the Secretariat located at Computer Network Information Center of the Chinese Academy of Sciences.

Main Activities of the Chinese National Committee for CODATA
- Investigate and know well about the domestic data activities and trends for establishing databases on science and technology (S & T). The CNC/CODATA has broadcast the domestic S & T data activities to the whole world.
- Hold an annual General Assembly of the CNC/CODATA, with the purpose of discussing its activities, introducing on-going tasks of its data groups and promoting the academic achievement of CODATA.
- Push domestic S & T information resources onto digital networks, thus to make more efficient use of the S & T data.
- Organize domestic working groups and relevant meetings within different academic spheres.
- Hold domestic and international academic activities.
- Organize participants to attend biennial CODATA international Conferences and other relevant meetings.
- Sponsor relevant publications.
- Initiate and organize the implementation of programs on scientific database construction and data sharing.

CODATA-China Members as of October 2006

Officers:
Prof. Li Jiayang, President, Chinese National Committee for CODATA. Vice-president, CAS
Prof. Yan Baoping, Vice-President, Chinese National Committee for CODATA. General Engineer, Computer Network Information Center, CAS
Prof. Xu Zhihong, National Delegate, Chinese National Committee for CODATA
Prof. Xiao Yun, Secretary General, Chinese National Committee for CODATA
Director, Scientific Database Center, Computer Network Information Center, CAS

Members (in alphabetic order):
Prof. Chen Yiyu, Director, National Natural Science Foundation of China
Prof. Cao Yongsheng, Institute of Crop Germplasm Resources, Chinese Academy of Agricultural Sciences,
Acad. Chen Jisheng, Academician, Chinese Academy of Engineering
Prof. Fu Bojie, Director, Bureau of Science and Technology for Resources and Environment, CAS
Prof. Fu Xiaofeng, Department of Basic Research, the Ministry of Science and Technology, P.R. China
Prof. Gui Wenzhuang, Director, Informatization Office, Chinese Academy of Sciences
Prof. Guo Huadong, Deputy Secretary-General, Chinese Academy of Sciences
Prof. Ji Jinkui, Director, Department of Informatization Promotion, Ministry of Information Industry, China
Prof. Liu Zhiyong, Director, Information Science, National Natural Science Foundation of China
Prof. Ma Juncai, Director, Information Center, Institute of Microbiology, CAS
Dr. Ouyang Hua, Vice-Director, Institute of Geographical Sciences and Natural Resources Research, CAS,
Prof. Shao Liqin, Inspector, Department of Fundamental Research and High-Technology, the Ministry of Science and Technology, P.R. China
Acad. Sun Honglie, Chinese Academy of Sciences
Prof. Yuan Haibo, Director of National Science and Technology Library
Prof. Zhang Xian' en, Director, Department of Fundamental Research and High-Technology, the Ministry of Science and Technology, P.R. China
Prof. Zhang Xiaozhong, Director, Electron Microscopy Laboratory of Tsinghua University
Prof. Zhu Jinning, Director, Department of International Affairs, China Association for Science & Technology

October 2006 Roundtable Meeting

On October 20th, 2006, the U.S. National Committee for CODATA and the Chinese National Committee for CODATA, under the auspices of their respective Academies of Sciences, organized the first U.S. - China Roundtable on Scientific Data Cooperation. This meeting was the first of a series of meetings that will be organized between the two countries. The four areas identified for framing the scope of discussion were: (a) scientific data policy, (b) cyber-infrastructure data applications, (c) health and biomedical data, and (d) environmental and geospatial data.

The meeting, which was held at Huarun Hotel, Beijing, China, was attended by approximately 50 experts representing academic and government institutions from both sides. The meeting was organized as follows:

- Session One: Background to the Roundtable Initiative
- Session Two: Parallel Breakout Sessions
- Session Three: Plenary Summary of Parallel Session Results

Below is the summary of the activities and discussion took place in these sessions:

Session One: Background to the Roundtable Initiative

The session was opened with welcoming remarks by Prof. Huadong Guo, Deputy Secretary-General of

CAS and Director General of Bureau of International Cooperation, and Prof. Michael Clegg, University of California at Irvine, and Foreign Secretary, U.S. National Academy of Sciences**.** Both Profs. Guo and Clegg emphasized the importance of this and other similar meetings for the scientific communities in both countries and for science conduct and progress in general. They also encouraged the participants to maximize the value of this meeting and future meetings by identifying and discussing main areas of mutual interest and by establishing and strengthening scientific coordination and collaboration channels between scientists and science and technology policy makers from both countries.

