Strategies for Preservation of and Open Access to Scientific Data in China: Summary of a Workshop

Paul F. Uhlir and Julie M. Esanu, Rapporteurs, U.S. National Committee for CODATA, National Research Council

ISBN: 0-309-66171-4, 152 pages, 6 x 9, (2006)

Strategies for Preservation of and Open Access to SCIENTIFIC DATA IN CHINA

Summary of a Workshop

Paul F. Uhlir and Julie M. Esanu, Rapporteurs

U.S. National Committee for CODATA
Board on International Scientific Organizations

NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, DC
www.nap.edu

THE NATIONAL ACADEMIES PRESS • 500 Fifth Street, NW • Washington, DC 20001

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

Support for this project was provided by the National Institutes of Health (under Grant No. 467-MZ-400266), the National Science Foundation (under Grant No. GEO-0407487), and the Open Society Institute (under Grant No. 40006515). Additional support was provided by the Ministry of Science and Technology of China, Chinese Academy of Sciences, Chinese National Natural Science Foundation, the Committee on Data for Science and Technology, the International Council for Science, and the United Nations Educational, Scientific and Cultural Organization. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the organizations or agencies that provided support for the project.

International Standard Book Number 0-309-10230-8

Copies of this report are available from the Board on International Scientific Organizations, 500 Fifth Street, N.W., Washington, D.C. 20001; 202-334-2807; Internet, http://www7.nationalacademies.org/biso/.

Additional copies of this report are available from the National Academies Press, 500 Fifth Street, N.W., Lockbox 285, Washington, DC 20055; (800) 624-6242 or (202) 334-3313 (in the Washington metropolitan area); Internet, http://www.nap.edu.

Printed in the United States of America

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Ralph J. Cicerone is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Wm. A. Wulf is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Ralph J. Cicerone and Dr. Wm. A. Wulf are chair and vice chair, respectively, of the National Research Council.

www.national-academies.org

STEERING COMMITTEE ON STRATEGIES FOR PRESERVATION OF AND OPEN ACCESS TO SCIENTIFIC DATA IN CHINA*

WILLIAM ANDERSON (Co-Chair), Praxis101
XIAN'EN ZHANG (Co-Chair), Ministry of Science and Technology of China
PETER ARZBERGER, University of California, San Diego
ROBERT CHEN, Center for International Earth Science Information Network
HUADONG GUO, Department of International Cooperation, Chinese Academy of Sciences
HEATHER JOSEPH, BioOne
CHUANG LIU, Global Change Information and Research Center, Chinese Academy of Sciences
BAOPING YAN, Computer Network Information Center, Chinese Academy of Sciences

Principal Project Staff

PAUL UHLIR, Project Director
JULIE ESANU, Program Officer
VALERIE THEBERGE, Communications Associate
AMY FRANKLIN, Program Associate

Representatives of Chinese Program Committee

JUN CHEN, China Association for Geographical Information Systems
JIANGUO HAN, National Natural Science Foundation of China
DINGCHENG HUANG, Institute of Geophysics, Chinese Academy of Sciences
ZONGQI SHEN, Ministry of Science and Technology of China
MIANZHEN TENG, Ministry of Science and Technology of China
JINNING ZHU, Chinese Association for Science and Technology

*Chinese names are presented here, and throughout the rest of this report, in the Western approach to personal names, with given name first and family name second.

U.S. NATIONAL COMMITTEE FOR CODATA

ROBERTA BALSTAD (Chair), Center for International Earth Science Information Network
WILLIAM ANDERSON, Praxis101
PETER ARZBERGER, University of California, San Diego
KATE BEARD, University of Maine
HELEN BERMAN, Rutgers University
CHRISTINE BORGMAN, University of California, Los Angeles
BONNIE CARROLL, Information International Associates
JANET GOMON, Smithsonian Institute*
SARA GRAVES, University of Alabama, Huntsville
MYRON GUTMANN, Inter-University Consortium for Political and Social Research
JAN HOPMANS, University of California, Davis*
MAUREEN C. KELLY, Consultant*
GARY KING, Harvard University*
REYNALDO MORALES, Los Alamos National Laboratory, retired
KRISHNA RAJAN, Iowa State University
JAMES SWEENEY, Consultant*
ALEXANDER SZALAY, The Johns Hopkins University

Staff

Paul Uhlir, Director
Julie Esanu, Program Officer (until May 2005)
Amy Franklin, Program Associate

*Member until June 30, 2004.

Preface

The U.S. National Committee for CODATA in collaboration with the Chinese National Committee for CODATA and the CODATA Task Group on Preservation of and Access to Scientific and Technical Data in Developing Countries, and together with the Chinese Ministry of Science and Technology, jointly organized an international workshop on "Strategies for Preservation of and Open Access to Scientific Data." Local logistical support and meeting planning was provided by the Chinese Academy of Sciences and the Chinese Association for Science and Technology. The International Council for Scientific and Technical Information provided additional substantive expertise.

The workshop, which was held June 22-24, 2004, in Beijing, China, served as an international and interdisciplinary forum to promote a deeper understanding of, and requirements for, long-term preservation and open access to digital scientific information resources. The meeting was organized into seven sessions, each led by two co-chairs and including several invited presentations. The first two sessions introduced the workshop participants to the Chinese strategies and initiatives for implementing scientific data sharing. The following sessions explored the policy and legal, institutional and economic, management and technical, and local and regional issues in preserving and providing open access to data in the life sciences and public health; data in the earth and environmental sciences; and scientific information, journals, and digital libraries. This publication presents a summary of the workshop.

The statements made in this summary are those of the individual rapporteurs based on presentations made at the workshop and do not necessarily represent the views of the steering committee, the U.S. or Chinese National Committees for CODATA, or the sponsoring organizations in China or the United States. This volume does not contain summaries of all of the presentations.

Paul F. Uhlir
Director, U.S. National Committee
for CODATA

Julie M. Esanu
Program Officer, U.S. National Committee
for CODATA (until May 2005)

William L. Anderson
Praxis101
Planning Committee Co-Chair

Chuang Liu
Chinese Academy of Sciences
Planning Committee Co-Chair

Acknowledgments

The U.S. National Committee for CODATA and the Board on International Scientific Organizations of the National Research Council of the National Academies wish to express their sincere thanks to the many individuals who played significant roles in planning the International Workshop on Strategies for Preservation of and Open Access to Scientific Data. The Workshop Steering Committee was chaired by Xian'en Zhang, Ministry of Science and Technology, China, and William Anderson, Praxis101, United States. Additional members of the Steering Committee were Peter Arzberger, University of California at San Diego, United States; Jun Chen, China Association for Geographical Information Systems; Robert Chen, Columbia University, United States; Huadong Guo, Chinese Academy of Sciences; Jianguo Han, National Natural Science Foundation, China; Dingcheng Huang, Chinese Academy of Sciences; Heather Joseph, BioOne, United States; Chuang Liu, Chinese Academy of Sciences; Zongqi Shen, Ministry of Science and Technology, China; Mianzhen Teng, Ministry of Science and Technology, China; Baoping Yan, Chinese Academy of Sciences; and Jinning Zhu, Chinese Association for Science and Technology.

We also would like to thank the following individuals (in order of appearance) who made presentations during the workshop (see Appendix A for symposium agenda): Jinpei Cheng, Ministry of Science and Technology, China; Roberta Balstad, Columbia University, United States; William Anderson; Zhihong Xu; Xian'en Zhang; QIN Dahe, China Meteorological Administration; Depei Liu, Chinese Academy of Medicine, Chinese Academy of Engineering; Qiheng Hu, Chinese Association of Science and Technology; Yasuyiki Aoshima, UNESCO; Peter Schröder, Ministry of Education, Culture, and Science, The Netherlands; Carthage Smith, International Council for Science; Michael Clegg, U.S. National Academy of Sciences; Peter Weiss, U.S. National Weather Service; Raymond McCord, Oak Ridge National Laboratory, United States; Menas Kafatos, George Mason University, United States; Qinmin Wang, Fujian Province, China; Jun Chen, Basic Geographical Information Center, China; Jerome Reichman, Duke University Law School, United States; Paul Uhlir, U.S. National Academies; Panqin Chen, Chinese Academy of Sciences; Tieqing Huang, Chinese Academy of Sciences; Lan Zeng, National Macro Economic Research Institute, China; Jun Li, National Macro Economic Research Institute, China; John Willinsky, University of British Columbia, Canada; Chuang Liu, Chinese Academy of Sciences; Belinda Seto, National Institutes of Health, United States; Yun Xiao, Chinese National Committee for CODATA; Shunbao Liao, Chinese Academy of Sciences; Theodore Carl Bergstrom, University of California at Santa Barbara, United States; Yixue Li, Shanghai Biomedical Center, China; Anne Linn, U.S. National Academies; Chengquan Sun, Chinese Academy of Sciences; Zukang Feng, Protein Data Bank, United States; Helen Doyle, Public Library of Science, United States; Shuichi Iwata, University of Tokyo, Japan; Honglie Sun, CODATA; Ling Yin, Chinese PLA General Hospital; Yiyuan Tang, Dalian Polytech University, China; Baoyan Liu, Chinese Academy of Medicine; Donglie Qin, The Capital Hospital University, China; James Edwards, Global Biodiversity Information Facility, Denmark; Xiaofeng Fu, Administrative Center for China's Agenda 21; Khudulmar Sodov, National Remote Sensing Center, Mongolia; Dingsheng Liu, Chinese Academy of Sciences; Jiansheng Qu, Chinese Academy of Sciences; Zhengxing Wang, Chinese Academy of Sciences; Paul Richards, Columbia University, United States; Raymond Willemann, GEM Technologies, United States; Dake Yang, China Earthquake Administration; Xiaolin Zhang, Chinese Academy of Sciences; Lulama Makhubela, National Development Agency, South Africa; Liansheng Meng, Chinese Academy of Sciences; and Pippa Smart, International Network for the Availability of Scientific Publications, United Kingdom.

This report has been reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise, in accordance with procedures approved by the National Academies' Report Review Committee. The purpose of this independent review is to provide candid and critical comments that will assist the institution in making its published report as sound as possible and to ensure that the report meets institutional standards for quality and objectivity. The review comments and draft manuscript remain confidential to protect the integrity of the process.

We wish to thank the following individuals for their review of this report: Heather Joseph, SPARC; Goetz Oertel, Consultant; John Rumble, Consultant; and Wang Zhengxing, Global Change Information and Research Center.

Although the reviewers listed above have provided many constructive comments and suggestions, they were not asked to endorse the content of the report, nor did they see the final draft before its release. Responsibility for the final content of this report rests entirely with the authors and the institution.

The U.S. National Committee for CODATA would like to recognize the contributions of National Research Council staff and consultants. Paul Uhlir, Director of International Scientific and Technical Information Programs, was project director of the symposium and served as the primary editor of this report. Julie Esanu helped to organize the symposium and edit the report. Valerie Theberge organized and coordinated the logistical arrangements, and Amy Franklin assisted with the production of the manuscript.

In addition, the committee would like to thank the other individuals who contributed to the success of the workshop: Wenneng Zhou, secretary-general of the local organizing committee, and Kathleen Cass, executive director of CODATA.

Contents

Appendixes

This study is dedicated in fond memory of
Peter Weiss,
National Weather Service

1

Introduction

As a major producer of scientific data,[1] and as a partner for international cooperative research, China has a great deal to offer to the world's knowledge base. Although China's research capabilities are rapidly improving, some significant problems remain. Among the recognized impediments are inadequate digital archiving and access policies and practices that inhibit progress and improved international cooperation. Particularly noteworthy in this regard was the high-level data access initiative announced in February 2003 by Guanghua Xu, the Chinese Minister of Science and Technology (MoST), and supported by the National People's Congress. This China Scientific Data Sharing Program includes "creating a law to ensure that scientific information is communicated more widely, and coordinating efforts by government departments to develop information centers and databases to facilitate the communication of scientific and technological information."[2] This new policy toward greater openness with publicly funded scientific data is part of a broader effort to modernize the national research and development infrastructure and its management in China.

In 2000, the U.S. National Committee (USNC) for CODATA and the Chinese National Committee for CODATA held two bilateral meet-

[1] In the context of this report, "scientific data" refer to scientific, technical, and medical data, and literature.

[2] Hepeng Jia, "China urges its researchers to share data," SciDevNet, March 14, 2003.

ings with senior science officials and data managers from both countries to discuss various data management and policy issues.[3] Following these two initial meetings of the U.S. and Chinese CODATA committees, the Chinese side was augmented by other experts from the Chinese Academy of Sciences and MoST who are leaders of the Scientific Data Sharing Program. The USNC for CODATA hosted a delegation of these Chinese data policy experts in the summer of 2002.

The fourth of these bilateral meetings of data experts was held in Beijing in October 2003. Focused on scientific resources sharing policy, that meeting provided some of the advance groundwork for the June 2004 workshop that is the topic of this report. It also re-confirmed the commitment of the Chinese science policy community to promoting greater openness regarding Chinese scientific data and identified the priority areas for additional focus.

The effective long-term preservation of and open access to digital scientific resources in all countries increases in importance as an essential component of the global public research infrastructure, which can now be integrated through the Internet. The challenges in storing and maintaining access to these growing collections of data and information are substantial, even in economically more developed countries. Moreover, although many of the challenges that require sustainable solutions are the same for digital data and information across all disciplines, others are distinct or unique to certain disciplines or types of information. And while all solutions are context dependent, some may be based on extending or emulating existing successful models, and others may require and benefit from entirely new approaches.

China faces substantial hurdles in this regard. Although many of its data resources and especially journal literature still reside in paper formats, China already has significant digital information preservation and access requirements that in many cases are not being successfully addressed. Factual databases and journals can provide an important research and economic tool for China—just as they do in more economically developed countries—for capacity building in science and education, for supporting sustainable development of commerce and industry, and for promoting good governance. Resolving the many difficulties in preserving and making

[3]For additional information, see http://www7.nationalacademies.org/usnc-codata/China_US_Data_Seminars.html.

broadly available the digital scientific data resources successfully today will provide great benefits for future generations; the costs of inaction are incalculable, but certain to be substantial. At the same time, it is important to recognize that even the most economically developed countries have encountered various difficulties with the preservation and open access issues. Careful consideration is needed to develop long-term plans for sustainable digital archiving in China, as in all other countries.

WORKSHOP ON STRATEGIES FOR PRESERVATION OF AND OPEN ACCESS TO SCIENTIFIC DATA

The international Workshop on Strategies for Preservation of and Open Access to Scientific Data was held on June 22-24, 2004, in Beijing. It built on the results of the four previous bilateral CODATA meetings and on the new Chinese scientific data access policy initiative noted above. The workshop explored in detail the various scientific and technical, legal and policy, institutional and economic, and management aspects that need to be addressed in successfully implementing sustainable and accessible archives of digital health and environmental data resources in China. It examined various models of open archiving that might be adopted or adapted for use within the Chinese context. It also provided much needed high-level attention to these typically under-appreciated problems by bringing together scientific information managers, digital archiving experts, national science policy and funding officials, and representatives of development organizations, who will be able to incorporate the results of this project into their future planning.[4]

The workshop was organized pursuant to the following statement of task:

1. Identify research areas in which preservation of and open access to digital scientific information require high-priority attention in China, and provide the underlying rationales for the areas chosen.

2. Identify and discuss the scientific and technical, institutional and economic, legal and policy, and management factors relevant to providing open access to digital scientific information resources (both the data and the literature), including an examination of different possible models and

[4]See Appendix B for the biographical summaries of all the speakers at the workshop.

their potential benefits and shortcomings in China, and drawing on examples of other digital archiving and access regimes in related areas.

3. Review and discuss the current status of access and archiving regimes for the types of scientific information identified in task 1.

4. Identify possible follow-up activities to improve open access and preservation for each major type of digital scientific information selected for discussion in task 1, taking into consideration the results of the discussions under tasks 2 and 3.

The workshop addressed these four tasks over two and a half days through a mix of invited presentations, focused panel presentations, and some discussion by all of the participants in both plenary and breakout sessions.[5] The primary areas of focus were on biomedical data, earth and environmental data, and related scientific, technical, and medical literature. Although many of the issues identified and discussed during the workshop were focused on the Chinese context, they also likely are relevant throughout much of the developing world. Because of significant language barriers and time constraints the identification of follow-up activities requested in task 4 could not be done through group discussions in the breakout sessions. Any potential follow-up activities were identified only in the context of individual presentations.

STRUCTURE OF THIS REPORT

Because this report is a summary of the workshop, it is limited in scope to the presentations and other information identified during the meeting. Chapter 2 presents two keynote speeches in their entirety by high-ranking officials at the Ministry of Science and Technology who describe the development and status of China's national Scientific Data Sharing Policy.

Subsequent chapters include summaries of the other speakers' presentations. Several international perspectives on the preservation of and open access to public scientific data are presented in Chapter 3. Chapter 4 discusses the cross-disciplinary issues—policy and legal, institutional and economic, and management and technical—that affect the preservation of and open access to public scientific information. The report concludes with a discussion of these issues in the areas of life sciences and public health data;

[5] See Appendix A for the workshop agenda.

earth sciences, environmental, and natural resources data; and scientific information, journals, and digital libraries.

The appendixes to the report provide additional background information, including the workshop agenda and the biographical summaries of the workshop speakers. The presentation materials used by the invited speakers are available in English on the USNC for CODATA Web site at http://www7.nationalacademies.org/usnc-codata/chinese_workshop.html.

2

Presentations on China's Scientific Data Sharing Policy and Project

THE DEVELOPMENT OF CHINA'S SCIENTIFIC DATA SHARING POLICY

Jinpei Cheng
Vice Minister, Ministry of Science and Technology of China

At the present rate of rapid advances in modern science and technology many areas, such as cosmology, earth system science, cognitive science, and nonlinear science, are becoming the new scientific frontiers. The integration of information science, bioscience, and materials science, as well as the interaction between natural sciences and social sciences, marks our entrance into a decade of unprecedented intensive knowledge and innovation.

Since opening to the outside world, the Chinese government has attached much importance to the development of science and technology, as evidenced by the national strategies of "National Renewal through Science and Education" and "Realization of a Prosperous China through Human Education." The Chinese government recognizes that knowledge and innovation are essential elements for realizing the goals outlined in these strategies, and the key components for obtaining and fostering national and international competitiveness. In recent years, China has achieved a series of advances in science and technology, including manned space flight, super-hybridized rice breeding, and supercomputer research and development. These achievements highlight some of China's ability for innovation. Science and technology are playing important roles in agricultural advancement, industrial technology upgrades, socially sustainable development, and the evolution of China's advanced-technology industries.

In 2002, the Chinese government established a new initiative with the central objective of building an affluent society throughout the country. In 2003, the Third Plenary Session of the 16th Central Committee of the Chinese Communist Party formulated and adopted five scientific development goals that collectively direct the national economic and social development. A large gap still exists, however, between the requirements for economic and social development in China and the capacity of its science and technology to meet these requirements. For example, insufficient investment in science and technology infrastructure, lack of world-class research teams, and an outdated research management system greatly constrain China's innovation and international competitive ability in science and technology. Among these constraining factors, the inefficient use of scientific and technological resources and the repetitiveness and duplication of research efforts have been prominent bottlenecks to China's advancement in innovation.

In 2002, the State Council authorized the Ministry of Science and Technology (MoST) to initiate a pilot project of the national science and technology infrastructure platform in coordination with 16 other ministries and departments. Based on reforms that strengthen data sharing and scientific resource system integration, this project focuses on increasing China's international competitiveness and science and technology innovation potential. China will use modern information technology and international resources to construct the public, fundamental, and strategic science and technology infrastructure platform. The primary objective is to create an environment that fosters scientific and technological innovation by providing the necessary infrastructure that best enables advancements in science and technology, and that best supports long-term developments that are the cornerstones of discovery and innovation. The sharing of scientific data is the core component in this project.

MoST regards scientific data sharing as a national science and technology infrastructure platform, considers it of national interest, and treats it as an important research component based on several factors. Scientific data are the most active and innovative resource in the information era. Scientific data have remarkable research and application possibilities, and decision-making potential. They are fundamental for meeting the needs for the advancement and innovation of science and technology, and for social development, economic growth, and national security. The value of scientific data resources is strongly correlated with their sharing characteristics in two respects. First, scientific research requires the use and sharing of data, information, and knowledge from other pertinent disciplines to make innovative advancements most efficiently and effectively. Solutions to world problems require an interdisciplinary approach. Second, the envisioned Chinese scientific data sharing platform would facilitate simultaneous and unlimited copying and uses of data. The scientific data-sharing platform will thus be based on the principle that all data be fully shared to facilitate the use of the data for fully realizing their maximum value.

Scientific data are derived from scientific and technological activities such as observation, monitoring, investigation, experiments, and research analysis in various organizations and institutions. The types of scientific data include numerical, spatial, graphical, and text data, which are complex, widely distributed, in multiple formats, and massive in scope.

Scientific data are a knowledge resource for the whole society, so China should manage the data sharing to best serve the entire country. China's massive data holdings are obtained and accumulated as a result of national investment plans. As such, they are a national asset and resource shared by the whole society. The capital accumulated by the taxpayers is used to obtain scientific data. Therefore, the taxpayers should have the right to access and share the data resources, which means that the producer of public data must distribute the data to the public and serve the whole country and society. It is on this basis that the data should be shared.

The insufficient use of China's massive data holdings has been an urgent problem. Over time, the Chinese government has organized different observations, surveys, and experiments in many scientific fields, and accumulated large quantities of scientific data. However, most of these data are not shared or used efficiently because of the lack of a policy of openness and the lack of a mechanism for sharing the data. This results in great waste since the national investment is used primarily for repeatedly collecting basic data. For example, there are 18 data receivers in the United States for the moderate resolution imaging spectroradiometer sensors on National Aeronautic and Space Administration (NASA) Earth Observing Satellites, while in China there are more than 30 stations, and their number continues to increase. Overall, the existing 5,000 to 6,000 scientific databases in China do not really support the development of the country and society. One can see that the problem of not sharing scientific data is pervasive because of policy, management, technology, and other factors. As a result, it is difficult for scientists to obtain the needed research information, and precious national resources are wasted due to repeated and redundant collection. Therefore, scientific data sharing and the means for China to use its resources more effectively are essential.