Following the welcoming remarks, Prof. Zhihong Xu (CNC/CODATA) and Dr. Roberta Balstad (USNC/CODATA) provided brief overview of both countries CODATA activities (for more information about these activities, please refer the USNC/CODATA and CNC/CODATA sections above).
The first session was concluded by Mr. Paul Uhlir's background information and objectives of the roundtable (please refer to background information section above) and the instructions for the parallel breakout sessions.

Session Two: Parallel Breakout Sessions

The purpose of these breakout sessions was to better identify the scope and organization of the activities under the Roundtable. Each session was facilitated by a Chinese and/or US co-chair, and the discussion was summarized by a Chinese and US rapporteur. Each group was asked to respond to the following set of questions:

1) What is the baseline of existing bilateral activities?
2) What would be the most useful scope and focus for future discussions under the Roundtable that are not already being addressed in other bilateral activities (i.e., what is the value added of such an activity)?
3) What are the highest priority data-related issues or activities in this topical area?
4) What are the goals regarding these issues/activities? What specific outcomes would be most useful?
5) What are the key governmental and non-governmental organizations and experts that should be involved in the specific topics identified?
6) What would be the best way to structure the Roundtable meetings to achieve these goals and outcomes? What should be done between the meetings?
7) What will be the measures of success for the Roundtable?
8) Are there any other issues not covered by the questions above?

Below is a brief summary of each of the parallel sessions.

Parallel Session A: Cooperation on Data Policy

Session A was co-chaired by Dr. Xian'en Zhang and Dr. Roberta Balstad, and was attended by seven Chinese and three American participants. Mr. Paul Uhlir and Dr. Liu Chuang were the assigned Rapporteurs for this session. There was a broad consensus among the participants that because there is no bilateral activity focused on data policy, there are many policy issues that could be jointly addressed that could be valuable for both sides and that both sides could learn from.
In an attempt to identify the highest priority data-related issues or activities in this topical area, the participants agreed that there are four issues regarding the Chinese data sharing legal and policy system that would benefit from greater attention:

Legislative data sharing guarantees. Currently, there are only draft regulations in China on publication of government info (FOIA). There are no specialized laws or regulations about data sharing policy. A legal administration system for scientific data sharing should be developed, covering data law, policy, and economic aspects.
National security and trade secret restrictions. The basic principle is that data should be shared, with exceptions for protection of national and trade secrets.
Intellectual property rights (IPR) balance with public domain. IPRs and user rights in data sharing need to be balanced. Public data should not be copyrighted. Government funders of research can and should set policies for data access. Private data producers may also be encouraged to make data freely available.

Creative Commons licenses (voluntary private contracts that encourage common use) can be used to mediate these rights in a more socially and scientifically beneficial way. This is especially useful for public scientific data and information products.

Better understanding and awareness raising at different levels. There is a general lack of understanding of the importance of data sharing in China. Data policy is of highest priority for MOST, but there are many obstacles to data sharing at the research institution and individual scientist levels.

In trying to identify the some potential areas of cooperation, the following were the top priority areas:

- Scientists frequently do not practice data sharing, more so in China than in the U.S.—what are the barriers and the reasons for not sharing? What are the methods of encouragement and punishment to change these attitudes? Different disciplines have different considerations and should be investigated.
- Evaluation of data management and access policies is an important issue. What are the methods for assessing the costs and benefits of different access regimes—their economic, social, and scientific effects (both direct and indirect and short and long term)? This is another area of very useful potential cooperation under the Roundtable.
- What are the institutional and funding roles and processes for establishing data policies in each country? There are different actors (stakeholders) in the data development process—government at different levels, the scientific community, and commercial interests. What are the short and long-term considerations? What lessons are learned on each side that can be adopted or adapted by the other side?

The following key governmental and non-governmental organizations and experts were identified to be involved in initiation and implementation of the specific topics identified

Chinese side includes:

- Ministry of Science and Technology (MOST)
- The Chinese Academy of Science (CAS)
- Agencies
- Individual data policy experts
- Universities

US side includes:

- Office of Management and Budget (OMB)
- Office of Science and Technology Policy (OSTP)
- Individual data policy experts
- The National Academies (NAS)
- Universities

One way to structure the Roundtable meetings to achieve these goals and outcomes would be by organizing specialized workshops on the topics selected and highlighting successful examples or case studies. Another suggestion was to organize focused study groups that would conduct short visits between meetings and report back at the Roundtable meetings. Also commission white papers between meetings to report to the Roundtable and make progress on agreed topics/stimulate further discussion. Use Web sites on both sides to disseminate results.

The rest of the session was devoted to presentations of current cooperative activities between the two sides, successful cases of data collection and management, and other related technical issues.

Parallel Session B: Cooperation on Cyber-infrastructure Data Applications

This session was co-chaired by Peter Arzberger and Gang Chen and was attended by six Chinese and three American participants. Both Raed Sharif and Chenzhou Cui were the assigned rapporteurs for this session. Other participants included Baoping Yan, Yaonan Zhang, Yaolin Shi, Kai Nan, and Chris Lenhardt.

The session started with a general discussion about the importance of coordination and cooperation between the US and China in the area of data management and policy, especially in fields such as ecological and environmental observation, high energy physics and medical data. The discussion then was focused on the

problems in the data sharing area emphasizing that these problems are not always technical and that more and special attention must be paid to the non-technical problems such as administrative and legal problems. The discussion of the specific outcomes that would be most useful for future cooperation raised the following points:

- Strengthening the current (and establishing new) channels of communication and exchange of information and data between scientific communities in both countries. Forming a task group to study and discuss issues concerning the development in intellectual property rights at both the national and international levels.

- In the area of earth science, promoting the organization of annual short summer courses for students and junior scientists, and other forms of academic exchange of professors and students.

- Identifying specific areas of mutual interest and working on joint projects especially at the university levels.

- Developing scientific training programs as well as cultural training programs for young scientists.

- Forming a task group to work as a coordinating body between scientists and policy makers from both sides.

- Rather than just having the roundtable participants address these issues, organizing a specialized workshop to identify success and failure examples. Short visits and white papers to be commissioned between the meetings are other desirable activities.

- Evaluating data management and access policies, especially the direct and indirect social and economic benefits and costs is an important issue and needs more attention and study.

Several key governmental and non-governmental organizations and experts were identified as potential stakeholders to be involved in initiation and implementation of the specific topics identified

- **International:** CODATA and ICSU
- **U.S.** : NAS, NSF, and USGS
- **China**: CAS and MOST

Among the Chinese participants there were four application areas discussed:

- High-energy physics (Gang Chen, Institute of High Energy Physics, CAS)
- Astronomy (Chenzhou Cui, National Astronomical Observatories, CAS)
- Geosciences (Yaolin Shi, Laboratory of Computational Geodynamics, CAS)
- Environmental science (Yaonan Zhang, Cold and Arid Regions Environmental and Engineering Research Institute, CAS)

There was one presentation by one of the US participants:

- Environmental Science (Chris Lenhardt, Center for International Earth Science Information Network)

Each of these participants gave an overview of their work. There was a discussion as to what would be areas of possible collaboration. In the following discussion we tried to categorize the issues of who decides what can be shared, and how far the sharing can go.

- In the High Energy Physics example, the project decides what can be shared. The data are shared among project participants. The public can use the data after some time. There is a community tradition of sharing. There is a mirror site of data in Italy, for example.
- In the Astronomy project, the project decides who can access the data. The data are shared among project participants, and are released after one year. There is a community tradition of sharing. It was noted that new equipment coming on line (in China) in 2008, which could be a source of data

to be shared.
- In Geosciences, geology data are published. For Earthquake / Seismic data, at the national level seismic data "almost" available on internet. However provincial sensor data are harder to get (evidently different equipment is used). Remote sensing data were not discussed.
- In the Environment area, a Hydrology project on the Black River (Heihe;黑河) was discussed. Some data are available (before 2000); the rest of data are available after 2007 when project ends. Data are shared in project and international partners are welcome – some exist now)

There were many possible areas of collaboration which were not discussed in this content, notably biomedical research, biodiversity, and land-use.