The development of international scientific data sharing policies has provided good references for the Chinese data sharing project. In recent years, many developed countries have carried out scientific data sharing activities. For example, in the United States NASA established distributed active archive centers for earth and environmental data, the White House developed an open data management policy for global change research in 1991, and the federal government has adopted other regulations and policies concerning data management, many of them since 1990. Some international scientific organizations have also strengthened the work of data exchange and sharing. For example, the World Meteorological Organization manages a data exchange system of global meteorological data. There is also the Global Disaster Information Network, and many other global data sharing projects have been established. These examples all provide approaches that China can adopt in carrying out its own scientific data sharing.

The quick growth of China's massive data holdings and the development of information technology have provided support for scientific data sharing. According to the statistics for the past 30 years, rapid developments of science and technology in the world have produced massive scientific data—much more than the data produced during all previous history. China has entered a new phase, with its national information infrastructure and information superhighway developing quickly. Series of networks are being established successively (e.g., the Science and Technology Network, Education and Science Network, Gold Bridge Network, Chinese Public Network, and the new broadband networks). These can all help to ensure that the scientific data are extensively, conveniently, and quickly shared. It is not only an urgent requirement for the scientific and technological innovation system to implement scientific data sharing and to create the new order of opening and distribution for scientific data, but also to promote China's participation in the global economy.

In the process of implementing its scientific data sharing policy, China must consider national strategic requirements and the world's technological evolution. There are five major steps that need to be emphasized as China moves forward on its data sharing strategy.

1. Change and enhance perspectives, and advocate resource sharing. For a long time, the attitude that public scientific data in China are the exclusive private possessions of individuals and departments has made the development of scientific data exchange and sharing difficult. In order to raise the awareness of scientific data sharing and to enhance its effectiveness, China must break the "information exchange barrier"; that is, change the traditional view that the scientific data resources are private property, and foster a new culture of scientific resource sharing.

Chinese researchers need to recognize that public scientific data are national resources that no organization or individual is allowed to keep privately. Public scientific data are derived from observations and experiments that are established through a national investment, as a national resource. This call for open scientific data sharing and reinforcement of the national interest subrogates the individual's interest to the national interest, facilitates the active circulation of scientific data among science departments, and develops its substantial value in the economy, society, and research.

China should encourage organizations to devote their efforts toward scientific data sharing through improved data management systems and mechanisms. MoST has established special funds to facilitate these developments of data sharing. At the same time, MoST also has created and implemented an evaluation policy for scientific data sharing to help ensure that data sharing develops in an orderly manner.

Through the demonstration of data sharing benefits, people gradually will be encouraged to change their traditional views and to make scientific data sharing part of their research activity. Some organizations that have initiated data sharing have provided scientific data conveniently and quickly to many users, which helps to change the data users' and the data managers' thinking and approach. This encourages more organizations and individuals to adopt a data sharing ethic and practice.

2. Plan and develop the scientific data sharing policy and system. In order to ensure the implementation of a scientific data sharing project, it is necessary to develop and integrate a national policy and operational system. "Sharing" requires a management system to regulate and ensure data exchange; to streamline scientific data linkages among data collection, integration, and application; and to promote the relationships between data owners, administrators, and users to ensure data sharing progress.

Scientific data sharing is not limited to one type of industry, department, or person; it needs the cooperation of the state and the whole society. The adoption of new laws can aid the national-level policy and regulation to help normalize the various social relationships associated with data sharing. Under a sound legal system, scientific data resources can be valued fully and serve scientific innovation and national development.

3. Strengthen the establishment of scientific data sharing standards. In order to share scientific data fully, a standard approach must be established. This should include establishing the technological system for the scientific data sharing platform, enacting standards for scientific data sharing and distribution, classifying data into different categories for different users, adopting the data distribution policy, and guiding data integration and communication. During the establishment of such a standard approach, it is necessary to import and use the related standards from other countries, as appropriate. It also is important to emphasize the combination of basic and universal standards, and their application.

4. Construct a service system for national scientific data sharing. It also is a very important task to implement scientific data sharing for existing data resources and for all new scientific data. Scientific data resources that are derived from government scientific and technological programs and distributed among research groups and individuals should be organized and shared efficiently.

Scientific data centers typically need to be organized at the department or organizational level. MoST has chosen some initial departments and organizations, such as resources and environment, medicine and sanitation, and agriculture, and established related data centers and networks. This preliminary work can demonstrate successful experiences and promote the development of scientific data sharing platforms.

5. Strengthen cooperation and improve the sharing of global data resources. International cooperation is essential for data resource sharing. Many important and complex scientific problems cannot be solved by only one country. Researchers need to share scientific data resources globally. China is reforming and opening, learning from other advanced experiences abroad, and encouraging its scientists to show their research results to the world. China also believes that the challenges it addresses can serve as an example for other countries.

MoST has signed bilateral and multilateral treaties on scientific and technological cooperation with more than 100 countries. On the basis of these treaties, the Chinese government is developing activities to encourage its scientists to cooperate with international scientific organizations and to take active part in data exchanges with international data centers. China also welcomes other international scientists to China to establish long-term cooperation in the scientific data and information fields.

The world's scientific community is rapidly entering into a new era dominated by digital technologies. Scientific data sharing, which is a basic strategy for this coming decade, can strongly improve worldwide scientific and technologic abilities with the combined efforts of international organizations, national governments, and individual scientists and engineers. A new knowledge society based on data and information sharing will be created soon.

INTRODUCTION TO THE CHINA SCIENTIFIC DATA SHARING PROJECT

Xian'en Zhang
Director General of the Division of Basic Research
Ministry of Science and Technology of China[1]

Scientific data have become a new resource in the information era and will play a key role in the process of scientific and technological innovation. More reliable, comprehensive, and richer scientific data mean more opportunities for original innovation. The fact that innovations in science and technology are, for the most part, interdisciplinary indicates that such collaboration is the future of science. Successful research depends on open access to data, information, and knowledge from various fields to the greatest extent possible.

The current status for scientific data sharing is far from meeting the demands of China's scientific and technical development, its emerging economy, and growing national power. The scarcity of scientific and technical data resources has become a major obstacle for innovation. The existing, limited resources cannot be used fully because of the outdated research

[1]Based on a presentation available at http://www7.nationalacademies.org/usnc-codata/Zhang_Xianen_Presentation.ppt.

management system. Funding also is not sufficient, and the supporting funding system needs to be improved immediately.

The China Scientific Data Sharing Project (China-SDSP) was launched in 2002 in response to this situation, as well as in response to meeting China's need for sustainable development based on science and technology in the information society. China-SDSP is a part of the National Facility Information Infrastructure for Science and Technology. This presentation provides a brief introduction to China-SDSP activities.

General Considerations and Objectives of China-SDSP

The major objective of China-SDSP is to establish a data sharing architecture that facilitates the use of scientific data by establishing laws, policies, and standards that are supportive of scientific data preservation and sharing and taking advantage of information and communication technology. China-SDSP is being developed through comprehensive planning at the national level; it is collecting and reorganizing data from government agencies, institutes, programs, and individual investigators while making full use of international scientific data resources through cooperation. China-SDSP should make all these data accessible to all interested users at an affordable cost, or free if possible. By utilizing modern information technology, integrating scientific data resources from all kinds of departments, establishing a data sharing policy, and improving management, scientific data resources can be integrated into a uniform framework of national scientific data sharing management. The methodology of the China-SDSP is to build up a data center cluster and a sharing-oriented network, with the goal of forming a multi-faceted, distributed scientific data sharing system that bridges the gaps between different agencies, institutes, and geographical regions. By 2020, the project is expected to achieve the following goals: to form a scientific data management and sharing system that is more user-friendly; to develop a set of supportive laws, policies, and standards; and to establish a core base of data service professionals through a career reward mechanism. By that time, 80 percent of scientific data funded by the government is expected to be made available to the general public.

Framework Architecture of China-SDSP

Logical Framework of the Project

China-SDSP has three major elements: master databases, scientific data centers and networks, and a Gateway Web site.

Master Databases. A major part of the development of the master database component is to reorganize the existing databases. Master databases disseminate authoritative and reliable data to users. The "master database" designation normally corresponds with some specific academic discipline, and contains and describes data that are important for innovations and the advance of science. Approximately 300 such databases have already been identified.

Scientific Data Centers and Networks. As with the development of the master databases, the establishment of scientific data centers and networks is also a result of re-organizing existing data centers and networks. Priority will be given to those centers or networks that have stable data sources and well-planned data archiving in fields such as natural resources and the environment.

Gateway Web Site. The Gateway Web site will link the master databases to all of the other information in every data center. It will deliver services, including content exchange, metadata, and information distribution in order to provide one-stop querying and information dissemination. Figure 1 provides a schematic diagram of the Gateway Web site.

Scope of Data Sharing Supported by China-SDSP

China-SDSP also functions as a catalyst. Its original purpose has been to integrate publicly funded data resources, but its long-term goal is to leverage all possible data resources from government to the private sectors, and make them available to the general public.

The data flow is illustrated in Figure 2. The data may come from three types of sources: large programs funded by the central government, programs funded by other agencies and institutes, and international cooperation programs. Data will be submitted and integrated into an easy-to-use form at following common guidelines and standards. Qualified users will have the right to access these datasets.

Six data sharing systems will be on the priority list: natural resources and

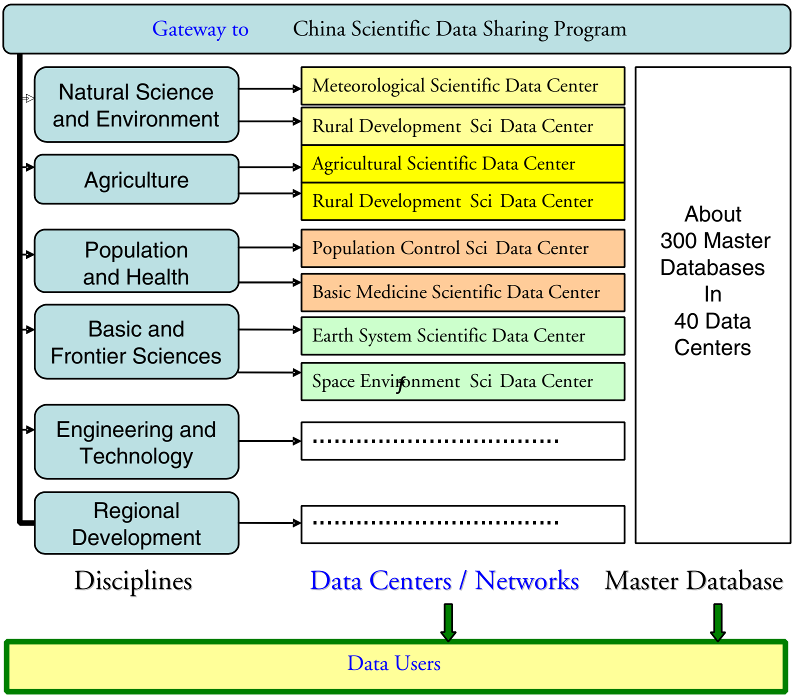

FIGURE 1 Architecture and framework of China Scientific Data Sharing Program.

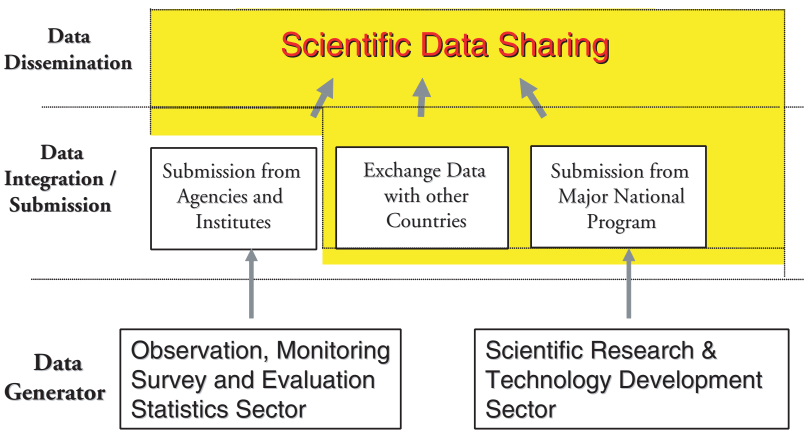

FIGURE 2 Data flow diagram for China-SDSP.

environment management, agriculture, population and health, basic and frontier science, engineering and technology, and regional development.

Service Architecture of China-SDSP

China-SDSP is expected to provide services in the following four ways:

Data management services will integrate up-to-date technologies, such as distributed databases, data warehouses, metadata, and networks, and build up a distributed database system that facilitates data submission, processing, archiving, and updating.

Content services will provide content query based on metadata so that users can obtain information about specific data in a timely and effective manner.

Data services based on a successful content query service will support data browsing and downloading of many kinds of data, either spatial or nonspatial, well-structured or nonstructured.

Extended services will provide tools for various users to search in massive amounts of data, integrate data from different sources, and mining of data to find new knowledge. Data centers may develop more customized services, such as subject-oriented computation.

Major Tasks

Major tasks that need to be addressed include scientific data resource development, standardization, and policy and law aspects.

Scientific Data Resource Development

Data resource development is the process of data collecting, integrating, re-organizing, and mining. The master databases are the natural result of this process. The major tasks are to improve existing data resources; safeguard or rescue endangered scientific data and records; develop the master databases for large research programs funded by the government; introduce international data resources based on their scientific values, quality, and usability; integrate multi-source data; and conduct value-added research.

Standardization

Standardization is a prerequisite for scientific data sharing in the digital era. There are two kinds of standards being addressed by China-SDSP: platform technical standards and data sharing standards. Data standards within major application areas will also be on the list of priorities.

Policy and Law Aspects

The policy and law aspects for scientific data sharing are also key tasks that need to be addressed. Policy making should be considered first as a means for guiding legislation, which is a long and complex process and requires sufficient feedback from sharing practices. Thus both theoretical investigations and case studies are important inputs for a sound data sharing policy. The relevant policy and law system will be developed step-by-step and should handle such issues as guidelines, data classification for sharing, copyrights, incentives, performance evaluation, and so on.

The Chinese data sharing policies and legislation should be improved for several reasons. The existing scientific and technical legislation has not fully represented the sharing spirit and principles. Targets for data sharing are not defined in current policies and regulations. No relevant description for a specific data sharing system could be found in the current policies and regulations. Finally, the existing policies and legislation provide no guidance to research institutions about data sharing, or their proper rights and obligations.

As of this time, over 20 countries have issued relevant policies for scientific data sharing management, thereby forming a primitive global system for scientific data exchange and sharing services in some domains. The policies and laws of these countries provide useful examples.

Tasks that should be addressed in the policy arena include the establishment and implementation of several types of guidelines: the implementation and management guidelines of the China-SDSP; data submission guidelines of major science and technology programs funded by the government; guidelines for scientific and technological data classification for data sharing; and performance evaluation (merit appraisal) of scientific data sharing activities.

Scientific data sharing legislation in China will be important for several reasons. It will provide one of the basic resources for transferring basic scientific and technical results (i.e., data) into real productivity. The implementation of scientific data sharing could produce new types of information business, such as data product processing and data distribution services. Scientific data sharing could also provide open scientific data resources and a fair competitive environment for data resources in the information industry. For the purpose of fully participating in worldwide high-level collaboration and competition, China should improve its national scientific and technical competitive ability and lay a foundation for efficiently supporting its emerging economy.

Data sharing legislation and regulations therefore need to define property rights in data; encourage data sharing; and establish a system for supervising data sharing, for the technical decisions about data sharing, and for data sharing evaluation. There are several legislative and regulatory actions that should be taken in constructing a domestic legal system for scientific data sharing.

Revise the Scientific and Technical Improvement Law. As the basic national law of science and technology and the standard for sharing scientific and technical resources, this law should represent the most important "sharing principle" for platform construction. It should make sure the data resources are open and shared with the public by data managers and public research managers, give rights to relevant bodies that need access to data resources, as well as define the characteristics of national scientific and technical data resources and regulate public scientific and technical data resource sharing on proper terms.

Draft a Scientific and Technical Resource Sharing and Protection Law. This law should define basic obligations for data resource managers and owners regarding open and shared data resources with the public, regulate basic qualifications and procedures for users, and stipulate the requirements for providing service and education. Such measures are needed to prevent improper or illegal activities and to provide the national scientific and technical resource management center with adequate funding, information, and timely data.

Regulation on Scientific Data Sharing. This regulation should define scientific data, clarify the obligations for scientific data owners and managers to open and share their resources, and establish a standard management system for scientific data resource and standard data formats. The persons in charge should be held accountable for the waste of resources and other types of mismanagement.

Regulation on Opening Government Information. This national-level regulation will clarify the principle for government to freely issue its information, similar to the resource sharing in the scientific and technical fundamental platform. The resource sharing on the platform could be greatly improved with the announcement and implementation of this major regulation.

There may be other relevant laws and regulations that need to be considered or amended as well.

Summary of the China-SDSP Work Plan

The following activities are being implemented during the initial experimental period, 2001-2005: overall planning and design; research and planning on the policy and legal framework; drafting and issuing relevant policies and regulations; developing technology and standards; establishing scientific data centers and networks and initiating a data sharing pilot project; identifying the optimal mechanisms for existing data consolidation and sharing; launching of the program Gateway Web site, selecting 25 data centers for the data sharing pilot project, and selecting other candidate centers for further development; summing up experiences from various aspects of the experimental period; and preparing a feasibility report about the overall implementation of public data sharing in the next period of activity.

In 2001, the meteorological data sharing project was launched,[2] which heralded the start of the scientific data sharing program in China.

The China-SDSP was formally initiated in 2002. By the end of 2002, another five data centers (land survey, hydrology and water resources, seismic, forestry, and agriculture) and three networks (earth system science data center network, modern agricultural technology and rural development network, and sustainable development network) had joined the pilot project.

In June 2003, a Coordinating Group and a Scientific Group were established for scientific data sharing. The main task of these groups was to develop the Plan for the China-SDSP by May 2004 with the following six major components: current status and major national requirements; overall

[2]See the summary of the presentation on "Progress in Meteorological Data Sharing in China" by Dahe Qin in Chapter 4.

considerations; principles and objectives; strategic framework and tasks; implementation and measurements; and supporting conditions and facilities.

As of June 2004, the Gateway Web site has become pivotal to the China-SDSP. Its function and technical specifications have been clarified and the overall design and specific modules are finished. Data from the five data centers and three networks are being integrated into this Gateway and will be available by the end of 2004.

In terms of policy making, a working group for data sharing was established and investigated the current status and trend of data policy, both at home and abroad, compiled relevant materials and information, established the "Guidelines of Data Submission from Major National Programs," began researching the framework of relevant law and policy, and finished the conceptual design for data classification for sharing.

A research group for data standards also has been established. It has investigated the current status and trend of data standards both at home and abroad, compiled relevant standards, and drafted the framework and guidelines for data sharing management. In general, the China-SDSP is still in the overall planning phase, accumulating the experiences of technology and policy making, as well as overseeing pilot data sharing projects.

During the implementation period, 2006 to 2010, the following work will need to be done: continue the establishment of data sharing technology, policy, and law; extend the program coverage of scientific data centers or networks and make them operational; gradually improve technology and standards; enforce the cooperation among data centers in different research areas; and enhance the capacity to develop high-level data products and to ensure quality. After each yearly performance evaluation of the 25 pilot data centers or networks, the qualified ones will be included in the "National Scientific Data Master Network" and will start regular operation. The amount invested in each center will depend on its merits and performance. Another 15 to 20 data centers will be launched, including 200 new master databases. By 2010, a mechanism is expected to be established through which data are submitted from various governmental agencies and programs and delivered to potential users efficiently.

In conclusion, it is important to emphasize that science respects no border. The establishment and implementation of the China-SDSP needs the support and assistance of the international community, and in turn will contribute to the development of global science and technology.

3

Summaries of Presentations on International Perspectives

The perspectives of five international organizations—the United Nations Educational, Scientific and Cultural Organization (UNESCO), the International Council for Science (ICSU), the Organisation for Economic Co-operation and Development (OECD), the Inter-Academy Panel on International Issues (IAP), and the Committee on Data for Science and Technology (CODATA)—were presented at the workshop. These presentations, summarized below, were focused on activities that are either under way or recently completed by these organizations in the areas relating to the preservation of and open access to public scientific data and information. The chapter concludes with brief overviews of the Chinese and U.S. National Committees for CODATA.

TRENDS IN DEVELOPMENT OF INTERNATIONAL SCIENTIFIC DATA AND INFORMATION[1]

The commitment of UNESCO[2] to the essential good of the free flow of information and access to knowledge sources is inspired by its Constitution, which states that "the wide diffusion of culture, and the education of

[1]Based on a presentation by Yasuyuki Aoshima, United Nations Educational, Scientific and Cultural Organization, available at http://www7.nationalacademies.org/usnc-codata/Aoshima_Presentation.ppt.

[2]For more information about UNESCO's many activities see http://www.unesco.org.

humanity for justice and liberty and peace are indispensable to the dignity of man and constitute a sacred duty which all the nations must fulfill in a spirit of mutual assistance and concern."

For several years, various Resolutions of the UNESCO General Conference and the Executive Board have urged member states and associate members to promote free and universal access to public domain information for the purpose of education, science, and culture. As a result, UNESCO has taken several proactive measures in order to encourage member states to establish a right of universal access to information and formulate policies and regulatory frameworks, which would determine the future orientations of the information society.

After an extensive round of negotiations, the General Conference of UNESCO adopted at its 32nd session in October 2003 the Recommendation concerning the Promotion and Use of Multilingualism and Universal Access to Cyberspace. The Recommendation recognizes the importance of promoting equitable access to information and knowledge, especially in the public domain, and reiterates UNESCO's conviction that the organization should have a leading role in encouraging access to information for all, supporting multilingualism and cultural diversity on the global information networks. The Recommendation proposes a number of specific measures fostering universal access to digital resources and services, and facilitating the preservation of their cultural and language diversity. UNESCO thereby encourages its member states to support equitable and affordable access to information and to promote the development of a multicultural information society.

The current evolution towards knowledge societies through an increased use of information and communication technologies (ICTs) facilitates the movement and handling of data as well as the process of generating and validating information and knowledge. Thus, the various aspects of access to scientific data and information in the digital world, including the very important questions concerning intellectual property rights, will continue to attract attention.