There are some natural barriers to closer collaborations, which are listed below in no priority order:
- **Communications**: Without knowing of what is happening in the other country, it is hard to collaborate. There are significant language barriers. For example, data may be available on a web site, but unfortunately most US researchers are not well versed in Chinese.
- **Infrastructure**: There was a strong and repeated expression of needs for improved bandwidth, especially in the areas of high energy physics and astronomy. There is also a need for facilities to store data.
- **Data discovery, integration, and archiving:** Mechanisms to discover data need to be improved. Then, if one knows where there data are, integrating them is a universal problem, because data are in different formats, schema, standards, language, and so on. Archiving efforts should be planned further in advance of their need.
- **Visas for Chinese researchers** to come to the United States can be a problem. Restrictions limit face-to-face collaborations (important to start a project) and can hurt both US and China science.
- **Education and public outreach**. Currently there is not much encouragement to do public outreach. If there were, it could improve the ability of the community to make data and results available.
- **Funding** for support for communication of ideas between computer scientists and application scientist was also mentioned as an issue (NSF-C). Also, so is support for software maintenance and users.

When asked the question of what would be the one improvement that was possible in the subsequent 9 to 12 months, the three responses were:
- Increased bandwidth to participate in international activities (e.g., to be able to transfer 15Tb in a 2 week period)
- Translate key web sites and make others aware of them (to allow others to use the data at those sites).
- Develop summer courses for graduate students to expose them to new tools and new collaborators.

Finally, the identification of a few key projects that are struggling with data sharing could be invited to the next meeting to describe technical, social, cultural, and policy issues that inhibit or promote sharing. These projects might provide interesting case studies for how data are shared.

Parallel Session C: Cooperation on Environmental and Geospatial Data

This session was co-chaired by Huadong Guo and Mark Parsons and was attended by seven Chinese and five American participants. Both Tingjun Zhang and Zhengxing Wang were the assigned rapporteurs for this session.

The discussion to identify the highest priority data-related issues or activities in this topical area resulted in the identification of the following issues:
- Help and acknowledge data providers
- Establish/promote data-related standards
- Involve data managers and technologists in the research
- Enhance capacity building and specialized interfaces

– Develop appropriate infrastructure for sharing
– Promote the integrity of the data
– Public communication
– Value added aspects

There was a general agreement among the participants that the Roundtable should include participants from high level policy and funding agencies, users, data providers, scientists, and data managers. It would be useful to choose one or two projects with clear societal benefits to explore/study mechanisms that are used in data exchange (examples include dust transport, land use and land cover change, water quality, air pollution monitoring, hazard warnings, and seismological monitoring). It is advisable to choose a project where policy issues are more important than technical issues and to make a clear case to be used in other fields.

The following key governmental and non-governmental organizations and experts were identified as potential stakeholders to be involved in the initiation and implementation of the specific topics identified:

- **International:** CODATA, WDCs,
- **High level–US**: NSF, NOAA, NASA, USGS, DOE, National Archives and Records Administration with experts for data rescue, and NAS
- **High level-China**: CMA, CAS, MOST, CNSF
- **Governmental funding agencies** – what are their concerns?
- **Experts:** data users, data providers, scientists, data managers,

When asked to identify the best way to structure the Roundtable meetings to achieve these goals and outcomes, and what should be done between the meetings, the participants agreed that soliciting inputs from research communities in both countries to identify and describe specific projects would be an effective way to plan for and manage this and future similar activities.

In trying to identify some criteria/indicators to measure the success of the Roundtable, the participants came up with the following list:

- How much data is being exchanged?
- Creating new data exchanges.
- Identify/establish a network of specialists.
- What are new some opportunities of new collaborations?
- Identify data sets that need to be rescued.
- Create better understanding of data exchange barriers.
- Increase technological capacity.
- Reward data providers.
- Improve description of data uncertainties.

Finally, the participants emphasized the following points:

- There is an urgent need to educate and explain to policy makers that data sharing is very important for sustainable development.
- It is important to recognize (and plan accordingly) that promoting data sharing needs financial support.
- There is a need for new/modified methods to recognize data providers.