Public-domain information, which is free of copyright and other intellectual property rights, often is not sufficiently well known to potential contributors and users. In some countries, there are growing restrictions on the availability and use of public data and information. The public-domain principle also can be extended conceptually by the adoption of "open access" to information, which is made freely and openly available by its rights holders. One example of open access is the open-source software license, by which computer programs are distributed free of charge by their authors for exploitation and further development. Another is the documentation produced and made available free of charge by the United Nations and its specialized agencies, as well as public data and official information produced and voluntarily made available by governments.

A significant aspect of the Recommendation concerning the Promotion and Use of Multilingualism and Universal Access to Cyberspace is the promotion of the social and cultural dimensions of universal access and the equitable balance between the interests of rights holders and the public interest. It is a major challenge for all those involved in the expansion of the emerging information society to contribute to maintaining a balance between copyright protection and access to information in the public domain. International copyright systems have been changed over time to adjust to the new digitally networked environment. In this sense, adopting coherent national legislative measures and ensuring international harmonization among countries is essential to avoid potential conflicts of interests in an increasingly globalized world. Through the adopted Recommendation UNESCO recognizes the importance of an equitable balance between the interests of rightsholders and those of users when valuable works and performances are exploited in the digital environment in the fields of education, science, and culture.

As one specific follow-up to the Recommendation, UNESCO published in 2004 policy guidelines regarding governmental public-domain information[3] in order to advise countries on policies for the development and promotion of such information, taking account of both national needs and international practices. These guidelines serve as an advisory tool for governments. They emphasize the importance of governmental public-domain information, which contributes to economic and social development, promotes democratic ideals through greater transparency of governance, enhances public health and safety, and serves essential scientific and technical research functions. The guidelines also encourage governments to define the scope of available governmental public-domain information according to national needs, adopt a "Freedom of Information" law, and

[3]See Policy Guidelines for the Development and Promotion of Governmental Public Domain Information prepared by Paul F. Uhlir. 2004. Paris: UNESCO, available at: http://portal.unesco.org/ci/en/ev.php-URL_ID=15863&URL_DO=DO_TOPIC&URL_SECTION=201.html.

develop and implement a comprehensive Governmental Public Information Policy Framework.

Freedom of expression and freedom of opinion, as enshrined in Article 19 of the *Universal Declaration of Human Rights*,[4] are essential premises of the information society. In building such a society, the ability for all to access and produce information, and generate ideas and knowledge is indispensable.

Scientific research leads to the development of new technologies and to the production of data and information that can benefit society as a whole.

While some countries are leading or keeping up with scientific and technical progress and with the digital information revolution, developing countries have to make huge efforts to gain access to the necessary infrastructure and to take full advantage of ICTs. It is indeed paradoxical that if ICTs facilitate communication from the global to the local levels and vice versa, they can also broaden the digital gap between those who participate in the information society and benefit from it and those who cannot.

The impact of ICTs in the production, use, and dissemination of scientific knowledge is immense. There are many opportunities to bridge the science gap, for example by improving networking among scientists locally and internationally, and by providing scientific information and knowledge to decision makers for better governance. Moreover, ICTs are central to scientific research itself. They enable scientists to perform fundamental and applied research, build partnerships and scientific international consortia, conduct experiments, manage data, coordinate laboratory activities, and communicate their findings to their colleagues and the public.

Based on the assumption that scientific knowledge produced through public investments is a "public good," scientific data and information should be as widely available and affordable as possible, since the more people that are able to share such information, the greater the positive effects and returns will be for society. The importance of scientific data to society therefore should be described clearly since all areas of research now require availability of high-quality data for their progress.

Finally, in light of all these developments, the sustainable, long-term preservation of digital data and information is a very important concern. For several years now, specialized agencies and organizations of the United

[4]See http://www.un.org/Overview/rights.html.

Nations have highlighted the fact that despite the growing efforts of the various stakeholders involved in generating, organizing, and providing access to digital scientific data and information, these resources are still at risk of being lost to future generations. The issue of digital archiving is thus essentially a matter of scientific and public policy, which should be of concern to decision makers.

TOWARDS INTERNATIONAL GUIDELINES FOR ACCESS TO RESEARCH DATA FROM PUBLIC FUNDING[5]

The ever increasing use of information communication technologies is bringing dramatic changes to the way our global science system operates. Digitization has become an essential part of the scientific process and the management of research. Yesterday's scientists studied nature. Today's scientists study digital data—digital data on nature to be sure. Sir Isaac Newton did not need more than a pencil and a pad to process his observational data into ground-breaking scientific laws. But today the next advance in high-energy physics requires a Large Hadron Collider that will produce 12 to 14 petabytes of digital data per year, the full capacity of about 16 million CD ROMs. These data will be analyzed by some 6,000 researchers scattered around the world, but tightly knit by the Grid computer network of our global science system.

The use of ICTs has made collections of scientific data in many respects comparable to musical scores: to be used time and again for a diversity of performances by a diversity of artists for different audiences of society. Optimum access to research data should enable researchers from all over the world to compose the full score for our knowledge-based international society. Consequently, access to the gold mine of research data quickly has become a major issue in international science policy and research management. The traditional exchange arrangements between scientific colleagues no longer suffice to guarantee the necessary openness of access to digital data resources. Optimum access requires formal agreements on the conditions of access at both the national and international levels. The main task of establishing an adequate regulatory framework lies within the re-

[5]Based on a presentation by Peter Schröder, Ministry of Education, Culture, and Science, The Netherlands, available at http://www7.nationalacademies.org/usnc-codata/Schroder_Presentation.pdf.

search community—at the national research councils, scientific institutes, universities, and funding agencies. The general principles to build data access regimes should be a responsibility of governments, however. Considering the international dimensions of the scientific effort in general and of access to data in particular, national data access policy regimes will only work when closely connected to international agreements.

At the ministerial-level meeting of the OECD Committee for Scientific and Technological Policy on January 30, 2004, the ministers responsible for science policy endorsed the "Declaration on Access to Research Data from Public Funding,"[6] including a draft set of principles and the recommendation that OECD should develop these principles into more detailed guidelines. The Declaration is an important step towards further international scientific cooperation.

The guidelines will address the technical, institutional, financial, legal, and cultural aspects of data access regimes. The basic premise of the guidelines will be open access to and unrestricted use of publicly funded research data, subject only to legitimate restrictions. At least nine areas will be addressed in the guidelines: transparency, formal responsibility, professionalism, protection of intellectual property, legal conformity, interoperability, quality, efficiency, and accountability.

In short, new governmental science policies on data access and new data policies of research funding organizations will influence the future course of international research practices. The development of international approaches to data access holds the promise of opening up exciting new dimensions of international scientific cooperation.

INTERNATIONAL PERSPECTIVES ON DATA AND INFORMATION FOR SCIENCE[7]

The mission of ICSU is to strengthen international science for the benefit of society. One of the key principles that underpins this mission is the "Universality of Science"; that is, all scientists should have the possibility to participate, without discrimination and on an equitable basis, in legitimate scientific activities whether they be conducted in a national, trans-

[6]See http://www.codataweb.org/UNESCOmtg/dryden-declaration.pdf.

[7]Based on a presentation by Carthage Smith, International Council for Science, available at http://www7.nationalacademies.org/usnc-codata/CarthageSmithPresentation.ppt.

national, or international context. Scientific data and information are the input and product of scientific research, and the practices and policies that dictate their use must reinforce the universality of science. In so doing they will improve science for the benefit of everyone.

Since its inception, ICSU, whose membership includes both international disciplinary science Unions and national interdisciplinary science bodies, has been involved in scientific data and information issues. A particular focus has been on the international and interdisciplinary issues relating to data production, management, and access. When ICSU was established in 1932, the challenges were very different than they are today. While data exchange was logistically more difficult and much slower in 1932, it is paradoxical that the incredible advances in information and communication technologies that have taken place in the last decade have also, in many ways, made the exchange of data much more complex.

In order to address the key management and policy challenges, ICSU has established a number of specific interdisciplinary bodies for data and information, including CODATA, a co-organizer of this workshop, and the International Network for the Availability of Scientific Publications, also represented at the meeting. Moreover, ICSU has developed the major international research programs on global environmental change and is currently planning large new scientific initiatives, such as the International Polar Year (2007-2008), which produce, collect, analyze, and disseminate large amounts of diverse scientific data from and for many sources around the world.

With regard to data policy, ICSU is a strong advocate of "full and open access"[8] to scientific data and "universal and equitable access"[9] to scientific publications. ICSU also has been very actively involved with CODATA, UNESCO, and other international science partners in the two phases of the United Nations World Summit on the Information Society (WSIS). On behalf of the international science community, these groups argued

---

[8]ICSU defines "full and open access" to data as equitable, nondiscriminatory access to all data that are of value to science. It does not necessarily equate to "free of cost" at the point of delivery, although this is certainly the ideal in many situations, particularly with regard to publicly funded data that are made available online.

[9]ICSU defines "universal and equitable access" to scientific publications as ensuring equal opportunities both to publish and to obtain scientific information for all scientists wherever they are located. It does not necessarily imply without any cost at the point of delivery.

very strongly for recognition of the critical role and needs of science in the information society. In particular, an agenda for action—Science in the Information Society[10]—was developed by ICSU and its partners. It highlighted the key issues relating to preservation of and access to scientific data and information. This agenda was very influential because its main recommendations were incorporated into the formal documents that were agreed by Heads of State at the end of phase I of WSIS, which was held in Geneva in December 2003.

In light of the rapidly changing international scientific (and political) landscape and in parallel to WSIS, ICSU recently completed a Priority Area Assessment of the international needs for scientific data and information[11]—production, management, access, and dissemination. The report addresses many of the future challenges related to data preservation and access and identifies clear actions for ICSU, the scientific community, policy makers, funders, and other stakeholders. Many of the existing practices, mechanisms, structures, and policies relating to scientific data and information need to be updated and improved if the optimum benefit from data and information is to be obtained for both science and society.

The Priority Area Assessment report also identifies a number of priorities and needs in the area of scientific data and information, including long-term strategic planning and investment, professional data management, and modernization of current infrastructure and systems, as well as new infrastructure in some areas. It also underscores the need for international cooperation and supportive data and information policies at the national and international levels.

In the area of data preservation, the report recommends that there must be a prioritization of what data to preserve. Other areas of importance are data rescue and recovery and the need for "openly available" metadata. The report also identifies long-term support for archiving and considerations regarding data integrity and data privacy as essential requirements for effective data preservation.

---

[10]See http://www.icsu.org for the full agenda for action and details of other ICSU publications and activities relating to the WSIS.

[11]International Council for Science. 2004. *ICSU Report of the CSPR Assessment Panel on Scientific Data and Information.* The background description, terms of reference, final report, and panel membership for the ICSU Priority Area Assessment on Scientific Data and Information can be obtained from http://www.icsu.org/1_icsuinscience/DATA_Paa_1.html/.

The ICSU Priority Area Assessment defines access in two ways: (1) to provide data and information and publish them, and (2) to be able to use and read the data and information. Minimal constraints on access should be maintained. The report recommends the development of stable systems for providing universal access to quality data and information. New economic models for providing access to scientific, technical, and medical literature need to be developed. Scientists also should become more involved in the development of policies, such as intellectual property rights, that affect access. Finally, the interests and needs of scientists in developing countries must be considered and addressed.

INTER-ACADEMY PANEL INITIATIVES ON PROMOTING ACCESS TO SCIENTIFIC INFORMATION[12]

The IAP is a global network of over 90 Academies of Science designed to promote their greater participation in science policy discussions as well as policies that strengthen scientific institutions. Toward that end, the IAP creates partnerships among its member institutions and works closely with other scientific organizations. The IAP is governed by an Executive Committee with a rotating membership of 11 member Academies and two co-chairs. In 2003-2006, the co-chairs are Professor Zhu Chen of the Chinese Academy of Sciences and Professor Yves Quéré of the French Académie des Sciences. The Third World Academy of Sciences and ICSU are also ex officio members.

Since its inception, the IAP has issued statements on population growth (1994), urban development (1996), sustainability (2000), human reproductive cloning (2003), science education (2003), health of mothers and children (2003), scientific capacity building (2003), science and the media (2003), and access to scientific information (2003).[13] The IAP also organizes its activities according to major programs and shorter, more focused initiatives. Currently, its programs are in the areas of capacity building for Academies, science education, health education for women, and water. It also has three initiatives focused on biosecurity, genetically modified organisms, and access to scientific information.

[12]Based on a presentation by Michael Clegg, foreign secretary of the U.S. National Academy of Sciences, available at http://www7.nationalacademies.org/usnc-codata/MichaelCleggPresentation.ppt.

[13]For more information on the Inter-Academy Panel, see http://www4.nationalacademies.org/iap/iapGA.nsf/.

Because of the importance of the issue of improving access to scientific data and information as a matter of science policy at the national and international levels, the IAP issued a policy statement on this topic in December 2003.[14] The IAP Statement on Access to Scientific Information focuses on access to numerical scientific data, databases, and scientific literature. It also attempts to address the high cost of scientific journal subscriptions. The statement recommends that:

➢ Electronic access to journal content be made available worldwide without cost as soon as possible, within one year or less of publication for scientists in industrialized nations, and immediately upon publication for scientists in developing countries;

➢ Journal content and, to the extent possible, data upon which research is based be prepared and presented in a standard format for electronic distribution to facilitate ease of use;

➢ Journal content be archived collectively, either by private or government organizations;

➢ Governments and publishers work together to raise awareness in the scientific community of the availability of free electronic access to scientific journals; and

➢ Scientific databases obtained by intergovernmental organizations be made available without cost or restrictions on reuse.

As noted above, the IAP also has launched a new initiative under the leadership of the U.S., Chinese, and Senegalese Academies of Science on access to scientific information in developing countries. The IAP will convene a meeting of interested members to define this new initiative.[15] In addition, it may also develop positions on subcategories of the broad topic of information access including electronic access, access to databases, responsibilities of the publishing industry, public-sector responsibilities, and the information communication technologies infrastructure supporting access to data and information.

---

[14]See the Inter-Academy Panel on International Issues. 2003. An IAP Statement on Access to Scientific Information, Mexico City, December 4. http://www4.nationalacademies.org/iap/iaphome.nsf/weblinks/WWWW-5U6HHG?OpenDocument.

[15]Since this workshop, the IAP convened two related meetings in Paris in November 2004 and in Dakar, Senegal, in January 2006.

FUTURE ROLE OF THE COMMITTEE ON DATA FOR SCIENCE AND TECHNOLOGY[16]

CODATA[17] is an interdisciplinary committee of ICSU concerned with various types of quantitative data resulting from experimental measurements or observations in the natural and social sciences, and the engineering disciplines. Particular emphasis is given to data management problems common to different scientific fields and to data sharing among these disciplines. CODATA's objectives focus on:

- ➢ Improving the quality and accessibility of data, as well as the methods by which data are acquired, managed, and analyzed;
- ➢ Facilitating international cooperation among those collecting, organizing, and using data;
- ➢ Promoting an increased awareness in the scientific and technical community of the importance of these activities; and
- ➢ Considering data access and intellectual property issues.

In addition to traditional activities such as meetings, workshops, and publishing, CODATA is now working to enhance such international activities by (1) promoting project-oriented approaches to highlight models to follow up on fruitful ideas; (2) articulating issues in global access to scientific and technical data through intensive commitments to global and societal activities, such as the WSIS; and (3) expanding human dimensions to enhance data flows beyond borders, disciplines, organizations, and generations.

CODATA has identified data preservation and data access as priority mandates. There is a diverse portfolio of activities focused on the preservation of scientific and technical data, primarily through the work of its Task Group on Preservation of and Access to Scientific and Technical Data in Developing Countries. The Task Group has been working to identify the scientific, technical, management, and policy issues related to the preservation of scientific and technical data, and the Task Group members have compiled a comprehensive, annotated bibliography of archiving resources.

---

[16]Based on a presentation by Shuichi Iwata, University of Tokyo, Japan, and President, CODATA, available at http://www7.nationalacademies.org/usnc-codata/Iwata_6_Presentation.ppt.

[17]For more information, see http://www.codata.org/.

They also have initiated a series of international, interdisciplinary workshops, including this workshop, to highlight the importance of permanent access to scientific information resources, and to examine the policy and legal, management and technical, and institutional and economic issues that are important to providing permanent access to digital scientific data. Finally, CODATA, along with the Task Group, has worked with the International Council for Scientific and Technical Information (ICSTI) to create an online portal of resources related to the Permanent Access of Scientific Data and Information.[18]

In the area of access to scientific data, CODATA has worked closely with ICSU as well to provide an international voice in support of "full and open" access to scientific and technical data as new legislative and treaty regimes have been considered and implemented. Between 1997 and 2003 the joint ad hoc ICSU-CODATA Working Group on Data and Information monitored the implementation of the European Union Database Directive in Europe, and participated as an official "Observer" in the discussions of a potential new database treaty at meetings of the World Intellectual Property Organization (WIPO). The ICSU-CODATA White Paper on Data Access (1997)[19] was a major defining document on this subject from the scientific community perspective at the WIPO. A workshop involving representatives from the U.S. and European Academies of Sciences and data law experts was held in conjunction with the 2000 CODATA Conference in Baveno and Stresa, Italy. Several sessions on these issues were convened during the 2002 CODATA Conference in Montreal as well.

Finally, CODATA, in collaboration with the U.S. National Committee for CODATA, ICSTI, ICSU, and UNESCO, organized a major international symposium on open access and the public domain in digital data and information for science in March 2003. This activity also helped to identify and analyze important issues for follow-up by the ICSU family of organizations and for the development of an Action Plan in this area by ICSU and UNESCO in preparation for the WSIS. The first phase of WSIS was held in Geneva in December 2003, and the second was held in Tunis in November 2005. CODATA discussed WSIS at its General Assembly in

[18]See http://www.nap.edu/shelves/data for the CODATA/ICSTI Portal on Permanent Access to Scientific Data and Information.

[19]See *Access to Databases: Principles for Science in the Internet Era* at http://www.codata.org/codata/data_access/principles.html.

Berlin in 2004. CODATA will continue to highlight the role of science in the information society in preparations for the Tunis phase of the Summit. The resulting documents from this process are expected to create solutions on "digital divide issues" based on the individual care principle with respect to the institutional, legal, ethical, emotional, and cognitive aspects of data.

CHINA'S NATIONAL COMMITTEE FOR CODATA[20]

China joined CODATA in 1984, with the Chinese Academy of Sciences serving as the national member organization. Since then, the Chinese National Committee for CODATA[21] has made efforts to promote China's scientific and technical data activities through its executive body—the Secretariat located at the Computer Network Information Center of the Chinese Academy of Sciences.

The mission of the Chinese National Committee for CODATA is to facilitate construction of scientific databases in China and to promote interdisciplinary and international exchanges, cooperation, and data sharing. The committee's main activities are to:

1. Investigate the domestic data activities and trends for establishing scientific and technical databases;
2. Convene an annual general assembly of the Chinese National Committee for CODATA, with the purpose of discussing its activities, providing an overview of the on-going tasks of its data groups, and updating the community on the academic achievements of CODATA;
3. Promote domestic scientific and technical information resources to make more efficient use of scientific and technical data;
4. Organize domestic working groups and relevant meetings within different academic spheres; and
5. Initiate and organize the implementation of programs on scientific database construction and data sharing.

[20]Based on a presentation by Zhihong Xu, representative of the Chinese National Committee for CODATA, China, available at http://www7.nationalacademies.org/usnc-codata/XuZhihong_Presentation.ppt.

[21]See http://www.codata.cn for more information.

U.S. NATIONAL COMMITTEE FOR CODATA[22]

The U.S. National Committee (USNC) for CODATA[23] provides a bridge between the scientific and technical community in the United States and the international CODATA on data issues. The USNC operates within the National Research Council's Board on International Scientific Organizations and is the principal organizational entity within that Board's Office of International Scientific and Technical Information Programs.

The USNC undertakes special studies and activities, counsels the U.S. National Delegate to CODATA on the U.S. position regarding official CODATA business, and provides a link between U.S. and international data compilation and evaluation activities, taking into account the needs of the scientific and technical user community.

Finally, the USNC works closely with the international CODATA on permanent access to scientific and technical data. It is a co-organizer of the aforementioned series of workshops relating to permanent access to scientific and technical data, including this one.

---

[22]Based on a presentation by Paul F. Uhlir, director of the U.S. National Committee for CODATA at the U.S. National Academies.

[23]See http://www7.nationalacademies.org/usnc-codata/.

4

Summaries of Presentations on Cross-Disciplinary Issues

Three parallel breakout panel discussions focusing on cross-disciplinary issues in open access to and preservation of scientific data and information from the viewpoint of China were convened during the course of the workshop. These sessions were organized according to (1) legal and policy, (2) institutional and economic, and (3) management and technical issues. The objective of these thematic breakout discussions was to examine different possible models in these areas and their potential benefits and shortcomings in China.

PANEL DISCUSSION ON LEGAL AND POLICY ISSUES

Introduction[1]

With regard to scientific data resources, most databases and data centers in China are managed directly or funded by government ministries and are subject to a relatively restrictive state information regime based on official secrecy requirements. This is a major challenge to the adoption of an open-access model because the past policies have been based on deeply rooted political, institutional, and cultural factors. Some of the restrictions have applied generally to the overall public information regime, while oth-

[1]Paul F. Uhlir and Julie M. Esanu, U.S. National Academies.

ers have been more specific to science and based on perceived political or economic sensitivities (e.g., domestic disease statistics or high-resolution geospatial data). The Chinese government, however, increasingly recognizes that many types of scientific data should be made openly available and usable, especially within the country, and not just for research purposes. As discussed in Chapter 2, the recent high-level focus by the Chinese government on the laws and policies regarding access to government-produced and government-funded academic research data has made this a very propitious time to examine these issues.

The case for change in access policies to governmental scientific data can be made at many levels, both internally and externally. The most effective approach is one based on the realization of national self-interest. A comparison with the policies of other countries can be effective as well. Particularly auspicious is the trend over the past decade by many developing countries to adopt Freedom of Information laws.[2]

Of course, there are legitimate public-policy reasons for limiting access to certain types of data, including appropriate national security restrictions, the protection of privacy and confidentiality, and the protection of private (as opposed to government) intellectual property rights. A related and very significant problem exists in getting scientists to contribute the data produced in the course of their research to public repositories. Barriers include the lack of an appropriate data center in which to deposit the data, no requirement by the funding source to deposit the data or to share them openly, insufficient recognition of the importance of data activities by the scientist's institution, a lack of effective incentives or rewards to make the data available, the desire of researchers to sell their data at unreasonable prices despite very weak market estimates, inadequate funding to prepare the data sufficiently to make them usable by others, and a lack of training to do so.