The rest of the session was devoted to presentations of current cooperative activities between the two sides, successful cases of data collection and management, and other related technical issues.

Parallel Session D: Cooperation on Health and Biomedical Scientific Data

This session was co-chaired by Depei Liu, Chinese Academy of Medical Sciences & Peking Union Medical College and Michael Clegg, University of California at Irvine and Foreign Secretary, U.S. National

Academy of Sciences. The assigned Rapporteurs for this session were Ling Yin, General Hospital of PLA, and Mary Jo Deering, NIH.

Below is the summary of the main points that were raised and discussed at this session:

- The U.S. participants described specific opportunities for collaboration, including the National Institute of Health's programs to support Chinese scientists at NIH and NIH grant-making programs; the National Cancer Institute's cancer Biomedical Informatics Grid (caBIG) project that offers informatics infrastructure, tools, and data resources freely available; the National Institute of Standard and Technology's Standard Reference Data Program.
- NIH and NCI policies and practices on data sharing were also proposed for consideration or use.
- China has achieved fast progress in medical data sharing. It has established a protective system, a standard and regulatory system, a resource planning system and a platform-support system for that purpose.
- The Data Sharing Network for Medical and Health Sciences that is still under construction consists of four scientific data centers and about 300 databases.
- The Data Center of Traditional Chinese Medicine has regional characteristics.
- CODATA could develop an inventory of specific agreements at each level: academic, agency, and governmental between the two countries.

When asked to identify the most useful scope and focus for future work, the participants suggested the use of a pilot project through which all the issues and opportunities can be explored. This pilot could be modest, but could pursue the following issues:

- Data structure and standards
- Database construction
- Interoperability
- Intellectual property, both of the software and of the data content.

The added value of such a pilot project is that it would be difficult to be conducted individually; the collaborative aspect is the greatest benefit. The area of neuroimaging was proposed as one area, since the National Institute for Biological Imaging and Biotechnology is in the process of funding a database in this area and would be willing to open it up for collaboration.

Among the highest priority data-related issues, the following were identified and discussed:

- China just created a Health and Biomedical Network; every year about 30,000 people use it just in China. It can be opened for access by American scientists.
- Collaborative work on standards, including metadata standards.
- GenBank and other publicly accessible resources could be examined for next steps. Identify proteomic or metabolomic data that could be shared.
- Study areas related to aging populations. For example, support cooperative data activities on specific diseases, like cancer, alzheimers, and others.
- Suggest that MOST support a pilot project on medical data sharing.
- Set up a China-US work group under CODATA to explore specific issues. Establish it at the next CODATA meeting. This group could look at policy issues as well as technical issues, including intellectual property and privacy/security issues.

Key governmental and non-governmental potential stakeholders to be involved in initiation and implementation of the specific topics:

- **China**: MOST, National Working Group of Data Sharing for Medicine.
- **USA**: NIH, NSF, and NAS.

Other Issues:

- Health and biomedical data do not receive as much attention in CODATA as other fields; propose elevating health and biomedical data within CODATA.

- China and the USA have cooperated several times in biomedicine research, but mostly among academic groups and experts. Cooperative programs between governments are not many, so it would be useful to promote this kind of cooperation.

Additional Areas of Possible Collaboration:

- Propose an inventory of existing projects on both sides.
- NCI welcomes Chinese collaboration on interoperable informatics projects, including developing infrastructure, tools, and data resources.
- Propose a plan for seminars to help each other understand respective needs and resources.
- Recommend training for students in bioinformatics or in how to construct databases and software. Focus on what kinds of technologies, procedures, policies.
- Exchange information about scientific and informatics meetings and promote attendance by other nation's scientists, informatics specialists, and students.
- Establish a pattern for China-US cooperation in biomedicine, under the general framework of CODATA.
- Set up working group for data sharing of medical and health sciences.
- This group to be responsible for organization of international academic activities of biomedical data sharing.
- Work out serial standards and norms for international data sharing of biomedicine.
- Conduct surveys and need analysis of international data resources of biomedicine and constitute plans for them.
- Institute administrative organs for data convergence, storage and exchange and patterns for data sharing service.
- Create a research environment for network coordination of biomedicine. Carry out global scientific research cooperation.
- On the basis of China-US bilateral cooperation in biomedicine, create an advanced biomedical data sharing network that can render services to all countries in the world.
- Strengthen cultivation and exchange of staff for biomedical data sharing.

The rest of the session was devoted to presentations of current cooperative activities between the two sides, successful cases of data collection and management, and other related technical issues.

Closing Plenary Session: Conclusion and Future activities

The meeting concluded with a summary of the main ideas discussed during the meeting and a call for more future cooperative activities and meetings between the two sides. Both Dr. Roberta Balstad and Dr. Baoping Yan led this concluding session and encouraged scientists and policy makers from both sides to pay more and special attention to these issues and to create an enabling environment for better organized cooperation. Finally, the session chairs thanked the participants for their time and effort to make this a successful meeting.

Appendix A: Meeting Agenda

U.S. – China Roundtable on Scientific Data Cooperation

Organizing Meeting
of the
Chinese National Committee for CODATA
International Cooperation Bureau
Chinese Academy of Sciences
and
U.S. National Committee for CODATA
Board on International Scientific Organizations
The National Academies, U.S.

China Resources (Huarun) Hotel
Beijing, China

20 October 2006

Preliminary Agenda

Session One: Background to the Roundtable Initiative

9:00 Welcoming Remarks
Huadong Guo (to be confirmed), Deputy Secretary-General of CAS, Director General of Bureau of International Co-operation;
Michael Clegg, University of California at Irvine, and Foreign Secretary, U.S. National Academy of Sciences

9:20	Brief Overview of Chinese CODATA Activities	**Zhihong Xu** China CODATA
9:35	Brief Overview of US CODATA Activities	**Roberta Balstad** Columbia University Chair, US CODATA
9:50	Background and Objectives of the Roundtable - Instructions for Breakout Sessions	**Paul Uhlir** US CODATA

10:10 *Coffee break*

Session Two: Parallel Breakout Sessions

A. Cooperation on Data Policy

10:30 Introductory comments

Co-Leader:
Xian'en Zhang, Director General of department of basic research, Ministry of Science and Technology; member of China CODATA
Roberta Balstad, Columbia University
Rapporteur:
Paul Uhlir, US CODATA
Chuang Liu, the Institute of Geographical Sciences and Natural Resources Research, CAS

11:00 Discussion of breakout session questions
- *General discussion*

B. Cooperation on Cyber-infrastructure Data Applications

10:30 Introductory comments

Co-Leader:

Gang Chen, the Instiute of High Energy Physics, CAS
Peter Arzberger, University of California at San Diego

11:00 Discussion of breakout session questions
- *General discussion*

Rapporteur:

Chenzhou Cui, the National Astronomical Oservations, CAS
Raed Sharif, Syracuse University and US CODATA

C. Cooperation on Environmental and Geospatial Data

10:30 Introductory comments

Co-Leader:

Huadong Guo Deputy Secretary-General of CAS, Director General of Bureau of International Co-operation, Session Co-Leader
Mark Parsons, University of Colorado, University of Colorado, U.S.

Rapporteur:

Tingjun Zhang, Center of International Research in Environmental Science, University of Colorado
Zhengxing Wang, the Institute of Geographical Sciences and Natural Resources Research, CAS

11:00 Discussion of breakout session questions
- *General discussion*

D. Cooperation on Health and Biomedical Data

10:30 Introductory comments

Co-Leader:

Depei Liu, Chinese Academy of Medical Sciences & Peking Union Medical College, Session
Michael Clegg, University of California at Irvine

11:00 Discussion of breakout session questions
- *General discussion*

Rapporteur:

Ling Yin, General Hospital of PLA
Mary Jo Deering, National Cancer Institute, National Institutes of Health

12:30 *Lunch*

Session Two: Parallel Breakout Sessions (continued)

14:00 ***Sessions A, B, C, D***
Discussion of breakout session questions
General discussion

15:30 *Tea break*

Session Three: Plenary Summary of Parallel Session Results

16:00 Presentations of responses to breakout sessions questions
- *General discussion*

Session Four: Observations on the Future of the Roundtable

17:00 Summary of next steps — **Roberta Balstad**, US CODATA; **Baoping Yan**, China CODATA