With regard to scientific, technical, and medical journals, these too are mostly published by government or government-sponsored organizations in China. Because they are meant to be read by the research community, they do not have many of the same official constraints based on national security considerations as the underlying data. They are, however, still published almost exclusively in print form, so their open availability on digital networks raises new policy issues for the Chinese journal publishers and research establishment.

[2]For more information, see, for example, http://www.freedominfo.org.

The presentations summarized in this section focus on some of the legal and policy barriers to more open access for publicly funded scientific data at both the national and international levels, outline some of the policy arguments in favor of greater unrestricted access, and offer some policy guidelines in support of open availability.

Global Trends to Restrict Access to Data from Government-Funded Research[3]

Scientific data produced from government-funded research constitute a fundamental element of the modern research infrastructure and, if well managed, can greatly accelerate scientific progress at the national and international levels. Newly emerging possibilities for enhancing this role of scientific data resources in the digital environment truly constitute another "endless frontier." High-level policy attention is necessary at the national and international levels in order to maximize the inherent value of data collections and to minimize the negative effects of restrictions on access and use.

Indeed, many economic, legal, and technological restrictions have been placed on public-domain scientific data throughout the world. From an economic perspective, the trends to privatize governmental public-good functions and to commercialize more of the academic sector's research activities have been under way over the past two decades, particularly in biomedical and engineering areas. While these trends can support significant research advances and economic benefits, they are not without their own economic and social costs. A further continuation of privatization and commercialization of upstream public-sector information resources can be viewed as potentially having greater associated costs than benefits.

Recent changes to international and national intellectual property laws, such as new digital copyright protection and the adoption of exclusive property rights protection for noncopyrightable databases in many countries, as well as the adoption of licensing agreements on onerous terms for research tools—including data—in academia are further diminishing the broad availability of public-domain data in science. Moreover, these highly protectionistic legal mechanisms are increasingly enforced by more effective digital rights management technologies. Such developments are intensifying the tensions that already exist between the policies that favor shar-

[3]Based on a presentation by Jerome Reichman, Duke University School of Law.

ing of scientific data and the perceived need to restrict access to and uses of data in pursuit of increased commercial opportunities. Restrictions on the dissemination of potentially sensitive research data and information based on national security considerations are further constraining the availability of substantial amounts of material in the public domain. Finally, the recent enactment of a powerful new database protection statute in Europe and proposals for equivalent legislation in the United States and in other countries might be expected to push these tensions into other areas of public research, which up to now have been less affected by the proprietary pressures from the commercialization and privatization trends.

A Contractually Reconstructed Research Commons for Scientific Data in a Highly Protectionist Intellectual Property Environment[4]

If the economic, legal, and technological pressures on public-domain scientific data that were identified in the previous section continue unabated, they will result in lost opportunity costs across the entire research enterprise. These pressures, which are especially pronounced in biomedical and engineering research, could elicit one of two types of responses. One is essentially reactive, in which the public research community continues to adjust as best it can on an ad hoc basis, without organizing a response to the increasing encroachment of a commercial and proprietary ethos on data produced by government-funded research. The other would require a science policy response to the challenge by formulating a strategy that would enable the scientific community to take more active control of its basic data supply. The idea is to reinforce, by voluntary means, a public space in which the data sharing ethic in public science can be promoted and insulated from some of the excessive privatization and commercialization trends, without impeding socially beneficial commercial opportunities. There are some contractual approaches that are now being considered in the United States and Europe, which the Chinese science policy community might consider as well in addressing this challenge in biomedical and other types of publicly funded research.[5]

---

[4]Based on a presentation by Jerome Reichman, Duke University School of Law.

[5]These approaches are examined in detail in an article by J.H. Reichman and Paul F. Uhlir. 2003. "A Contractually Reconstructed Research Commons for Scientific Data in a Highly Protectionist Intellectual Property Environment," *Law and Contemporary Problems,* Duke University School of Law, vol. 66, Winter/Spring.

Balancing the General Public Interests and Copyright in Scientific Information Management[6]

In both the policy and legal arenas, there is a rising sense of research and scholarship falling within the more general public's right to know and supporting initiatives for increasing and opening access to research. There are two policy aspects to consider, as government policies can determine how scientific knowledge circulates and as policies are affected by the research that is consulted in their formation. Shifts are taking place in policies affecting science, and these shifts are motivated by the basic human right to know as recognized, for example, by the United Nations Universal Declaration of Human Rights. They also are motivated by greater demands for accountability and transparency in the public administration of funding in areas such as government research grants. In Canada, for example, one of the principal granting councils for the social sciences and humanities is transforming itself into a "knowledge council," which gives a high priority to the public impact and awareness of research. According to a University of British Columbia study of policy makers' actual use of research,[7] online access is having a substantial impact and is increasing the amount of research consulted, even as the policy makers are largely restricted to "open access" or free materials due to budgetary restrictions and the limited number of subscriptions held. Online access also has expanded policy makers' circle of consultation, as they are relying less on a small set of academics to advise them. The role of research in policy making is an issue raised in many countries and in many contexts. Greater public access to the research literature and to the underlying data sources would help support more informed and rational policy making.

In terms of legal issues, two pertinent areas of law are Freedom of Information legislation and copyright. In the United States, for example, recent legislation has brought federally funded scientific data produced in universities that are used to support the formation of federal government regulations within the purview of the Freedom of Information Act (FOIA). Previously, the FOIA applied only to data produced within the federal government itself. Charges also have recently been made in New York against a

---

[6]Based on a presentation by John Willinsky, University of British Columbia, Canada.

[7]See Willinsky, J. 2003. "Policymakers' online use of academic research," *Education Policy Analysis Archives*, *11*(2), January 11. Retrieved January 28, 2005, from http://epaa.asu.edu/epaa/v11n2/.

major drug company for suppressing research unfavorable to its medication. These examples are indicative of rising public expectations of science and that people have a right to know what is known.

Copyright protection in scholarly publishing is occasionally portrayed as a matter of protecting authors from plagiarism. Open access reduces considerably the likelihood of getting away with plagiarism. The more substantial legal issue, however, concerns the basic principle of copyright, namely, to protect the interests of the author and the public. Here a new argument can be introduced in favor of open-access scholarly publishing, serving the interests of the author and the public better than publishing models that depend on subscriptions and copyright control, which actually reduce both the author's and the public's rights in scholarly publication and communication. In short, increasing access to research has much to contribute to the policy and legal considerations in scholarly publishing.

Borders in Cyberspace: Maximizing Social and Economic Benefit from Public Investment in Data[8]

Many nations are now embracing the concept of open and unrestricted access to public-sector information—particularly scientific, environmental, and statistical information of great public benefit. Federal information policy in the United States is based on the premise that government information is a valuable national resource and that the economic benefits to society are maximized when taxpayer-funded information is made available inexpensively and as widely as possible. This policy is expressed in the Paperwork Reduction Act of 1995[9] and in Office of Management and Budget Circular No. A-130, "Management of Federal Information Resources."[10] The policy actively encourages the development of a robust private sector, improved access to critical information in the academic and research sector, and offers to provide publishers with the raw content from which new information services may be created, at no more than the cost of dissemination and without copyright or other restrictions.

---

[8]Based on a presentation by Peter Weiss, J.D., U.S. National Weather Service. See also Weiss. 1997. "International Information Policy in Conflict: Open and Unrestricted Access versus Government Commercialization," in *Borders in Cyberspace*, Kahin and Nesson, eds., MIT Press.

[9]See http://www.cio.gov/archive/paperwork_reduction_act_1995.html.

[10]See http://www.whitehouse.gov/omb/circulars/a130/a130trans4.html.

In a number of nations, particularly in Europe and in many developing countries, publicly funded government agencies treat their information holdings as a commodity to be used to generate revenue in the short term. They assert monopoly control on certain categories of information in an attempt—almost always unsuccessful—to recover the costs of its collection or creation. Such arrangements tend to preclude other entities from developing markets for the information or otherwise disseminating the information in the public interest. The world scientific and environmental research communities, and especially developing nations, are particularly concerned that such practices have decreased the availability of critical data and information. Moreover, firms in emerging information-dependent industries seeking to utilize public-sector information find their business plans frustrated by restrictive government data policies and other anticompetitive practices.

Recent economic research and initiatives at the European Commission, the United Nations Educational, Scientific and Cultural Organization, and the Organisation for Economic Co-operation and Development, as well as in individual countries, such as China's Scientific Data Sharing Program, are helping to create an international framework for open and global data sharing. There has been an emerging recognition in Europe as well that open access to government information is critical to the information society, environmental protection, and economic growth. A "government commercialization" policy for public information cannot succeed in the face of social and economic evidence and evenhanded application of competition policies. Conversely, open government information policies foster significant, but not easily quantifiable, social and economic benefits to society.

In order to achieve a successful international framework for open access to public scientific information, governments should:

1. Support full, open, and unrestricted international access to scientific data for public interest purposes—particularly statistical, scientific, geographical, environmental, and meteorological information of great public benefit. Such efforts to improve the exploitation of public-sector information contribute significantly to maximizing its commercial, research, and social values.

2. Allow the private sector to take an active role in using public-sector information to meet the diverse needs of citizens and users for such products and services. Meeting these needs requires entrepreneurial and pub-

lishing skills that are most evident in the private sector. Market needs are best served by open and unrestricted access to public-sector information.

3. Prohibit copyright protection for government information, limit fees to recouping the cost of information dissemination only, and eliminate restrictions on reuse. This will allow diverse entities to make new and innovative uses of public-sector information. However, attribution of data sources should be made, e.g., through the use of electronic watermarks or appropriate citations.

4. Avoid asserting a monopoly—either public or private—on public-sector information. Governments and societies both lose when governments treat their information as a commodity to be sold or allow a private-sector entity to "capture" the information on an exclusive basis.

5. Develop and maintain strong freedom of information laws to foster greater transparency and public trust in government.

Policy Considerations on Government Information Sharing in China[11]

Information policy research may be divided according to government, public, and commercial information. This section focuses on government information policies in China.

As in the United States and other countries, government information sharing activities in China must be based on policy studies. The key issue is whether government information should be free and open or not. The opening up of government information is subject to competing policies of state security, commercial confidentiality, and personal privacy. There are conflicts and coordination between the state and the private sectors, and among different groups. Information sharing is also subject to information system security.

There are several key values and principles informing the sharing of government information. Information is a strategic national asset of potential value to society and the economy. Government policy should seek to promote information sharing and the information sector. However, this needs to be done by balancing the costs and benefits to the providers and users of the information. The principles of open access, public-domain sta-

[11]Based on a presentation by Jun Li, National Macro Economic Research Institute, China.

tus, and free use of information must be balanced against the protection of rights of the producers of the information. Government information sharing requires consideration of the following aspects:

➢ Planning the national system as a whole;
➢ Developing directories and catalogs for government information;
➢ Establishing a chief information officer system for government information, and responsibility for information access and restrictions;
➢ Development of a one-stop information service;
➢ Understanding the social and economic benefits of access and sharing;
➢ Marketing and exploitation of information;
➢ Classification of information by users;
➢ Integrating government information research across the government; and
➢ Enhancing exploitation and usage of government information to increase information availability and benefits to government, and to drive the information industry.

Comparative Aspects of Policies for Open Access to Scientific Data in the United States, European Union, and China[12]

Effective management of scientific data has become a vital component of the research infrastructure in the information era. Many countries have developed their own policies and mechanisms to manage and share their scientific data and, as a result, a set of laws and regulations governing these activities has been gradually formed and improved. The United States, the European Union, and China have different policies regarding access to publicly funded scientific data. The United States supports "full and open" access to many kinds of scientific data and considers publicly funded data as a public good. However, European dissemination policies are based on the market value of public-sector scientific data. As a potentially big producer and user of scientific data, China needs to clarify its policies and

---

[12]Based on a presentation by Chuang Liu, Global Change Information and Resource Center, Institute of Geography and Natural Resources Research, Chinese Academy of Sciences. See also, Chapter 18, "Recent Developments in Environmental Data Access Policies in the Peoples' Republic of China," by Chuang Liu in *Open Access and the Public Domain in Digital Data and Information for Science: Proceedings of an International Symposium,* National Academies Press, Washington, DC, 2004.

establish different management mechanisms for various data produced or funded by the government. China has already begun to do so with its recent Scientific Data Sharing Program, as discussed in Chapter 2.

Data Sharing in Scientific Databases of the Chinese Academy of Sciences[13]

Ever since its initiation in 1983, the Chinese Academy of Sciences' Scientific Database and Applications System (SDAS) has developed quickly in its construction, technology application and development, information services, and other functions. It has become the largest scientific database cluster in China, with 45 collaborative institutions providing over 8 terabytes of data through 313 specialized databases. There are still greater challenges to scientific data sharing and services, however, requiring innovation for the traditional project management and application models. Therefore, from the beginning of the tenth Five-year Plan of the Chinese Communist Party, the SDAS has focused on research on data sharing policies with standard criteria, in addition to its data resources and system platform construction, to meet the growing external and interdisciplinary demand for data sharing through remote access, research collaboration, and information integration.

Setting up sharing policies for the SDAS is a major project to promote scientific data exchanges, enable further applications, and establish a series of fundamental standards for continuous data development. The methods for data sharing and management involve the establishment of sharing principles, classification, distribution requirements, collective management, and the protection of data owners' rights and interests. The principle of "full and open" data access is being clarified and costs and revenues are distributed on a reasonable basis. Based on current legal sources in China, the rights, obligations, and proper conduct relevant to data sharing are defined in three main categories—the data producers, distributors, and end users. The sharing policy of the SDAS is not merely an ordinary administrative management regulation, but the creation of a new scientific tradition based on data sharing in research developed from changing legal sources. The policy encourages data sharing, which can provide theoretical and practical

[13]Based on a presentation by Yun Xiao, Computer Network Information Center, Chinese Academy of Sciences, available at http://www7.nationalacademies.org/usnc-codata/Xiao_Yun_Presentation.ppt.

guidelines for data sharing services and the continuous development of scientific databases, consistent with the trends of the knowledge economy.

The Data Sharing Policy of the Chinese Ecosystem Research Network[14]

The Chinese Ecosystem Research Network (CERN) has 36 field observation and research stations across China, and each station has produced a large amount of data through monitoring, experiments, and research. Users worldwide can share most of those data, in accordance with the CERN Data Sharing and Management Rule, which was issued in 2002 by the Chinese Academy of Sciences.[15] This rule protects the rights of the data producers and permits these data to be shared widely, following the principle of keeping a balance between rights and obligations. The regulation divides data into two types, monitoring data (e.g., observational data from sensors) and data from research projects. In addition, it specifies five classes of users: related national departments, CERN members, members of the Chinese Academy of Sciences, domestic research and other nonprofit institutions, and others. This last category includes non-Chinese researchers. Users in each class have the same rights and obligations.

The rule established initial periods of exclusive use for data producers, typically ranging from one-half year to two years. The producers have priority to use their own data within the initial protection periods; other users can access those data following those periods, or even within the periods, if they obtain permission from the producers and provide attribution. Data producers may make data available on their own initiative.

Data Sharing Policy of the National Institutes of Health[16]

The sharing of biomedical data is essential for expedited translation of research results into knowledge, products, and procedures to improve hu-

[14]Based on a presentation by Panqin Chen and Tieqing Huang, Bureau of Science and Technology for Resources and Environment, Chinese Academy of Sciences.

[15]Both CERN's data policy and metadata are available on the CERN Web site at http://www.cern.ac.cn:8080/index.jsp.

[16]Based on a presentation by Belinda Seto, National Institute of Biomedical Imaging and Bioengineering, U.S. National Institutes of Health, available at http://www7.nationalacademies.org/usnc-codata/SetoPresentation.ppt.

man health. The U.S. National Institutes of Health (NIH) endorses the timely sharing of research data to serve these and other important scientific goals, particularly research data from NIH-supported studies for use by other researchers.

The NIH established new guidelines through its Data Sharing Policy, which became effective on October 1, 2003.[17] The NIH expects timely release and sharing of final research data for use by other researchers. In general, timely release means no later than the acceptance for publication of the main findings from the final dataset. No timeline is specified; however, the time of release will vary depending on the nature of the data collected. The policy also recognizes that investigators who collected the data have a legitimate interest in benefiting from their investment of time and effort, and have a right to an initial period of exclusive use. NIH also expects grant applicants to include a plan for data sharing or to state why data sharing is not possible, especially if $500,000 or more of the direct cost is requested in any single year. Finally, NIH expects contract offerors to address data sharing regardless of cost.

The NIH recognizes that there are different ways to share data. The researcher may provide the data in publications, share the data upon request, place datasets in public archives, or place data in restricted-access data centers. The NIH also provides support for data sharing, either directly through the grant application through a budget line item and budget justification or through administrative supplements.

Nonetheless, there are challenges to sharing data from research on human subjects because of data sensitivity, the need to protect confidentiality, and the danger of deductive disclosure. Investigators need to carefully consider release of data in studies with very small samples or studies collecting very sensitive data. However, even these data can be shared if safeguards exist to ensure confidentiality and protect the identity of subjects.

The solution is a multi-tiered system that divides data into three categories: open public-use data, restricted-use contractual data, and a cold room for on-site data use for certain authorized users only. Additional specific procedures are used to help ensure the privacy of all human subjects.

[17]See http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html. Additional details about the NIH Data Sharing Policy and privacy restrictions are available from the NIH Office of Extramural Research Web site on Data Sharing Policy at http://grants.nih.gov/grants/policy/data_sharing; from the Office for Civil Rights Web site with Departmental guidance on privacy at http://www.hhs.gov/ocr/hipaa/; and from NIH and other research agencies about the Privacy Rule at http://privacyruleandresearch.nih.gov.

PANEL DISCUSSION ON ECONOMIC AND INSTITUTIONAL ISSUES

Introduction[18]

Scientific data and journal activities in publicly funded research throughout the world have public good characteristics that make their broad availability especially beneficial. Some of these aspects were already discussed in the previous section on Legal and Policy Issues. In this section, the first two presentations briefly examine some of the economic and institutional issues associated with the publishing of scientific journals. The final one looks at the interface of the public and private sectors in the area of environmental data.

The Peculiar Economics of Scientific Information[19]

The technology of the production and distribution of information is very different from that of producing and distributing consumer goods like automobiles or shoes. The marginal cost of providing an automobile or a pair of shoes to one more consumer is about the same as the average cost per consumer of producing these goods. In contrast, while there are costs to gathering or creating information, the marginal cost of supplying this information to an additional consumer is small. In fact, with access provided on digital networks, this marginal cost of dissemination to each additional user is almost zero. Because of this simple technological fact, efficient pricing methods are very different for information goods than for ordinary physical goods. Prices for ordinary commodities serve the dual purpose of repaying the producers for the costs they have incurred and of restricting consumers from consuming goods that are worth less to them than the costs to others of providing these goods. With information goods this does not happen. The social cost of allowing an extra reader online access to information is negligible. Thus it will not be possible to repay the costs of producing information in the first place by pricing at marginal cost. If, on

[18]Paul F. Uhlir and Julie M. Esanu, U.S. National Academies.

[19]Based on a presentation by Theodore Carl Bergstrom, University of California, Santa Barbara, United States, available at http://www7.nationalacademies.org/usnc-codata/BergstromPresentation.ppt and http://www7.nationalacademies.org/usnc-codata/BergstromPresentation7.ppt.

the other hand, prices are set high enough so that revenue from users will repay the total supply cost of the information, then many potential users will be excluded from access to this information even though they would benefit.

Institutions responsible for provision of scientific information, particularly journals, that have been prominent in the academic community include not-for-profit scientific societies and university presses, for-profit publishers, and government agencies. In recent years, there have been extensive discussions regarding the benefits and costs of the not-for-profit versus the for-profit business models, open access versus subscription-based publishing, and the practice of some of the "nonprofit" societies of using profits from their journals to subsidize other activities.

The professional scientific societies publish journals, books, and datasets. They usually charge subscriptions to libraries and individuals to cover their costs. Some make a profit from their journal subscriptions and use it for other activities (e.g., *Science* of the American Association for the Advancement of Science). University presses are used to enhance the prestige of the university.

For-profit commercial publishers own journals and charge subscriptions to libraries and individuals in order to make profits. Because they own copyrights to the articles in their journal, they have a monopoly on these articles. The prices charged by the commercial publishers are usually much higher than not-for-profit journal prices.

There are several potential advantages of the professional society journals. Societies can lend prestige and attract top authors. Society journals are not-for-profit and usually very cheap. The society journals also tend to maintain very high quality standards. The professional society journals also have some disadvantages, however. Societies tend to be conservative. They often are slow to adopt new ideas and are controlled by an aging elite. There is little incentive for officers of society journals to invest new effort, or expand into new fields.

One advantage of commercial journal publishers is their entrepreneurship. For example, in the 1970s the economics profession was growing rapidly, but society journals expanded slowly. Specialized field journals were rare, so North-Holland (and other companies now merged into Elsevier) started about 30 new journals. About 12 of these could now be called leading field journals. Nevertheless, there are substantial disadvantages from commercial journals. High prices to libraries mean access only for the wealthiest users. There is also great inefficiency in restricting access to publicly funded scientific information when the marginal cost of providing access online to each additional user is near zero.

Overall, the policies for electronic journals that publish the results of publicly funded research in the networked environment should seek to:

➢ Promote open Internet access to scientific papers by supporting open-access archives for scientific work and requiring government supported research to be posted in these archives.

➢ Encourage evaluation and quality standards without discouraging innovation.

➢ Avoid intellectual monopolies and cliques.

➢ Encourage independent scientific societies and university presses—more than one per discipline, to prevent monopoly and cliques.

➢ Encourage innovative "journal substitutes" with alternative forms of evaluation and certification.

➢ Search for new solutions to match new technology.