17:30 *End of meeting*

18:00 *Reception*

Appendix B: Meeting Participants

Chinese Participants:

Name	Institution	
Baoping Yan	Computer Network Information Center, CAS	
Baoyan Liu	China Academy of Traditional Chinese Medicine	
Bingxin Yang	Xiangshan Science Conferences	
Bisong Liu	Informatization Standardization Institute, China National Institute of Standardization	
Changlin Wang	the Institute of Remote Sensing Application, CAS	
Jing Li	National Disaster Reduction Center of China, Beijing Normal University	
Mianzhen Teng	Deputy Director General of department of basic research, MOST	
Ping Zhang	Peking University Law School	
Ruifeng Liu	Director of China Earthquake Information Center, Earthquake Scientific Data Sharing Project	
Ruijun Wang	Director of one department of R&D Infrastructure and Facility Centre, MOST	
Shaohua Lin	Director of National Marine Data and Information Service	
Xiaobo Li	the Ministry of Land and Resources	
Xin Lin	Department of Policy, Regulations and Reform, Ministry of Science and Technology, PRC	
Yaxi Guo	Department of Scientific and Technological, the China Meteorological Administration	
Yiyuan Tang	Institute of Neuroinformatics & Lab for Brain and Mind, Dalian University of Technology	
Zhengxing Wang	the Institute of Geographical Sciences and Natural Resources Research, CAS	
Jin'an Cai	China Earthquake Administration	
Chenzhou Cui	the National Astronomical Observatories, CAS	
Dingsheng Liu	the Remote-Sensing Satellite Ground Station, CAS	
Dongsheng Zhao	Academy of Military Medical Sciences	
Xiangtao Fan	the Institute of Remote Sensing Application, CAS	
Xiaofeng Fu	Researcher, Department of Basic Research, Ministry of Science and Technology, PRC	
Gang Chen	the Institute of High Energy Physics, CAS	
Huadong Guo	Deputy Secretary-General, CAS	

Chuang Liu	Researcher, Institute of Geographic Sciences and Natural Resources Research, CAS	
Shunbao Liao	WDC for Renewable resources and environment; the Institute of Geographical Sciences and Natural Resources Research, CAS	
Depei Liu	Chinese Academy of Medical Sciences & Peking Union Medical College, Session	
Xin Li	WDC for Glaciology; Cold and Arid Regions Environmental and Engineering Research Institute, CAS	
Yaolin Shi	CAS Laboratory of Computational Geodynamics	
Jiulin Sun	the Institute of Geographical Sciences and Natural Resources Research, CAS	
Gonghuan Yang	Chinese Center for Disease Control & Prevention	
Yaonan Zhang	researcher, Cold and Arid Regions Environmental and Engineering Research Institute, CAS	
Ling Yin	General Hospital of PLA	
Xian'en Zhang	Director-General, Department of Basic Research, Ministry of Science and Technology, PRC	
Zhihong Xu	Researcher, Institute of Process Engineering, CAS	

U.S. Participants:

Name	**Institution**	
Roberta Balstad	Columbia University and Chair of US CODATA	
Michael Clegg	University of California at Irvine, and Foreign Secretary, U.S. National Academy of Sciences	
Paul Uhlir	The U.S. National Academies	
John Wilbanks	Science Commons	
Peter Arzberger	University of California at San Diego	
Chris Lenhardt	Center for International Earth Science Information Network, Columbia University.	
Mark Parsons	University of Colorado	
John Faundeen	Earth Resources Observing System Data Center, U.S. Geological Survey	
Tingjun Zhang	Center for International Research in Environmental Science, University of Colorado	
Raymond Willemann	Incorporated Research Institutions for Seismology	
David Clark	National Geophysical Data Center, National Oceanic and Atmospheric Administration	
Mary Jo Deering	National Cancer Institute, National Institutes of Health	
Sandy Ressler	National Institute for Standards and Technology	
Belinda Seto	National Institute for Biomedical Imaging and Bioengineering, National Institutes of Health	
James Herrington	Fogarty International Center, National Institutes of Health	
Bruce Ross	U.S. Embassy in Beijing.	
Raed Sharif	Syracuse University and The U.S. National Academies	