Launching an Open-Access Journal[20]

The rapid spread of the Internet and concurrent innovations in electronic publishing have dramatically transformed scientific publishing and the dissemination of technical information and data. Electronic publishing has spawned an entirely new economic model for the sharing of scholarly research, just as the Internet has transformed other areas of commerce and information management. This model, known as open-access publishing, is predicated on the fact that an article published on the Internet can be read by ten readers or ten million with virtually no additional cost to the publisher. Many new open-access publishers, including the U.S.-based Public Library of Science (PLoS) and the U.K.-based BioMedCentral, generate revenue through one-time publication charges, which are generally paid from the author's research grant or from institutional or library funds, rather than by charging subscriptions or fees for readers. Covering the cost of publication once through publication charges then allows the content to be made free to all readers.

New open-access journals face many challenges, just as any new journal trying to compete with the most prestigious ones in their field would. In

[20]Based on a presentation by Helen Doyle, Public Library of Science, United States, available at http://www7.nationalacademies.org/usnc-codata/HelenDoylePresentation.ppt.

addition to questions about the viability of the open access economic model, these challenges include: convincing authors to submit their best work; establishing a rigorous peer review system and respected editorial board; creating a high-quality production; and marketing the journal to the appropriate audiences. Conversely, there is also great opportunity for new electronic open-access journals to be innovative in their publishing practices. Some recent innovations include managing manuscripts and the peer review process electronically; developing interactive Web functionalities for different content and audiences; and encouraging the creative reuses of the content allowed by a more liberal copyright license. These challenges and opportunities are all being confronted by the PLOS's new open-access journal, *PLoS Biology*, which was launched in October 2003.

Involving the Private Sector in the Environmental Enterprise[21]

The conduct of environmental (atmosphere, ocean, land) research and applications requires extensive and thus costly observations taken from all parts of the world and participation by the government, academic, and private sectors. Ten years ago, the roles of these sectors in the United States were reasonably straightforward: government agencies collected observations and disseminated information to the public; academia used government-collected data for research; and private-sector organizations used observations collected by governments for developing information products targeted to paying customers.

However, advances in science and technology have made it possible for private-sector organizations and academia to perform many government tasks, including collecting data, running models, and disseminating information. These overlapping activities create potential inefficiencies and friction, especially because the sectors have different goals and require different data policies.

On the one hand, U.S. government agencies and environmental science researchers require full and open access to data (i.e., data are available without restriction for any use for no more than the cost of reproduction). On the other hand, private-sector organizations generally need to restrict

[21]Based on a presentation by Anne Linn, U.S. National Academies, available at http://www7.nationalacademies.org/usnc-codata/AnneLinnPresentation.ppt.

access to data in order to generate a financial return and their policies are usually proprietary. Two recent National Research Council reports provide guidelines for resolving these different data policy requirements and for easing friction between the sectors.[22]

The *Fair Weather: Effective Partnerships in Weather and Climate Services* report examined conflicts among government, academia, and the private sector dealing with weather data. It concluded that establishing rigid boundaries between the sectors and defining what each should do is counterproductive.

The *Resolving Conflicts Arising from the Privatization of Environmental Data* report examined these issues for all environmental data and provided criteria for purchasing data from the private sector and for transferring government data collection and product development to the private sector. The report concluded that transferring government data collection and product development to the private sector can be beneficial as long as the following conditions exist:

➢ Avoiding market conditions that will give private companies a monopoly;
➢ Preserving full and open access to key data sets and products;
➢ Assuring that a supply of high-quality information will continue to exist; and
➢ Minimizing disruption of ongoing uses and applications.

For economic and data policy reasons, however, public funding for data collection and analysis should continue, focusing contributions of the private sector primarily on distribution of value-added products and collection of certain observations.

[22]National Research Council (NRC). 2003. *Fair Weather: Effective Partnerships in Weather and Climate Services*, National Academies Press, Washington, DC; NRC. 2001. *Resolving Conflicts Arising from the Privatization of Environmental Data*, National Academy Press, Washington, DC.

PANEL DISCUSSION ON MANAGEMENT AND TECHNICAL ISSUES

Introduction[23]

Although the technical aspects of digital scientific data and information activities are typically quite well understood and do not raise inordinate barriers except, perhaps, related to costs, the proper management of such activities, especially data preservation and dissemination, poses some unique hurdles. In the discussion below we identify some of the problems that ought to be considered in properly planning scientific data center activities.

Operating a Twenty-First-Century Data Center[24]

In the past several decades, large-scale data resources have assumed an increasing role in scientific research, particularly research on Earth and its environment. There are a number of reasons for this, including advances in computational technologies, software, and observational capabilities, and a growing emphasis on empirical and interdisciplinary research. One of the consequences of the increasing dependence on data resources across fields of science is that in the coming years, the scientific community must devote a larger share of its resources and energies to data management and preservation than it has in the past.

A recent Priority Area Assessment on Scientific Data and Information, presented to the strategic planning committee of the International Council for Science (ICSU) in June 2004, emphasized that the scientific community needs to develop strategies for data management over time periods of decades to centuries.[25] Effective planning for long-term data management requires clarification of the role of data centers versus archives, obtaining regular scientific advice on data management and archiving decisions, and developing long-term financial support for data center and archival operations.

The report also stressed the critical importance of professional management of data. That is, it is no longer sufficient for the scientists who analyze scientific data to be responsible for managing those data; profes-

[23]Paul F. Uhlir and Julie M. Esanu, U.S. National Academies.

[24]Based on a presentation by Roberta Balstad, Center for International Earth Science Information Network, Columbia University, United States.

[25]The final Priority Area Assessment on Scientific Data and Information report is available from ICSU at http://www.icsu.org/1_icsuinscience/DATA_Paa_1.html.

sional data managers, with professional expertise, are needed. Professional data managers must be knowledgeable about the technological drivers of data center operations. They have to understand the financial implications of hardware, software, and training. They must be able to manage the rapid pace of change and the timely updating of data, software, and hardware. And they should develop career incentives and rewards for effective data production and management.

Finally, the report recommended that there be a common international approach to data and information management. The benefits of an international approach to scientific data need to be properly understood and communicated so that common strategies, standards, and software interoperability can be developed.

Managing the Effects of Programmatic Scale and Enhancing Incentives for Data Archiving[26]

The challenge of a digital scientific archive is to engage scientists in the process of archiving their data and provide the mechanism for archiving. The functions of an archive are to store data safely and reliably; build a catalog and structure; maintain storage across technology generations; review new data (quality assurance, metadata); "advertise" contents; find data for users with query and browse logic; and distribute data by providing access and references to documentation.

An effective scientific data archive operates on several presumptions. Information sharing is important. Multidisciplinary data access will foster more robust scientific discoveries. Archiving can always be improved. The number of permanent data archives will increase, which will increase the value of all archives connected on an interoperable basis through digital networks.

Even if an appropriate archive exists, however, there are many reasons why scientists do not archive their data. They lack incentives and proper acknowledgement; they may be concerned about giving up publication rights; their research may have poor planning or insufficient resources for data management and preservation; they may not believe the archive will

[26]Based on a presentation by Raymond McCord, Oak Ridge National Laboratory, United States, available at http://www7.nationalacademies.org/usnc-codata/RaymondMcCordPresentation5.ppt.

get long-term support; and they may lack training or be unsure about the metadata content they should provide.

Nevertheless, research managers can provide many good arguments and incentives to scientists for archiving their data, including the following points:

➢ Recognition for archiving. Scientists need to receive some career benefits for their archiving-related work. Consider scientific journals that also provide companion "data publications."

➢ Emphasize good scientific practice.

➢ Promote professional development and training. Provide daily interactions between scientific and information specialists. Allow a reasonable time for initial discovery. Provide support for long-term "stewardship."

➢ Provide institutional incentives. Archiving should be required by the sponsor. Data archiving should be "in the plan" and resources available to support it. Interweave archiving with the planning and publication processes.

➢ Exploit technological advances. It is technically easier now and there are more options.

➢ Plan for managing change. Change is inherent in research, but managing change without prior planning can become consumptive. Changes may cause confusion and diminish data usefulness. (See the next section for more details about managing change in archiving.)

Perhaps the best argument is that effective archiving supports better science. Archiving extends data usefulness. Archived data increase the volume and diversity of our information base for doing research.

Managing the Effects of Change on Archiving Research Data[27]

The archiving of scientific data and information is made more difficult by the evolving changes associated with research accomplishments. Research discoveries lead to a continual series of revisions to sampling schemes, measurement methods, and scientific objectives. All of these changes add to

---

[27]Based on a presentation by Raymond McCord, Oak Ridge National Laboratory, United States, available at http://www7.nationalacademies.org/usnc-codata/Raymond McCordPresentation.ppt.

the scope and complexity of information that must be recorded and logically organized as part of a successful data archive. Recording and communicating these changes for future data users is facilitated by additional supporting information and an evaluation of the rules that define the information. Most of the available information technology (hardware, software, and implementation methodology) originates from business applications, which are designed to accommodate fundamentally different patterns of change. Managers of scientific data archives will need to adapt the traditional designs of information systems to meet these special features of research data and its users. Management also needs to encourage extra effort during initial design and later operations to accommodate the future changes that will occur.

The structure and content of scientific data and information can be very complex. Successful archiving of data requires that the variation in this complexity be minimized. The efficient operation of information systems and effective communication with future data users are enhanced by minimizing the variation in the logic, concepts, and keywords used in the metadata. Some of the complexity is inherent to the variety of measurements and materials included in the research and cannot be avoided. Additional complexity occurs as archived information is aggregated into more extensive systems and accessed by broader user communities.

There are many management and institutional issues that must be considered to avoid unnecessary complexity and uncertainty in the archived information. The effects of the varying dimensions of programmatic scale (volume, diversity, longevity of data and research programs) need to be considered. Institutional impediments and incentives also affect the willingness of scientists to contribute information to archives. The documentation and archiving of data should be integrated with the publication process as part of the "modern scientific method" and should receive similar incentives. Management should reinforce these practices by insisting on early planning for data archiving and providing specific rewards for these activities. Other management issues include protecting initial discovery opportunities, supporting long-term stewardship of data (answering questions after the project is completed), and providing "cross-training" of archive personnel in both scientific and information disciplines.

There are several key rules for creating data sets for archiving:

➢ *Unique occurrences.* Each type of measurement should be represented in a consistent way and each measurement event should be represented by only one value. If multiple versions of datasets accumulate, provide version information, explain version differences, and document the effective data range for each version.

➢ *Identifiers.* Each value should be associated with a parameter name and each measurement value should have a quality indicator and link to a method description. Whenever possible, remove multiple aliases for the same identifier (e.g., sample identifier, site identifier or name, measurement name, etc.).

➢ *Place and time.* Each value should be associated with a unique place name with a quantitatively defined location (geographic coordinates). Each value also should be associated with a date and time. Do not confuse data and time for measurements with data and time for storage revisions or data and time ranges for measurement or encoding methods.

➢ *Data storage and transport.* Data should be stored or managed with a database management system or self-documenting data format. Include data analysis software in the data management suite. This is useful for comparing versions of data that accumulate over time. Also include data format conversion software in data management suite, which is useful for migrating data from one storage technology to another.

Finally, there are a number of best practices for preparing ecological and ground-based data sets to share and archive:[28]

- ➢ Assign descriptive file names;
- ➢ Use consistent and stable file formats;
- ➢ Define the parameters;
- ➢ Use consistent data organization;
- ➢ Perform basic quality assurance;
- ➢ Assign descriptive data set titles; and
- ➢ Provide the necessary documentation.

Special Considerations for Archiving Data from Field Observations[29]

Archives depend on logical rules for information structures and consistent codes for metadata. Different types of data and information pose different types of challenges and archiving requirements, however. For ex-

---

[28]Cook et al. 2001. Bulletin of the Ecological Society of America. Available at http://www.daac.ornl.gov/DAAC/PI/bestprac.html.

[29]Based on a presentation by Raymond McCord, Oak Ridge National Laboratory, United States, available at http://www7.nationalacademies.org/usnc-codata/RaymondMcCordPresentation5.ppt.

ample, scientific data from field investigations are fundamentally different from laboratory observations. Laboratory studies are conducted under controlled conditions, whereas field observations are collected from incompletely controlled environments. Data archives for field observations must include additional design features to accommodate, but minimize and rationalize, this additional complexity. These features include a need to address multiple schemes for location information, temporary changes in methods, unmeasurable events, evolving reference lists, and a containment strategy for exceptions.

Location information typically involves multiple geographic coordinate systems. Conversions from one system to another may not be reversible without some loss of information. It is important, therefore, to test changes before large-scale conversions are made. A visualization capability is essential. Location information also usually encounters multiple naming schemes, such as unofficial "folk" names; divergence in naming schemes at local, regional, and national scales; and connecting historical name changes.

Archives also must take account of temporary changes in data collection methods. For example, the field sampling protocol may be insufficient, resulting in a need to restructure the metadata to record the temporary change. Field observations at remote sites can experience various anomalies and instrument malfunctions. A robust scheme for missing value representation is needed and data analyses must correctly exclude missing values.

There also may be unmeasurable events that are either too small or too large. How do you record values that are below the detection limit (but not zero)? You can set all values to the minimum detection limit, set all values to the midpoint between detection and zero, set all values to zero, or retain estimated value, but include a quality flag. You will need to select one of these strategies and document the choice, realizing that the choice can have significant impacts on summary statistics.

How do you record the biological population when there are too many individuals to count? You can record some arbitrary large number or flag it as unmeasurable. Similar problems can occur with wide ranges in chemical concentrations. Different schemes may have impacts on results from statistical analyses or on setting quality assurance limits.

Evolving reference lists pose further challenges. Taxonomic lists assume agreement on a single and accepted classification scheme, which may not be true. Individuals who are involved only infrequently also may not be fully identifiable. Later samples may enable fuller identification, which may require recoding earlier records to match newer identification. Chemical constituents may have similar classification problems.

A containment strategy is thus needed to deal with the exceptions. The "90/10" rule is a useful guideline: approximately 90 percent of the data can be described by a few logical rules, while approximately 10 percent of the data cannot be described by rules and contain numerous and isolated exceptions. Put the information that cannot be described by rules in an alternative structure that can be labeled as "user beware," support detailed and various documentation, and accommodate and communicate numerous exceptions. However, when there are "too many" logical rules, the archiving process will become inefficient and tedious, so it is important to make adjustments as needed.

Toward a Balanced Performance Appraisal System in the Digital Era for Data Archiving and Sharing in China[30]

Although data archiving and sharing are not new problems in China, they became worse with the advent of the digital era. Traditionally, the data producers and the data users were not very different in terms of their academic recognition and reputation. They collected, archived, and used data, and most of the data products were published and thus were available to the public at affordable prices. The published data sets were also regarded as academic achievements by their peers and funding agencies. This mechanism encouraged some scientists to archive and publish their research data.

In the early years of the digital era, however, a number of factors began to hinder data sharing, some of which have continued to the present. First of all, data management as a whole became more expensive and needed more investment. Second, digital publishing of research data is not the same as in the print format and the authorship of digital data products is less clear. Third, those scientists who are both data producers and data users (as is true in most cases) have a competitive edge over their peers if they can use their data exclusively forever. Fourth, public funding agencies may not be able to make data available to other scientists, because even the agencies sometimes have obtained the data according to restrictive contracts. Fi-

[30]Based on a presentation by Zhengxing Wang, Global Change Information and Research Center, Institute of Geography and Natural Resources Research, Chinese Academy of Sciences.

nally, China's scientific community has not been fully aware of the fundamental role that data have played in scientific inquiry. Data collection, processing, validation, archiving, and sharing all have not been included in China's "academic performance appraisal system." In a system in which print publishing gets nearly all the information funding and digital data archiving and sharing get much less, one has little incentive to continue working in data archiving. Therefore, it is critical to develop a more comprehensive and balanced performance evaluation system to foster and sustain digital data archiving and data sharing in China.

Earth Science Data and Information Management in Western China[31]

Over the past few decades, the western regions of China have been the focus of important earth science research and a lot of earth science data and information have been accumulated there. Recent studies have focused on the plateaus, mountains, deserts, vegetation, hydrology, and ecosystems in that region from the perspective of different fields of study, including ecology, environmental and earth sciences, sociology, and regional development. The historical data and information resources are valuable for such studies.

Based on the analysis of the distribution of information resources and users, and on user requirements and the support of data, many functions need urgent attention in the national framework of earth science data sharing in China. For example, reproducing single copies of data and information, accelerating the sharing of data and information, meeting the demands of potential users, and promoting the use of information to benefit the western region of China all require attention.

There are several actions that could be taken in this regard. One should be to assist institutions that have not organized and digitized their historical data, and encourage them to provide an index and to become members of the data sharing system. Another is to establish a special management program to reduce the costs of data collection and promote their application. Finally, it would be very useful to establish a digitally networked clearinghouse for earth science data and information that involves the relevant organizations and persons in western China.

[31]Based on a presentation by Chengquan Sun, Scientific Information Center for Resources and Environment, Chinese Academy of Sciences, available at http://www7.nationalacademies.org/usnc-codata/Sun_chengquan_Presentation.ppt.

Data Integration and Management: The Protein Data Bank Perspective[32]

The Protein Data Bank (PDB) is the worldwide repository for the structures of biological macromolecules.[33] The PDB provides a rich history from which to explore the practices of biological data management, because its data set has many characteristics found in other biological data—diversity, complexity, and variable quantity and quality of annotation.

A database resource is only as good as the data it contains. Data representation, acquisition, annotation, and distribution are all essential functions. The data representation (metadata) used by PDB is the macromolecular Crystallographic Information File (mmCIF) standard dictionary. The mmCIF dictionary conforms to a subset of encoding rules embodied in a Self-defining Text Archival and Retrieval (STAR) syntax. STAR has provisions for defining scope, resting, looping, and other aspects. Conforming to STAR is a Dictionary Definition Language (DDL) that defines how dictionaries are described. DDL has provisions for fully characterizing the terms in the domain and is relational in nature; that is, there is the notion of relations (categories), attributes (specific data names), primary and secondary keys (mandatory data items), and so on. The data defined by mmCIF consist of name-value pairs where each name must be defined in the mmCIF dictionary. The mmCIF dictionary can be characterized as having the features of extensible markup language (XML) document type definition or schema. Although the dictionary was written in the STAR format, the ontology and its derivations are independent of STAR or any other particular file format. It can be automatically converted to other formats. It provides the foundation of integrated software systems for building robust automated data pipelines. The PDB has been actively involved in various aspects of automated and accurate data acquisition, annotation, and distribution.

Biological data management concerns more than just the technical aspects, however. There are sociological and political issues as well. A key element for success is good communication among those running the resource, who need to have diverse skill sets, and among every member of the team and the communities they represent. Community feedback must be

[32]Based on a presentation by Zukang Feng, Protein Data Bank, United States, available at http://www7.nationalacademies.org/usnc-codata/ZukangFengPresentation.ppt.

[33]See http://www.wwpdb.org/.

treated seriously and lead to a prioritized set of action items to be addressed by the resources available. The technology must take advantage of the most recent innovations in hardware and software. These technological developments, however, must be introduced so as to enable and not disrupt the users of the resource. It is critical to maintain an interactive dialogue with the user community about desired new functionalities and the feasibility of their implementation. Beyond all else is the need for good data and a robust data representation that is flexible enough to meet the needs of the changing science.

5

Summaries of Presentations on Thematic Issues

There were several examples of existing scientific data and information activities in China, the United States, and internationally highlighted at the workshop. These activities can benefit from the policies and management practices outlined in the previous chapter. They also provide practiced experience at the working level and models for other similar activities in other institutional, disciplinary, and national contexts. They thus represent a bottom-up approach to the evolution of national and international policy and practice with regard to public scientific data and information resources. The examples summarized below are organized according to three thematic areas that were the focus of the workshop: (1) life sciences and public health data; (2) earth sciences, environmental, and natural resources data; and (3) scientific information, journals, and digital libraries.

EXAMPLES OF LIFE SCIENCES AND PUBLIC HEALTH DATA ACTIVITIES

The Chinese Management and Sharing System of Scientific Data for Medicine[1]

The initiation of the Management and Sharing System of Scientific Data for Medicine (the Medical Data Sharing System) was a key project in

[1]Based on a presentation by Depei Liu, Chinese Academy of Medicine and Chinese Academy of Engineering, available at http://www7.nationalacademies.org/usnc-codata/Liu_Depei_Presentation.ppt.

2003 of the China Scientific Data Sharing Program (SDSP) under the National Basic Platform for Science and Technology in the Ministry of Science and Technology. Oversight of this Medical Data Sharing System has been undertaken by the Chinese Academy of Medical Sciences, the Chinese Center for Disease Prevention and Control, the Chinese People's Liberation Army General Hospital and Graduate Medical School, and the Chinese Academy of Traditional Chinese Medicine.

Various scientific data resources in medicine are being integrated together by the Medical Data Sharing System. The system covers most fields of medicine, including basic and clinical medicine, public health, traditional Chinese medicine, other special areas of medicine, and pharmacology. The system also has a database specifically for SARS and respiratory diseases.

The Medical Data Sharing System has important roles in many respects. It is used to:

1. Serve medical research and teaching, and improve innovation in medical sciences and technology;
2. Improve the overall level of prevention, diagnosis, and treatment of diseases;
3. Enhance the personal health consciousness of people, thus promoting good health in society;
4. Strengthen China's ability to plan for and respond to sudden incidents in public health; and
5. Provide an improved basis for government policies, as well as promote the development of the national economy and the medical system.

The organizational and management model for this system consists of a leading group, an expert group, and a working group. The leading group consists of the leaders of participating institutions and the professionals in charge of scientific administration who are responsible for the decision making, organization, and coordination of the system. The expert group is composed of experts from the China-SDSP and professionals in project-supported institutions. The expert group is responsible for steering and authorization of the general design, selection of participating institutions, and inspection of the project's progress. The working group consists of members selected from the working group of the China-SDSP and employees of the participating institutions. It is responsible for drafting the feasibility report, general project design and implementation, and development of the main database.

The Medical Data Sharing System benefits all sectors of society, including the government; medical, scientific, and educational institutions; private enterprises; and the general public. The system promotes a cooperative environment and will be operated on a free and not-for-profit basis.

International Medical Scientific Data Sharing[2]

There are many kinds of medical scientific data that can be shared internationally. They include primary data, data products, and related information obtained from the activities of medical clinics, teaching, and research. Medical data are necessary resources for the development of medical sciences and have characteristics common to other kinds of scientific data, including reusability and potential long-term value.

The Human Genome Project has provided a successful example of access to and sharing of biomedical scientific data. This project used modern informatics techniques in human genomic research and organized top scientists and facilities to cooperate on the same project based on the same standards. In recent years, similar data sharing projects have been initiated in other areas of science internationally, such as neurology, human anatomy, and proteomics, all of which are described in the sections that follow. These projects have produced large changes in the behavior of medical researchers toward greater cooperation, but there are many unmet needs for medical scientific data access and sharing by the research, teaching, and biomedical industry communities.

Several problems can be identified. The lack of data access and sharing is especially acute in developing and least-developed countries. This results in redundant research and inefficiencies. Medical data resources are separated and many lack quality and uniform international standards and data exchange protocols. There also is a lack of adequate investment and attention to these problems.

The following suggestions are focused on improving the status of international medical scientific data access and sharing:

1. It would be helpful to set up an international coordination committee for access to and sharing of medical scientific data, responsible for developing plans for such activities and proposing appropriate guidelines.

[2]Based on a presentation by Ling Yin, People's Liberation Army General Hospital and Graduate Medical School, China.

2. In order to strengthen international cooperation in the establishment of medical scientific databanks it is necessary to establish relevant standards and uniform criteria for data contributions by the globally distributed scientists and facilities. New databank groups for medical scientific data similar to Genbank and the Protein Data Bank need to be established to make the data sharable, extendable, and authoritative.

3. To truly implement data access and sharing among the different medical specialties, China needs to establish an integrated international service system to break the barriers among basic, clinical, and preventive medicine, public health, pharmacology, and other areas. This will offer one-stop service via index and catalogue inquiries for data services, and a networked environment for therapeutics, prevention, control, and research on specific diseases.

4. China also needs to speed up the recruiting of expert personnel to implement access to and sharing of medical scientific data, and to offer effective training programs.

5. Finally, funding needs to be secured from multiple sources to guarantee long-term financial support and to facilitate global access to and sharing of medical scientific data and information.

China's Contributions to the Organisation for Economic Co-operation and Development's Neuroinformatics Data Sharing Initiative[3]

The human brain is the most complex system known. Achieving a better understanding of it is a key scientific challenge for the 21st century. Having developed sophisticated methods to investigate the brain in the finest possible detail, neuroscientists now face the challenge of managing the enormous amounts of raw data and the many useful inferences drawn from them. The neuroinformatics field therefore would benefit greatly from increased data sharing. Neuroinformatics research already has received billions of U.S. dollars and Euros to establish individual databases and platforms and to lay the groundwork for data sharing in the future.

---

[3]Based on a presentation by Yiyuan Tang, Institute of Neuroinformatics, Dalian University of Technology; Ling Yin, Neuroinformatics Center, PLA General Hospital and Graduate Medical School; and Xiaowei Tang, Neuroinformatics Center, Zhejiang University, China, available at http://www7.nationalacademies.org/usnc-codata/Yin_Ling_Presentation.ppt.

In 2000, the Organisation for Economic Co-operation and Development's (OECD) Global Science Forum approved the international Working Group on Neuroinformatics (WG-NI), which included 21 members and observers from OECD countries. Their goal was to establish a global knowledge management system and Internet research environment in which the data and results from research on the human neurological system would be available. Such an information initiative would be helpful for improving public health, scientific research, medical education, and the pharmaceutical industries. The three principal aims of the WG-NI group were to promote: (1) the establishment of databases and the integration of data resources; (2) data sharing policies for OECD members and observers and the development of working guidelines and rules for neuroinformatics data; and (3) the establishment of international neuroinformatics research networks.

The Chinese government also began to pay more attention to the Human Brain Project (HBP) and to neuroinformatics research. The Xiang Shan Science Conference for HBP and neuroinformatics, held in September 2001, was a major event that included senior Chinese government representatives from the National Natural Science Foundation of China and the Ministry of Science and Technology. The conference participants agreed that China should start work on the Chinese neuroinformatics project as soon as possible. Since 2001, the Chinese government has awarded several grants in the neuroinformatics field and supported the development of a neuroinformatics platform and digital network according to the OECD's WG-NI standards.

In 2003, the WG-NI requested the establishment of an international coordinating mechanism, the International Neuroinformatics Coordinating Facility (INCF). The proposed role of INCF was to optimize the accumulation, integration, standardization, exploitation, and sharing of very large amounts of highly diverse primary data and of large, structured neuroscience databases that are being generated worldwide by researchers who study the brain. The first INCF meeting was held in April 2004 at OECD in Paris to establish the INCF secretariat, initiate substantive activities, and develop a proposed funding scheme (the Program in International Neuroinformatics). This new international program is intended to promote international collaboration among researchers whose work will be funded by existing (or possibly new) national programs, eliminate national and disciplinary barriers, and provide a more efficient approach to global collaborative research and data sharing.

In order to accelerate Chinese neuroinformatics development, the neuroinformatics project has been suggested to become an important part of the China Science Data Sharing Project, especially for promoting international collaboration. Consistent with INCF's working plan, the Chinese neuroinformatics community will develop tools for manipulating and managing the data and standards and mechanisms for sharing these data among global researchers. The Chinese also will design and develop special-purpose analytical tools and algorithms, and create computational models of brain structure and function that can be validated using diverse data. These actions will advance the understanding of the human brain and may be expected to lead to breakthroughs in the prevention and cure of nervous system disorders and to improvements in the quality of life for humanity.

Long-Term Studies of Human Anatomy Using the Digital Human and Scientific Data Sharing[4]

Human anatomy is a cornerstone of modern medicine. In 1543, Vesalius published the seminal anatomy book, *On the Structure of the Human Body*, which was one of the starting points of modern medicine. Digital anatomy and the digital human represent a new revolution in medicine. The U.S. National Library of Medicine (NLM) began to discuss a long-range plan for the digital human in 1985. In 1989, the NLM Board of Regents submitted a long-range plan for the next 10 to 30 years of electronic imaging in biomedical research. The first stage of the plan was the Visible Human Project (VHP). The plan encouraged scientists to conduct research in various fields, including anatomical structure informatics, graphics technologies in biomedical imaging, basic medical research (e.g., developmental biology, neuroscience, cell and histological science, and molecular structure), clinical applications (e.g., image-guided surgery, actinotheraphy, anaesthesia, radiology, organ system imaging, orthopedics), and the development of related medical equipment using digital technologies.

In 1991, the University of Colorado signed a VHP contract with NLM. They completed two cryomacrotoming data sets (one male and one female) separately in 1994 and 1995. Since that time, the digital anatomy community has developed important applications for medical teaching and

---

[4]Based on a presentation by Donglie Qin, BME College, Capital University of Medical Sciences, China.

clinical practice, including breakthrough advances in medical imaging, such as more realistic rendering of three-dimensional images.

In 2001, the Federation of American Scientists presented the digital human initiative, which is based on human body digital simulations at three levels: the microlevel (molecular, gene, cell, and nanoscale), the meso-level (tissue and organ), and the macrolevel (whole body). Since 2001, the emerging Grid, Internet 2, and other advanced information technology capabilities and imaging devices have been used in the VHP and similar digital human activities.

In addition to the VHP, there are now many other digital human projects outside the United States, including the European Union, Japan, Korea, Singapore, and China. In China, the proposed draft for a digital human initiative was submitted to the government by scientists in 2001. Two experimental whole-body data sets and several high-resolution organ data sets (heart, kidney, and liver) have been completed. The next generation digital data sets and applications in medical teaching and clinical practices are expected to occur within the next 5 to 10 years. This brief history of digital human research underscores the need for implementing scientific data sharing in support of both research and applications in these digital human initiatives.

The Protein Data Bank: A Key Biological Resource[5]

As noted in the previous chapter, the Protein Data Bank (PDB) is the single international repository of three-dimensional data for biological macromolecules and currently contains over 25,000 entries. The concept of the PDB began to be formed during the late 1960s and early 1970s with community discussions about the need for such a resource. Protein crystallography was still in its infancy, but it was apparent to the producers of these structures as well as to the potential users that every structure contained valuable information that needed to be archived and maintained for posterity. In June 1971, key representatives of the two communities attended the Cold Spring Harbor Symposium on Quantitative Biology and agreed that the time was right to create the PDB. The PDB was established in October of that year at the Brookhaven National Laboratories as an archive for biological macromolecular crystal structures.

---

[5]Based on a presentation by Zukang Feng, Protein Data Bank, United States, available at http://www7.nationalacademies.org/usnc-codata/ZukangFengPresentation7.ppt.

In the 1980s, the number of deposited structures began to increase dramatically. This was primarily due to the improved technology for all aspects of the crystallographic process, the addition of structures determined by nuclear magnetic resonance methods, and the changes in the research community's views about data sharing. By the early 1990s, the majority of journals required a PDB accession code and government funding agencies adopted the guidelines published by the International Union of Crystallography requiring the deposition of data for all protein structures. The archive's growth has been accompanied by increases in both data content and the structural complexity of individual entries over the years.

In October 1998, the management of the PDB became the responsibility of the Research Collaboratory for Structural Bioinformatics (RCSB) at Rutgers University, together with the University of San Diego Supercomputing Center and the National Institute of Standards and Technology. The vision of the RCSB is to create a resource based on the most modern technology that facilitates the use and analysis of structural data and thus creates an enabling resource for biological research. Its mission is to provide the most accurate, well-annotated data in the most timely and efficient way possible to facilitate new discoveries and advances in science.

The challenges that the PDB is addressing include the continuing increase in the number and complexity of structures, the need to develop new methods for structure determination, satisfying user demands for response to complex queries and better annotation, integrating PDB data with other genomic and proteomic information, and serving a growing and more diverse community of users. The PDB's strategy for meeting these challenges involves the adoption of new technologies, creating extensible and portable data systems, making the archive as uniform as possible, improving communication with the users, and helping to create and enforce community policies and standards.

The Safeguarding and Sharing of Traditional Chinese Medicine Database Resources[6]

A big effort to develop traditional Chinese medicine database resources began in the 1980s. Since that time, nearly one hundred such databases of various sizes have been constructed by numerous universities, colleges, and

---

[6]Based on a presentation by Baoyan Liu and Meng Cui, China Academy of Traditional Chinese Medicine.

institutes. The digitization of information in traditional Chinese medicine information has been completed on a preliminary basis, including modern and ancient literature databases, factual and structural databases, and data warehouses.

The main organizations involved in the modern literature and factual databases are the Traditional Chinese Medical literature center and branch center, which are affiliated with the State Administration of Traditional Chinese Medicine. The national structural databases are being developed now. In 2001, the scientific experiments information database of traditional Chinese medicine was initiated using data warehouse technology and a virtual research center platform, which is now operational. The ancient literature database contains various e-books.

The development of traditional Chinese medicine databases already has made significant achievements and established a basis for the digitization of traditional Chinese medical information at the national level. Because this scientific and technical literature, and the underlying data and scientific experiments, are the result of fundamental research work for the public welfare, it needs the public support and funding of the government to guarantee its continuation. If this initiative obtains the support of the government, the main database can be completed quickly and provide a sharing mechanism as a public good.

Open Access to Scientific Data on Biological Diversity: An Urgent Need for China[7]

Biological diversity (also called biodiversity) is generally divided into three categories: genetic-level diversity, species-level diversity, and ecosystem-level diversity. The focus here is on data in the latter two categories in China.

China's species-level biodiversity is immense. For example, it contains about 30,000 different plant species, and nearly 500 mammal species. In fact, it has been estimated that the 17 so-called "megadiverse" countries, which include China, contain 70 percent of the world's species of plants and animals within their borders. China has databases for some of its species-level biodiversity. For example, the State Environmental Protection

[7]Based on a presentation by James Edwards, Global Biodiversity Information Facility, Denmark, available at http://www7.nationalacademies.org/usnc-codata/JamesEdwardsPresentation.ppt.

Agency of China lists 67 databases of plants, animals, and microorganisms.[8] However, these are mostly small databases, except for birds and plants, and most do not have georeferenced data or follow international standards. Moreover, unlike some of the other megadiverse countries—for example Costa Rica and Mexico—China has not initiated systematic efforts to develop computerized databases about its biota, or to access the great wealth of biodiversity information about China that is contained in the world's natural history collections. As a result, China cannot currently use this considerable body of knowledge for informed decision making.

The Global Biodiversity Information Facility (GBIF) is an international consortium aimed at making the world's primary biodiversity data freely and openly available over the Internet to benefit society, science, and a sustainable future. Begun in 2001, GBIF's members as of June 2004 include 41 countries and 24 international organizations, each of which agrees to set up a computer node to share primary biodiversity data. Control of the data, including the decision on what information to make available, resides with the data providers in each country or organization. GBIF's role is to aid the data providers in setting up their databases and to provide a portal that allows users to search all the databases at once.[9] GBIF is thus a network of participant nodes and other partners that agree to use common standards for data and metadata, encourage the generation and contribution of additional data and information for the network, and assure that data providers retain control of their own data.

As of June 2004, the GBIF data portal (initiated only a few months earlier) is serving nearly 24 million records containing information about specimens in natural history collections, as well as observational data. These records are being served by 63 data providers from around the world. Even though China is not yet a member of GBIF, the portal already contains approximately 45,000 records of plants and animals, representing more than 9,000 species that were collected in China.

The data being served by GBIF can be a valuable resource for many scientific and societal problems, including tracking invasive species, predicting the spread of emerging infectious diseases, optimal design of protected areas, and making decisions about where to undertake field trials of genetically modified crops. Other innovative examples of how megadiverse countries have used these data are being compiled by GBIF.

---

[8]See http://www.zhb.gov.cn/english/.

[9]See http://www.gbif.net.

China has been much more successful at developing and archiving ecosystem-level biodiversity data than species-level data. The Chinese Ecosystem Research Network (CERN)[10] is a consortium of 33 field research stations and one synthesis center. CERN was established in 1988, and currently provides access to a wide range of ecological and environmental data, including more than 3,000 historical datasets.[11] CERN has also developed a comprehensive data sharing policy and joined the International Long-Term Ecological Research Network. China should be encouraged to give similar attention to developing and archiving its species-level biodiversity information.

The NIH Roadmap for Medical Research[12]

The U.S. National Institutes of Health (NIH) launched the Roadmap for Medical Research initiative in 2003. It is focused on important public health challenges such as acute and chronic medical conditions, an aging population, health disparities in society, emerging diseases, and biodefense concerns. The roadmap provides a framework of priorities and a vision for a more efficient, innovative, and productive research system. It also establishes a set of initiatives that are central to improving the quality of healthy life for people in the United States and around the world.

One research priority of the NIH roadmap is the reengineering of the clinical research enterprise. This reengineering effort has multiple components, including integration of a clinical research network and facilitating data mining to advance data sharing goals. Examples of health databases and technologies that are part of this NIH focus are the lung image database, a network for translational research for optical imaging, a biomedical informatics research network, bioinformatics roadmap centers, and an insight segmentation and registration toolkit in support of the VHP.

The National Institute of Biomedical Imaging and Bioengineering, a component of the NIH, is addressing specific issues in the clinical research network initiative. There are a number of barriers to creating a successful

[10]See "The Data Sharing Policy of the Chinese Ecosystem Research Network" in Chapter 4 of this report.

[11]See http://www.cern.ac.cn:8080/index.jsp.

[12]Based on a presentation by Belinda Seto, National Institute of Biomedical Imaging and Bioengineering, U.S. National Institutes of Health, available at http://www7.nationalacademies.org/usnc-codata/BelindaSetoPresentation7.ppt.

network, which can include fundamental differences in informatics infrastructure and communication tools used at various research sites. To the extent that interoperability can be implemented and data and tools shared, studies can be initiated more quickly. From the perspective of the Institute, there is a need to create imaging databases and repositories where researchers can access such data. However, access to databases and data mining also requires user-friendly informatics tools. Approaches that combine images, genomic, gene expression, and patient medical records data will ultimately deliver patient-specific information at a time and place where clinical decisions are made regarding risk, diagnosis, treatment, and follow-up.

The overall implementation strategy involves the development and standardized validation of application-specific software for integration and knowledge extraction of heterogeneous, clinically relevant data. Specific functions that need to be addressed include:

➢ Quantitative data integration, knowledge extraction, and clinical interpretation;
➢ Linking imaging and other databases with software tools;
➢ Managing software in the scientific and clinical workflow;
➢ Partnerships between industry and academia for software development and dissemination;
➢ Database development specifically for software validation and regulatory approval; and
➢ Standards related to interoperability of imaging and other databases, and including results of quantitative analysis of metadata.

EXAMPLES OF EARTH SCIENCES, ENVIRONMENTAL, AND NATURAL RESOURCES DATA ACTIVITIES

Progress in Meteorological Data Sharing in China[13]

Meteorological data are a vast resource that applies to many fields. Such data are indispensable for economic and social development, scientific and technological innovation, and general human welfare. The collec-

[13]Based on a presentation by Dahe Qin, China Meteorological Administration, available at http://www7.nationalacademies.org/usnc-codata/Qin_Dahe_Presentation.ppt.

tion of meteorological data by the China Meteorological Administration (CMA) is supported operationally by polar-orbiting and geostationary satellites; a ground-based observation network consisting of various types of sensors; the synthesized operational transmitting system, which covers the entire world; a supercomputer; and many other facilities and equipment. The total amount of archived data is over 100 terabytes.

As a member of the World Meteorological Organization, the CMA has cooperated extensively with the international community in the exchange of meteorological data products. In December 2001, under the support of the Ministry of Science and Technology (MoST), the CMA initiated the meteorological data sharing services program as part of the China SDSP.

Developments in meteorological data sharing over the past three years include the integration of data resources, the compilation of technical standards, and the construction of service systems. The project has provided online and in-house services for scientific research, education, national constructing projects, and the public. Its data services are based on policy, technical standards, an operational data management system, and the classification of data and users. The work and experience of this project can be used as a reference in developing the China-SDSP and its other projects.

The next step is for the meteorological data sharing project to take a leading role in the China-SDSP. To accomplish this, it must:

➢ Increase the capability of its service functions, and provide more and better services for national economic development and social welfare;

➢ Build upon the advances in science and technology, improve data security, and provide high-quality data products for sharing;

➢ Optimize the allocation of resources in the climate system and establish a "climate system" data sharing platform; and

➢ Develop extensive cooperation with domestic sectors and with international meteorological Web sites, institutes, and organizations.

The project is expected to increase the sharing of meteorological data and promote the proper configuration and effective utilization of national information resources. The project also plans to quicken the implementation of the Chinese Climate Observation System, advance climate system data sharing as part of the Global Climate Observation System, and help in the enactment of the new national data sharing law and policy.

The World Data Center for Renewable Resources and Environment[14]

The World Data Center (WDC) system was established by ICSU in 1957. It consists of 52 discipline centers, which are distributed in the United States, Russia, Europe, Japan, India, Australia, and China. China joined the WDC system in 1988 and it now has nine discipline centers (Astronomy, Geology, Geophysics, Glaciology and Geocryology, Meteorology, Oceanography, Renewable Resources and Environment, Seismology, and Space Sciences).

The WDC for Renewable Resources and Environment (WDC-RRE) is maintained by the Global Change Information and Research Center at the Institute of Geographic Sciences and Natural Resources Research in the Chinese Academy of Sciences. The mission of the WDC-RRE is to cooperate actively with ICSU to promote exchange and sharing of data in the fields of natural resources and the environment. The WDC-RRE attaches great importance not only to data collection, but also to data exchange and services to users. It seeks to play an important role in supporting scientific research, public decision making, scientific popularization, personnel training, and international cooperation.

The WDC-RRE is funded by the Chinese Academy of Sciences and MoST. The WDC-RRE carries out its work under the direction of the head of the Global Change Information and Research Center, and is engaged in the following activities:

➢ Researching the present situation regarding renewable resources and environmental data inside and outside China;
➢ Investigating user requirements for such data;
➢ Working out a metadata standard for WDCs in China;
➢ Establishing the Web site of the WDC for RRE;[15] and
➢ Producing data and providing data services.

---

[14]Based on a presentation by Shunbao Liao, Geosciences and Natural Resources Institute, Chinese Academy of Sciences, available at http://www7.nationalacademies.org/usnc-codata/Liao_Shunbao_Presentation.ppt.

[15]See http://eng.wdc.cn:8080/Metadata/index.jsp.

Information System for Earth Science Data of China[16]

The Information System for Earth Science Data is a project initiated in 2002 by the Scientific Information Center for Resources and Environment (SICRE) of the Chinese Academy of Sciences. Since that time, SICRE has established regulations and policies for the collection, processing, management, and sharing of earth science data. The draft of the Chinese metadata requirements for this system is compatible with international approaches and is based on a synthesis of the main metadata criteria for earth science data. The system has begun to provide data and information services online.

The China-SDSP will be a great benefit to the researchers and the scientific activities in China. Presently, more detailed polices are being established and the common understanding of data sharing is developing. One challenge is that the standardization and digitization of earth science data is complicated, and there is a great demand for it.

This tension between the supply and demand for these data will be mitigated once the Clearinghouse of Earth Science Data is established. The Clearinghouse will reproduce and organize the primary data with lower costs and less time. With such a "top-down" process, the actualization of China's scientific data sharing can be realized in all respects. So far, an effective data management and information catalogue has been completed and the data service function is available for researchers and decision makers. It is with these goals in mind that the Clearinghouse of Earth Science Data in China has been launched.

Present Status and Future Development Strategy of China's Sustainable Development Information Network[17]

China's Sustainable Development Information Network (CSDIN) was created by the Administrative Center for China's Agenda 21 and nine academic institutions and scientific organizations in 1997 with the support of the Chinese MoST. The main goal of CSDIN is to provide data and infor-

[16]Based on a presentation by Jiansheng Qu, Scientific Information Center for Resources and Environment, Chinese Academy of Sciences.

[17]Based on a presentation by Xiaofeng Fu, Administrative Centre for China's Agenda 21, Ministry of Science and Technology, China and Xintong Li, State Key Laboratory of Resources and Environment Information System, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences.

mation for research, management, and decision-making related to sustainable development strategies that are being implemented in China. A related goal is to improve the public awareness of sustainable development issues through information dissemination, and to achieve the objectives for long-term development that were established in China's Agenda 21 program.

The data and information in the CSDIN system are focused on natural resources, biodiversity, forestry, agriculture, macroeconomics, environmental protection, environmental technology transfer, and natural disasters, which together may be called sustainable development information. CSDIN's multidisciplinary data have been integrated and analyzed for implementing sustainable development strategies implemented in China, such as national or regional sustainable development capacity evaluation methodologies and experiences. As of June 2004, there were 50 gigabytes of data and 17 databases.

The formation process and outcomes of the data sharing policy, and the standards for promoting information sharing in China are key aspects of CSDIN. The policy regulates CSDIN's data access and Internet dissemination activities according to data pricing, user status (e.g., government sector, the general public, and private enterprises), and data use. For the data standards, certain metadata and data dictionary standards, a geo-grid standard, and a data classification and encoding standard have been proposed. These standards will be used in the whole country. These CSDIN standards provide important application experiences for the Chinese national e-government and scientific data sharing program. Other major elements for the development of CSDIN include the technology for the legacy system's reconstruction, the architecture of the data warehouse, the development of a geospatial database, and technologies for geo-information services.

The future development of CSDIN will focus on geo-information service standards and technical specifications; implementation of a more user-friendly interface; enabling interoperable spatial, thematic, and temporal geo-information services; and building a sustainable development decision support system.

Progress Toward a National Spatial Data Infrastructure in China[18]

As society is becoming more and more information dependent, multi-scale digital spatial data are urgently needed by a variety of users for supporting their planning, monitoring, management, and decision making in

[18]Based on a presentation by Jun Chen, National Geomatics Center, China.

a broad range of applications. In this context, China's National Spatial Data Infrastructure (NSDI) is providing the digital geospatial framework. It consists of vertically and horizontally integrated geospatial databases and communication networks, as well as necessary institutional arrangements for effective flow and exchange of geospatial information.

There are four main components of China's NSDI—data sets, the clearinghouse, policy and standards, and institutional arrangements. There are multiscale fundamental geospatial databases at the national, provincial, and municipal levels. For example, 1:1,000,000 and 1:250,000 scale databases have already been completed, and a 1:50,000 scale database is expected to be developed. Finer-scale databases are being produced at the provincial and local levels.

Several key factors influence the sharing of geospatial data. These include the data resources that are available, the policies that apply to these activities, and technical and institutional coordination. User needs also must be considered. Administrative licensing regulations in 1999 created three pricing categories for data users: government agencies can obtain data freely, not-for-profit institutions at 10 percent of full commercial cost, and corporations at full commercial prices. E-government initiatives are now driving a variety of applications that require greater data integration.

Uses of Seismic Data and the Importance of Open Access to Major Data Centers in Seismology[19]

Earthquakes originate at places where stress levels in the Earth have become too high. Rocks rupture, and slip occurs on a fault surface in order to release stress. The resulting seismic waves spread throughout the Earth's interior. Eventually they reach the Earth's surface where they may be recorded.

Every day, there are about one or two hundred earthquakes, large enough for their seismic signals to be recorded more than 1,000 kilometers from the earthquake source. Also every day, some earthquakes are large enough for their signals to be recorded all over the world. The only way to study such earthquakes effectively is to work with data recorded by seismographic stations in different countries.

[19]Based on a presentation by Paul Richards, Columbia University, United States, available at http://www7.nationalacademies.org/usnc-codata/PaulRichardsPresentation.pdf.

Scientists and engineers use seismic signals to make earthquake catalogs and bulletins, which provide basic information for various kinds of research and applications, including the study of earthquake hazards, the physics of the earthquake sources, and the structure of the Earth's interior. The great progress in seismology in all these fields has been stimulated principally by the availability of improved and improving data.

Both China and the United States are developing major new networks of seismometers at a cost of hundreds of millions of U.S. dollars to each country. The experience in the United States indicates that data centers that do not have open data policies are rarely able to attract researchers who apply state-of-the-art methods of data analysis. Such data centers consequently find it difficult to maintain high-quality operations. The best research is usually associated with centers that make their data openly and easily available. For example, users draw attention to errors that inevitably arise in the data, they find ways to correct the data, they share information about how to use the data center effectively, and they contribute to new ways to process the data. Information from users of the data thus is needed to provide guidance for the management of a data center. From this perspective, it appears that an important part of providing international scientific leadership in seismology is making seismic data easily available to all interested potential users.

China too has excellent data sets of seismic waveforms, which will yield new insights into earthquake physics, tectonics, and the Earth's internal structure. New methods of locating earthquakes have recently been applied to limited datasets. They indicate the potential for China to produce one of the best bulletins of seismicity in the world, covering a large region (more than 10,000,000 square km). Bulletins are a starting point for hazard management, as well as for scientific projects in the study of the Earth's structure and earthquake physics.

Presently there are handicaps, however, in that station coordinates are not made easily available, and waveform data are accessible for only a very limited number of stations. For many years, China has not allowed even the locations of seismographic stations to be known to western scientists, except for a network of 24 stations. A consequence of these restrictions is that the locations of earthquakes in China are not as well known as they would be if more data were made generally available.

Existing Infrastructure for International Exchange of Seismic Data[20]

As noted in the preceding section, knowledge of earthquake hazards can advance as a result of unrestricted sharing of seismic data, including seismic station information, bulletins, and waveforms. The infrastructure to arrange for and carry out international data exchange has existed and been used successfully for many years. The International Association for Seismology and the Physics of the Earth's Interior (IASPEI) includes commissions to discuss specific arrangements and to establish data format standards. The International Seismological Centre (ISC) collects, merges, and redistributes seismic bulletin data. The Federation of Digital Seismic Networks (FDSN) helps broadband seismic networks to coordinate their activities. The United States and China help to fund and participate in each of these organizations, but the amount of data that the Chinese have shared lags far behind most other countries with similarly extensive earthquake monitoring.

It is useful to have data from more stations for a number of reasons in addition to the ones outlined in the preceding section. Seismology is advancing beyond computing "formal" estimates of uncertainty that are based on questionable statistical assumptions. To estimate "absolute" location uncertainty, seismologists must calibrate new techniques using very accurate locations. Many seismologists accept a location computed by a local network as absolutely accurate within 5 kilometers only if it is computed from arrival times at stations at least one of which is within 30 kilometers of the earthquake; at least ten of which are within 250 kilometers of the earthquake; and with a very good distribution around the earthquake.

Local networks may be reconfigured frequently to meet new needs, so using their data depends on detailed knowledge of the network. Thus, seismologists recognize that there may be advantages from arranging some exchanges of local network data individually.

Regional networks cover a broader area but with somewhat more sparsely distributed stations. Regional networks usually have a more standardized configuration, so data from them can be used with greater confidence by more seismologists. Many seismologists accept a location computed by a regional network as absolutely accurate within 20 kilometers if it is computed from arrival times at seismic stations that are all within

[20]Based on a presentation by Raymond J. Willemann, GEM Technologies, United States, available at http://www7.nationalacademies.org/usnc-codata/RaymondWillemann Presentation.ppt.

1,000 kilometers of the earthquake and have very good distribution around the earthquake.

Achieving good distribution around an earthquake usually requires at least ten seismic stations. Because China is so large, several hundred stations are needed to obtain a good distribution of regional stations around all of the earthquakes in China.

Two recent changes offer an opportunity to improve the international contributions of seismic data from China. First, IASPEI passed at its 2003 General Assembly in Sapporo, Japan, a resolution urging all seismic networks to share information about all seismic stations. The resolution states:

> RECOGNIZING the need to accurately locate earthquakes and determine earthquake size, and compile complete earthquake bulletins,
>
> URGES all operators of seismic stations and networks to deposit unique station codes with the international registry maintained by the International Seismological Centre and by the World Data Centre for Seismology, Denver, and to freely share the coordinates of all seismic stations,
>
> URGES all operators of seismic stations and networks to keep accurate record of instrument response and performance.

China might respond to this IASPEI resolution by starting to send the ISC a computer-readable bulletin from the China Earthquake Administration that is complete for a network of several hundred seismic stations in China. Second, the FDSN has broadened the definition of membership to include many more regional and local networks of broadband seismic stations. In response, numerous provincial networks in China might join the FDSN. This would provide the networks an opportunity to contribute data from selected stations to the FDSN archive and access to software and other assistance from the FDSN to establish their own data centers to distribute their own data on the Internet.

Digital Fujian[21]

The digital Earth has the allure of a diamond; people from different angles all can appreciate its prism and reflection. Consistent with the con-

[21]Based on a presentation by Qinmin Wang, Department of Science and Technology, Fujian Province, available at http://www7.nationalacademies.org/usnc-codata/Wang_qinmin_Presentation.ppt.

cept of digital Earth, thinking globally and acting locally has become the approach for organizing the regional e-government information sharing program in Fujian Province. The governor of Fujian Province formed the Digital Fujian program to promote regional modernization during the 2001-2005 time period. The program includes the establishment of an information network, the creation of data resources (a data warehouse with metadata, data processing, modeling, and other related functions), the formation of data standards and a data sharing policy, the training of qualified technicians, and the funding of an information applications infrastructure.

After the three years of development, the Digital Fujian program has established successfully the e-government information networking and data sharing platform between the provincial government and the city and county government departments. It now includes approximately one terabyte of standardized data resources among 21 government agencies, 9 information application systems, 20 government information application projects, a provincial information technology technician training center, and a set of information sharing standards and regulations. Another key technical issue of the program is how to extract useful information from the huge Earth observation satellite databases. Therefore, data mining, data analysis, data integration, information extraction, information presentation, and intelligent decision making are also important parts of the program.

Local and Regional Earth System Science Applications and Associated Infrastructure: The Mid-Atlantic Geospatial Information Consortium[22]

The Mid-Atlantic Geospatial Information Consortium (MAGIC) is a federated consortium of universities in that region of the United States. Their mission is to develop a distributed remote sensing, applications, geospatial data and information system, serving a variety of users at the local, state, and regional levels. It extends wide usage of National Aeronautics and Space Administration (NASA) data. It focuses on the use of such data for NASA priorities and on the dissemination of such data through interoperable information systems that are coupled to NASA's

[22]Based on a presentation by Menas Kafatos, George Mason University, United States, available at http://www7.nationalacademies.org/usnc-codata/MenasKafatosPresentation.ppt.

systems and that promote open solutions and standards. It prototypes applications for regional effects of climate phenomena and change, land use, pollution mitigation, agriculture, health, wetlands, and forestry, coupling to NASA priorities.

There are many factors related to providing open data access that must be considered by MAGIC, including the following:

➢ *A federated data and information management system.* Support depends on individual partners and changes are not reflected at all sites.

➢ *Standards.* MAGIC uses standards developed by the Federal Geographic Data Committee, WebGIS (Geographic Information Systems), and the Web Mapping Service.

➢ *Diversity of users.* MAGIC serves users at the local, state, regional, and federal levels, and from the private and public sectors. These users have different needs, different expertise, and different infrastructure.

➢ *Large number of data types.* There are over 2,000 databases with many different types of data (e.g., GIS, swath, gridded, socioeconomic). These data sources require different tools and metadata to make them useful and interoperable with federal government agencies and across all other MAGIC sources. Help desks are needed for user support.

➢ *Infrastructure.* High-speed network access is required, but not all partners have the same access capabilities.

➢ *Funding.* The NASA Earth System Science Applications program provided initial funding. Funding priorities change, however, and the users are a mix of public- and private-sector entities. Ultimately, the funding sources should reflect this diversity of users and include the state government and industry.

➢ *Training.* Remote-sensing and other geospatial data are difficult to use. Users therefore frequently require training, including in the application of geographic information systems to remote-sensing data.

➢ *Free versus proprietary data.* NASA Earth System Science data are freely and openly available. Other government (e.g., from the U.S. Geological Survey and the National Oceanic and Atmospheric Administration) data are low-cost and unrestricted. Local and state data are not free, and commercial private-sector data are proprietary. An important question is, how can a federated, Web-based system make access free to different users? This key question is still being addressed.

THEMATIC ISSUES IN SCIENTIFIC INFORMATION, JOURNALS, AND DIGITAL LIBRARIES

Policies and Mechanisms for Literature Resource Sharing—The Practice of the Chinese National Scientific and Technical Library[23]

Information resource sharing benefits society and the economy by the spread of knowledge at a low cost. Intellectual property rights (IPRs) stimulate and protect creative innovation, but restrict the public diffusion and application of knowledge. They are thus positive on the one hand, but negative on the other. Enhancing the resource sharing requirement may limit the interests and benefits of the IPR owner, whereas increasing the IPRs of the owner may limit the benefits that society could gain from resource sharing. A proper balance between the two needs to be found to maximize the social benefits. The policy and legal system is facing an increasing challenge in reconciling these opposing effects.

Information resource sharing is a powerful instrument for reducing the digital divide, but the threshold for doing it is quite high due to the price of literature. At the same time, the strengthening of IPRs is the current trend in international law and trade. This is related with the fact that the main exporters of knowledge and information products are in developed countries. For developing economies, the level of IPR protection should be synchronized with the level of development of the economic and legal system. Setting the protection standards too high can be harmful. There are choices and tradeoffs in enacting these policies.

If the rigorous application of copyright law requires payment for every drop of knowledge, then the capacity of public libraries will be very limited by the shortage of funds. The urgent problem is: How to observe international IPR regulations while at the same time improving the scientific information resource capacity building and meeting the demands of social progress?

The Practice of the Chinese National Scientific and Technical Library of China

The National Scientific and Technical Library (NSTL) of China is a virtual scientific literature service center initiated in June 2000 by the MoST

[23]Based on a keynote presentation by Qiheng Hu, Vice President, Chinese Association for Science and Technology, available at http://www7.nationalacademies.org/usnc-codata/Hu_Qiheng_Presentation.ppt.

and the Ministry of Finance in cooperation with five other ministries, and approved by the State Council. The NSTL is composed of seven member libraries: the Center for Literature and Information of the Chinese Academy of Sciences, the China S&T Information Institute, the Machinery Industry Information Institute, the Metallurgy Industry Information Standard Institute, the Chemical Industry Information Center, the Literature Center of the Chinese Academy for Agricultural Science, and the Information Center of the Chinese Academy for Medical Science. The NSTL office is in charge of coordinating and managing the services.

The principles for operation of the NSTL are: "unified purchasing, normalized processing, combined networking, and resource sharing." The main goals are to:

➢ Build and share the scientific literature and information resources among all members through a convenient network so that this virtual library can provide better service to the research community in the areas of basic science, engineering, agriculture, and medicine;

➢ Develop a high-level scientific literature collection and service center;

➢ Demonstrate effective applications of information technologies in scientific literature and information services;

➢ Become a pivotal force in cooperation with the broader Chinese library system and the leader in the scientific library system of China;

➢Play a major role in exchanges with the international library community; and

➢ Establish the information resource base for research, training, and the popularization of scientific education.

The responsibilities of NSTL are diverse, consistent with its main goals. These responsibilities include: planning and funding of comparatively complete literature resource acquisitions; coordination of the literature collection to avoid unnecessary redundancies; formulation of standards, norms, and formats for the unified database; providing service to users throughout the country with the NSTL network service platform; development and application of in-depth resources; and domestic and international exchanges and collaborations.

The responsibilities of the NSTL are directed by a 19-member Council, which is the decision-making and leading body, and also represents NSTL members and users. The NSTL Director is appointed by the Council. The Director General is in charge of operations and hires the office staff. The Council and the Director are advised by two advisory groups, the

Expert Advisory Committee for Literature Resource Construction and the Expert Advisory Committee on Computer-Network Construction.

The NSTL has made significant progress in several areas, three of which are highlighted here. First, it has increased the total number and variety of the scientific literature resources in China. The investments in scientific literature have been used more effectively since the NSTL member libraries stopped the counterproductive duplication and competition in purchasing international scientific periodicals as a result of NSTL planning. Prior to 2000, there were only about 3,000 English-language scientific journals and periodicals in the NSTL member libraries. The NSTL increased the total number to 10,653 in 2000. By 2003, the NSTL holdings of journals and periodicals in English numbered approximately 13,500 with an additional 5,000 conference proceedings and handbooks in English. Also in English are 15 kinds of network edition periodicals and 65 kinds (physics and chemistry) of network edition subscriptions with access controls. The NSTL also has a rich variety of literature in the Chinese language, including more than 4,000 kinds of journals and periodicals, over 22,000 conference proceedings, and 470,000 theses for Masters and PhD degrees.

Second, NSTL has constructed a 1 gigabit/second broadband network linking the NSTL's seven member libraries, thereby making the separate libraries a united library with digitized resources in the networked environment. The NSTL service network is also linked with the National Library, the China Education and Research Network, and the China Scientific and Technical Network via 100 megabit/second connections.

The third major area of NSTL's progress is in developing the rapid growth of online digitized resources and related services. In 2000, before the launch of the NSTL network, there were 1.7 million information items of online data. This increased to 27 million items in 39 databases by December 2003. The NSTL provides many services including: literature searches, full-text provision, periodicals cataloging, common directory query, full-text database access for the network-edition literature, literature directory database searching, expert consulting and information services, online resources search engine, preprints system in Chinese, and a portal for preprints in English. Services that are provided to users via the Internet include free literature searching and online payment for services (around 44 percent of payable services are paid online). Requests for full-text provision are processed within two working days.

All users can access and download for free the 15 network-edition periodicals in English for which NSTL buys access. The other 65 kinds of periodicals on physics and chemistry are available for 80 universities and research institutes according to the subscription license, based on access controls. Over 99 percent of the periodicals in English are accessed each year.

Despite these accomplishments, the NSTL, like all public libraries, continues to face a serious problem. The benefit of using information technology to decrease the costs of knowledge sharing is greatly counteracted by the rising costs of the imported information resources. From the experience of the NSTL, the network edition's cost is typically very high and is proportionate to the scale of information sharing. There may be no other choice than to go back to printed editions for such high-cost imported information. Does this sound like progress in the age of digital networks?

Perspectives on the Future of the Library and on the Economics of Open Access[24]

The big question facing the circulation of scientific information is: What is going to become of the emerging open-access movement? There are growing numbers of arguments in support of increasing access to research through a variety of open-access models. The case for increasing access has components that can be roughly categorized as epistemological, historical, developmental, political, and public—as well as economic and legal, which are discussed elsewhere—and is directed primarily toward faculty members, students, librarians, policy makers, and the public.

The epistemological argument for open access, for example, has to do with how dependent a knowledge claim is on being fully open to review and critique. Anything that unduly restricts the circulation of knowledge, especially among "legitimate" participants in its construction, reduces that body of knowledge's claims to validity and reliability. If the current subscription publishing model can be shown to contribute to suboptimal levels of access, then those models are not what we might call epistemologically conducive to the development of knowledge.

The concern with exploring new publishing models leads to the historical argument, which draws on precedents from an earlier era of publishing innovation, using Isaac Newton as a leading instance. Newton is well known for being a highly secretive scientist and a reluctant author. Neverthe-

[24]Based on a presentation by John Willinsky, University of British Columbia, Canada.

less, after he had tracked for only a few years the birth and emergence of the scientific periodical, with the launching of *Philosophical Transactions* in 1655, he understood that this publicly circulated, relatively inexpensive 16-page journal represented something important to science. He allowed one of his letters, on optics, to be published in that journal. It was a move he came to regret and did not do again, but this early experience in open access went a long way in shaping what became the norms of the scientific article.

The developmental (or developing countries) argument for increasing access has everything to do with the parallel development in China of both economic growth and scientific papers published, with an increase by a factor of 10 since the 1980s. Developing countries are suffering a knowledge gap, even as their university population grows. The universities in the West are contributing to that gap, as their work becomes increasingly expensive to access (with some generous exceptions negotiated with publishers by the International Network for the Availability of Scientific Publications and some other organizations).

The political and public arguments for open access to research are about people's basic right to know, especially in matters of publicly funded research and scholarship. The value of exercising that right is affirmed by the health revolution brought about by public access to medical information. Public access to research also speaks to greater accountability demanded of professionals (e.g., physicians, educators, lawyers) and the increasing role of interest groups in selectively presenting information to the public, against which full access to the research would act as a safeguard.

An Open-Access Future[25]

The open-access movement has gained momentum over the past several years, with increased visibility and recognition from the various stakeholder communities, including research and publishing communities. Since the Budapest Open Access Initiative[26] began collecting signatures in February 2002, more than 3,500 individuals and organizations have signed on with their support for free access to information. The Directory of Open Access Journals[27] at Lund University, which contained over 1,100 journals

[25]Based on a presentation by Helen Doyle, Public Library of Science, United States.
[26]See http://www.soros.org/openaccess/.
[27]See http://www.doaj.org/.

as of June 2004, has announced the launch of its second phase, allowing sophisticated searching of the full text of the Directory's articles. The U.K.-based open-access publisher BioMedCentral now publishes over 100 open-access journals. Since its launch in October 2003, *PLoS Biology*[28]—the first peer-reviewed, open-access journal of the Public Library of Science (PLoS)—is demonstrating remarkable strength as a competitive new journal: submissions are increasing; readership, measured as visits to the site and as downloads of individual articles, is growing; and *PLoS Biology*'s reputation as a high-quality, peer-reviewed journal is improving among scientists, publishers, librarians, and other stakeholder groups.

In addition to a transformation of the economics of scientific publishing, open-access publishing also represents a modernization of traditional copyright laws that are based on an outdated print-based economic model. In the open-access definitions used in both the Bethesda Principles[29] and the Berlin Declaration,[30] an open-access article can be reused and redistributed freely and without permission from the publisher, for any responsible purpose. Authors retain their copyright. In the case of PLoS journals, the copyright license is the Creative Commons[31] "attribution license," which preserves the author's right to be acknowledged for the original work.

Many journals that are labeled open access are in fact free access, meaning that the restrictions on use and distribution are the same as for many subscription-based journals. It is worth noting that researchers themselves virtually never benefit financially from publication of their peer-reviewed articles. Several recent policies that may appear to be a liberalization of subscription policies are in fact small concessions to the growing demand for greater access from the scientific community that produces the articles, concessions made at little economic risk to the publishers.

The sharing of data, reagents, and ideas is fundamental to the scientific process itself. Open-access publishing, including both the unfettered distribution and searching afforded by online free access and the unlimited creative reuses permitted by less restrictive copyright licenses, will facilitate the advance of science and medicine.

---

[28]See http://biology.plosjournals.org/perlserv/?request=index-html&issn=1545-7885/.

[29]See http://www.biomedcentral.com/openaccess/bethesda/.

[30]See http://www.zim.mpg.de/openaccess-berlin/berlindeclaration.html.

[31]See http://creativecommons.org/.

Other Opportunities in the Changing Information Environment[32]

It has been recognized that research and access to research results play a vital role in the development of all countries, and that transitional and developing countries often cannot obtain and make use of research that would benefit them. In response to this "information divide," the International Network for the Availability of Scientific Publications (INASP) was established in 1992 as a program of ICSU to provide support to networking and partnerships with the aim of bridging the increasing information divide between the developed and developing world. It now operates a range of programs to support access to information for researchers, health workers, and rural development experts. Access to research information is facilitated through its Programme for the Enhancement of Research Information (PERI). This program not only provides access to international research, but facilitates activities to support national publications to increase their visibility and long-term sustainability. Research information is provided in many different forms (e.g., datasets, publications), but the scholarly journal remains one of the prime vehicles for accrediting and disseminating research information. Of course, both national and international information is of importance to research. PERI provides support for (1) access to global information, (2) increased visibility for local publications, (3) training in the use and management of online information, (4) support for publishers and editors, and (5) research and networking support. INASP negotiates access to as many required resources as possible with content owners and publishers. The exact cost of each resource is related to the GDP of the country, and although many of the resources are available without cost as part of PERI, others are obtainable at up to 98 percent discount on the normal subscription rates. Nearly all of the resources are available on a countrywide license basis, meaning that anyone in an educational, research, or nonprofit environment is eligible to access them.

The Journals OnLine (JOL) project supports a methodology to enable national publications to have an online presence, to increase their visibility, and to promote communication with readers and authors. It has been particularly successful in Africa with the African Journals Online (AJOL) ser-

[32]Based on a presentation by Pippa Smart, International Network for the Availability of Scientific Publications, United Kingdom, available at http://www7.nationalacademies.org/usnc-codata/PippaSmartPresentation.ppt.

vice that now includes 184 journals, and also supports full-text, online publishing. This service has been operational since 1998 and was recently re-launched on a new platform. The new software enables individual publishers to load, edit, and correct their own content to further support development and use of online communication. The JOL methodology and software are available for other countries and regions to adopt for their own publications.

Of course, there continue to be many challenges for information access, and INASP is continually updating its activities to respond to requests from partners and to increase its capabilities to develop sustainable methodologies.

Scientific Information and Digital Libraries: Can Developing Countries Become Key Players in the Information Society?[33]

In most developing countries, politics rather than markets drive knowledge diffusion. Whether new technology is adopted or not by a country is largely determined by political will, its patterns of power and influence, and resource-allocation policies. Politics are thus a decisive variable driving the transition of developing countries to the information society.

In addition to the political hurdles in most developing countries, there are many other factors working against preservation of and open access to scientific information in developing countries. These factors may be characterized as institutional, economic, and social, although many of them intersect and are difficult to separate out.

Generally speaking, the institutional culture creates barriers to the creation and diffusion of knowledge. For example, government officials tend to believe that access to knowledge is automatic once connectivity to digital networks is ensured. There is no institutional framework for data preservation mechanisms (e.g., the demise of the South African Data Archive). Moreover, there is no legislative mandate that obliges researchers that are publicly funded to make available research findings publicly accessible. The lack of comprehensive strategies and policies for data management are a major barrier to open access in developing countries.

From an economic standpoint, a large percentage of the population in

[33]Based on a presentation by Lulama Makhubela, National Development Agency, South Africa, available at http://www7.nationalacademies.org/usnc-codata/Lulama Presentation.ppt.

developing countries currently falls below the poverty line. Income distribution remains highly unequal, and poverty and inequality continue to exhibit strong geographic and racial biases. Public funding for the development and connectivity of information and communication technologies (ICTs)—from both national and global sources—continues to be inadequate. The resulting lack of access to ICTs further restricts the potential use of even the information that is otherwise openly accessible online.

The social barriers are no less daunting. Human resources in the scientific and technical domains are not adequately developed. Moreover, there is a "brain drain" of highly trained scientists and professionals from the developing to the more developed countries. The organizational complexity of scientific communities and the esoteric language of scholarly journals create further roadblocks to the broad transfer of knowledge. Within the broader society, the reading environment is poorly developed. Only a small percentage of the population uses library facilities, and most rural areas do not have libraries. Educational gaps mean that access to information does not necessarily result in access to knowledge. The simple availability of information technologies has shown to be insufficient to guarantee proper diffusion of scientific knowledge across society. Achieving mastery of technological change in the economy and society remains elusive.

Some actions that may be considered to help ameliorate the barriers outlined above include the following:

➢ There must be a greater effort to develop human resources in research, especially in national tertiary education programs, and to connect research more effectively with productive sectors and the specific needs of society. The curricula for research professionals should include digital data and information management principles and techniques.

➢ The adoption of rigorous and more collaborative approaches in addressing the lack of leadership and professional management, which has resulted in the poor implementation of data preservation in many areas, is vital. This needs to be coupled with a greater effort on developing a better understanding of the value of data and information preservation for future access among working scientists.

➢ A strategic approach is also needed to leverage resources and maximize effectiveness by increased collaboration among developing countries, and to document and disseminate best practices. This goal can be promoted by forming data archiving groups at different geographic levels to overcome the isolation of individual data archivists and promote beneficial exchanges. This will help harness the human intellectual capacity with collective expertise in issues of data preservation and open access. Cooperation should be encouraged for richer quality and more impact, rather than merely for the sake of cooperation.

➢ At the broader social level it is important to promote the use of scientific knowledge and of information technologies among all population strata, as well as to reach out to the literate poor. Investments in technological innovations deserve high priority because in some cases they can overcome the constraints of low incomes and weak institutions.

In conclusion, the discourse on scientific data and information resources and on digital libraries needs to be understood in the political, economic, and social context of developing countries. The future global information society may be one of widespread and beneficial international collaboration, or one of highly stratified access to knowledge. In order for the first scenario to prevail, the external barriers to access to knowledge need to be reduced, and the perverse internal dynamics preventing many developing countries from joining the global scientific community as active participants must be changed. A search for effective strategies for preservation of and open access to scientific information in developing countries will remain utopian unless those countries themselves become actively integrated in the broader information society.

Appendixes

A

International Workshop on Strategies for Preservation of and Open Access to Scientific Data

International Conference Room
Century Golden Resources Hotel
Beijing, China
June 22-24, 2004

AGENDA*

Tuesday, 22 June

Session 1: Opening Remarks

Co-Chairs:

- ZHANG Xian'en, Director General, Department of Basic Research, Ministry of Science and Technology of China
- Shuichi IWATA, President, CODATA and Professor, University of Tokyo, Japan

9:00 Welcoming Remarks by Workshop Co-Chairs
CHENG Jinpei, Vice President, Ministry of Science and Technology of China
Roberta BALSTAD MILLER, Chair, U.S. National Committee for CODATA and Director, Center for International Earth Science Information Network (CIESIN), Columbia University, United States

*The workshop agenda presented Chinese names with family names first and in capital letters, and given names second.

9:15 Workshop Objectives
William ANDERSON, Principal, Praxis101 and Co-Chair, CODATA Task Group on Preservation of and Access to S&T Data in Developing Countries

9:30 Keynote Address—Strategies of China Scientific Data Sharing
CHENG Jinpei, Vice President, Ministry of Science and Technology of China

Rapporteurs:

- FU Xiaofeng, Director, Information Division, Administrative Center for China's Agenda 21 and Member, CODATA Task Group on Preservation of and Access to S&T Data in Developing Countries
- Steve ROSSOUW, Member, CODATA Executive Committee

10:00 Break and Pictures

Session 2: Overview of Thematic Areas in China

Co-Chairs:

- CHENG Jinpei, Vice Minister, Ministry of Science and Technology of China
- Michael CLEGG, Foreign Secretary, National Academy of Sciences, United States

10:30 The Chinese National Committee for CODATA: Its Mission and Future
XU Zhihong, Representative of the Chinese National Committee for CODATA

10:50 China's Scientific Data Sharing Program (China-SDSP)
ZHANG Xian'en, Director General of Department of Basic Research, Ministry of Science and Technology of China

11:20 Progress and Perspectives of Climate Data Sharing of China
QIN Dahe, President, China Meteorological Administration

11:45 The Management and Sharing System for Scientific Data in Medicine
LIU Depei, President, Chinese Academy of Medicine, and Vice President, Chinese Academy of Engineering

12:10 The Mechanism for and Management of the National S&T Library of China
HU Qiheng, Vice President, Chinese Association of Science and Technology of China

Rapporteurs:

- XU Feng, National Information Center of China
- John WILLINSKY, Professor, University of British Columbia, Canada

12:10 Lunch

Session 3: International Perspectives on Preservation of and Open Access to Public Scientific Information

Co-Chairs:

- SUN Honglie, Vice President of CODATA, Chinese Academy of Sciences
- Roberta BALSTAD MILLER, Chair, U.S. National Committee for CODATA and Director, CIESIN, Columbia University, United States

13:30 Trends in Development of International Scientific Data and Information
Yasuyuki AOSHIMA, Director, Beijing Office, United Nations Scientific, Educational, and Cultural Organization (UNESCO)

13:45 Towards International Guidelines for Access to Research Data from Public Funding
Peter SCHRÖDER, Ministry of Education, Culture, and Science, The Netherlands

14:00 International Perspectives on Data and Information for Science

Carthage SMITH, Deputy Executive Director, International Council for Science (ICSU)

14:15 Initiatives on Promoting Access to Scientific Information
Michael CLEGG, Inter-Academy Panel (IAP) Member and Foreign Secretary, National Academy of Sciences, United States

14:30 General Discussion

15:00 Break

Rapporteurs:

- WANG Zhengxing, Deputy Director, Global Change Information and Research Center, Institute of Geography and Natural Resources, Chinese Academy of Sciences
- Pippa SMART, Head of Publications, Publishing Initiatives, and Publishing Training, International Network for the Availability of Scientific Information (INASP), United Kingdom

Session 4: Policy and Management

Co-Chairs:

- SUN Shu, Director, Department of Earth Sciences and Director, Expert Committee of China Scientific Data Sharing Project, Chinese Academy of Sciences
- Theodore Carl BERGSTROM, Professor, University of California at Santa Barbara, United States

Policy and Legal Issues

15:20 Borders in Cyberspace: Maximizing Social and Economic Benefit from Public Investment in Data
Peter WEISS, J.D., U.S. National Weather Service

15:45 Reflections on the Legislation for National Scientific Data Sharing
ZHANG Jing'an, Director General, Department of Policy, Regulation, and Reform, Ministry of Science and Technology of China

Institutional and Management Issues

16:00 Managing the Impacts of Change on Archiving Research Data
Raymond McCORD, Oak Ridge National Laboratory, United States

Local and Regional Issues

16:20 Local and Regional Earth System Science Applications and Associated Infrastructure: The Mid-Atlantic Geospatial Information Consortium
Menas KAFATOS, Professor, George Mason University, United States

16:40 Digital Fujian
WANG Qinmin, Vice Chairman, Fujian Political Consultative and Director, Department of Science and Technology, Fujian Province, China

17:00 The Role of the CODATA Task Group on Preservation of and Access to S&T Data in Developing Countries
LIU Chuang, Co-Chair, CODATA Task Group on Preservation of and Access to of S&T Data in Developing Countries

Rapporteurs:

- YIN Ling, The PLA General Hospital, China
- Carthage SMITH, Deputy Director, ICSU

17:20 Preparation for Day Two and Announcements

17:30 Adjourn

18:10 Reception and Dinner

Wednesday, 23 June

8:30-12:15 Individual Presentations and Panel Discussions

Session 5: Breakout Panel Discussions on Cross-Disciplinary Issues

5-1: Policy and Legal Issues

Co-Chairs:

- CHEN Jun, Director, Basic Geographical Information Center of China
- Peter SCHRÖDER, Ministry of Education, Culture, and Science, The Netherlands

Panelists:

Maximizing Social and Economic Benefit from Public Investment in Data
Peter WEISS, J.D., U.S. National Weather Service

Progress and Policy Issues on Geo-Spatial Framework of Digital China
CHEN Jun, Director, Basic Geographical Information Center of China

Global Trends to Restrict Access to Data from Government-Funded Research
Jerome REICHMAN, Duke Law School, United States, and Paul UHLIR, U.S. National Academies

Data Sharing Policy of Chinese Ecosystem Research Network
CHEN Panqin, HUANG Tieqing, Deputy Director, Department of Natural Resources and Environment Management, Chinese Academy of Sciences

On Government Information Resources Sharing
ZENG Lan, LI Jun, National Macro Economic Research Institute

Balancing the General Public Interests and Copyright in Scientific Information Management
John WILLINSKY, University of British Columbia, Canada

Comparative Study on Strategy and Policy for Open Access to Scientific Data Among the United States, European Union, and China
LIU Chuang, Professor, Institute of Geography and Natural Resources, Chinese Academy of Sciences

Rapporteur

- Paul UHLIR, J.D., U.S. National Academies

12:25 Lunch

5-2: Institutional and Economic Issues

Co-Chairs:

- XIAO Yun, Secretary General, Chinese National Committee for CODATA
- Dora Ann CANHOS, Centro de Referência en Informação Ambiental (CRIA), Campinas, Brazil and Member, CODATA Archiving Task Group

Panelists:

Data Sharing Policy of the National Institutes of Health
Belinda SETO, Deputy Director, National Institute of Biomedical Imaging and Bioengineering, National Institutes of Health, United States

Data Sharing in Scientific Databases of the Chinese Academy of Sciences
XIAO Yun, Secretary General, Chinese National Committee for CODATA

Operating a Twenty-First-Century Data Center
Roberta BALSTAD MILLER, Director, CIESIN, Columbia University, United States

WDC for Renewable Resources and Environment
LIAO Shunbao, Associate Professor, Global Change Information and Research Center, Institute of Geography and Natural Resources, Chinese Academy of Sciences

The Peculiar Economics of Scientific Information
Theodore Carl BERGSTROM, Professor, University of California at Santa Barbara, United States

Rapporteur:

- Henda van der BERG, National Research Foundation, South Africa

5-3: Management and Technical Issues

Co-Chairs:
- LI Yixue, Director of Shanghai Biomedical Center
- Lulama MAKHUBELA, Research & Development Directorate, National Development Agency, South Africa

Panelists:
Information Infrastructure for Life Science Studies in Shanghai
LI Yixue, Director, Shanghai Biomedical Center

Managing the Impacts of Programmatic Scale and Enhancing Incentives for Data Archiving
Raymond McCORD, Oak Ridge National Laboratory, United States

Involving the Private Sector in the Environmental Enterprise
Anne LINN, Director, Committee on Geophysical and Environmental Data, U.S. National Academies

Earth Science Data and Information Management in Western China
SUN Chengquan, Director, Information Center for Natural Resources and Environment, Chinese Academy of Sciences

Data Integration and Management: A PDB Perspective
FENG Zukang, Protein Data Bank, United States

Launching a New Open-Access Journal
Helen DOYLE, Director, Development & Strategic Alliances, Public Library of Science, United States

Rapporteur:
- William ANDERSON, Praxis101, United States

Session 6: Summaries of Cross-Disciplinary Issues and Discussion

Co-Chairs:
- GUO Huadong, Associate Secretary General and Director, Department of International Cooperation, Chinese Academy of Sciences

- Khudulmur SODOV, Director, Information and Computer Center, National Remote Sensing Center, Mongolia

13:30 Cross-Disciplinary Rapporteur Reports

Policy and Legal Issues
Paul UHLIR, J.D., U.S. National Academies

Institutional and Economic Issues
XIAO Yun, Secretary General, Chinese National Committee for CODATA

Management and Technical Issues
LI Yixue, Director, Shanghai Biomedical Center, China

14:15 General Discussion of Cross-Disciplinary Issues

15:00 Break

15:30 The Future Role of CODATA
Prof. Shuichi IWATA, University of Tokyo, and President, CODATA

16:00 Open Journal Systems: A Live Demonstration of Open-Source Journal Management Software
John WILLINSKY, University of British Columbia, Canada

16:45 Summary Remarks
SUN Honglie, Vice President, CODATA

Rapporteurs:

- LI Zengyuan, Deputy Director, Institute of Natural Resources Information, Chinese Academy of Forestry
- Julie ESANU, Office of International S&T Information Program, U.S. National Academies

16:55 Preparation for Day Three and Announcements

17:00 Adjourn

Thursday, 24 June

Session 7: Breakout Panel Discussions on Thematic Issues

9:00-12:00 Individual Presentations and Panel Discussions at Separate Locations

7-1: Life Sciences and Public Health Data

Location:
Chinese PLA General Hospital

Co-Chairs:

- YIN Ling, Director, Neuroinformatics Center, Chinese PLA General Hospital
- Vishwas CHAVAN, National Chemical Laboratory, India

9:00 A Contractually Reconstructed Research Commons for Scientific Data in a Highly Protectionist Intellectual Property Environment
Jerome REICHMAN, Duke Law School, United States and Paul UHLIR, U.S. National Academies

9:20 International Medical Scientific Data Sharing
YIN Ling, Chinese PLA General Hospital

9:40 The NIH Roadmap for Medical Research
Belinda SETO, Deputy Director, National Institute of Biomedical Imaging and Bioengineering, National Institutes of Health, United States

10:00 OECD-GSF-NI (Neuroinformatics) Data Sharing
TANG Yiyuan, Dalian Polytech University, China

10:20 The Safeguarding and Sharing of Traditional Chinese Medicinal Database Resources
LIU Baoyan, Chinese Academy of Medicine

10:40 Break

10:55 Protein Data Bank: A Key Biological Resource
FENG Zukang, Protein Data Bank, United States

11:15 Long-Term Study for Digital Human and Scientific Data Sharing
QIN Donglie, The Capital Hospital University, China

11:35 Open Access to Scientific Data on Biological Diversity: An Urgent Need for China
James EDWARDS, Global Biodiversity Information Facility, Denmark

11:55 Discussion

Rapporteurs:

- HUANG Funan, Chinese PLA General Hospital
- Paul UHLIR, J.D., U.S. National Academies

12:10 Adjourn

7-2: Earth Sciences, Environmental and Natural Resources Data

Location:
Institute of Geography and Natural Resources, Chinese Academy of Sciences

Co-Chairs:

- LIU Chuang, Institute of Geography and Natural Resources, Chinese Academy of Sciences
- Dora Ann CANHOS, CRIA, Brazil and Member, CODATA Archiving Task Group

9:00 International Policy and Cooperation in Digital Earth
GUO Huadong, Associate Secretary General and Director, Department of International Cooperation, Chinese Academy of Sciences

9:15 The Present Status and Future Development Strategy of China's Sustainable Development Information Network
FU Xiaofeng, The Administrative Center for China's Agenda 21

9:30 Environmental Database Development in Mongolia
Khudulmur SODOV, Director, Information and Computer Center, National Remote Sensing Center, Mongolia

9:45 Data Archiving for Long-Term Preservation in the Satellite Ground Station of the Chinese Academy of Sciences
LIU Dingsheng, Satellite Ground Station, Chinese Academy of Sciences

10:00 Special Considerations for Archiving Data from Field Observations
Raymond McCORD, Oak Ridge National Laboratory, United States

10:15 The Information System for Earth Science Data Sharing of China
QU Jiansheng, Information Center for Natural Resources and Environment, Chinese Academy of Sciences

10:30 Towards a Balanced Performance Appraisal System in the Digital Era for Data Archiving and Sharing in China
WANG Zhengxing, Global Change Information and Research Center, Institute of Geography and Natural Resources, Chinese Academy of Sciences

10:45 Break

11:00 Uses of Seismic Data and the Importance of Open Access to Major Data Centers in Seismology
Paul RICHARDS, Columbia University, United States

11:15 Existing Infrastructure for International Exchange of Seismic Data
Raymond WILLEMANN, GEM Technologies, United States

11:30 Seismic Data Sharing in China
YANG Dake, China Earthquake Administration

11:45 Discussion

Rapporteurs:

- ZHU Xiaohua, Global Change Information and Research Center, Institute of Geography and Natural Resources, Chinese Academy of Sciences
- Roberta BALSTAD MILLER, Director, CIESIN, Columbia University, United States and Chair, U.S. National Committee for CODATA

12:10 Adjourn

7-3: Scientific Information, Journals and Digital Libraries

Location: Library of Sciences, Chinese Academy of Sciences

Co-Chairs:

- ZHANG Xiaoling, Director, Library of Sciences, Chinese Academy of Sciences
- Steve ROSSOUW, Cape Technicon University, South Africa

9:05 An Open-Access Future
Helen DOYLE, Director, Development & Strategic Alliances, Public Library of Science, United States

9:25 Developing an Open Digital Library: The Practice of the Chinese National Science Digital Library (CSDL)
ZHANG Xiaolin, Director of Science Library, Chinese Academy of Sciences

9:45 Are Economic Journals Becoming Obsolete?
Theodore Carl BERGSTROM, University of California at Santa Barbara, United States

10:05 Arguments on the Future of the Library and Economics of Open Access
John WILLINSKY, University of British Columbia, Canada

10:25 Information, Journals, and Digital Libraries: Can Developing Countries Become Key Players in the Information Society?
Lulama MAKHUBELA, Research & Development Directorate, National Development Agency, South Africa

10:45 Break

11:10 Arguments on Open Access to Scientific Information
MENG Liansheng, Chinese Academy of Sciences

11:20 The International Network for the Availability of Scientific Publications (INASP)—Opportunities in a Changing Information Environment
Pippa SMART, Head of Publications, Publishing Initiatives and Publishing Training, INASP, United Kingdom

11:40 Discussion

Rapporteur:

- Julie ESANU, Office of International S&T Information Programs, U.S. National Academies

12:10 End of Meeting

B

Biographical Summaries of Workshop Speakers and Steering Committee Members

William L. Anderson is a co-founder of Praxis101, where his consulting practice focuses on user-centered information systems architecture, participatory design, software engineering practice innovation, and organizational learning. Before founding Praxis101 he worked for Xerox Corporation in distributed system architecture, technology strategy, and advanced product development. He pioneered co-development and customer collaboration on one of the first digital libraries, a joint project between Cornell University and Xerox. He has published papers on digital library product development, participatory design of product prototypes, and software development practices and tools. Prior to Xerox, Dr. Anderson worked in the telecom, image management, and pharmaceutical industries. Dr. Anderson holds a Ph.D. in Theoretical Chemistry from Rensselaer Polytechnic Institute. He is a member of the U.S. National Committee for the Committee on Data for Science and Technology (CODATA) and co-chair of the CODATA Preservation of and Access to Scientific and Technical Data in Developing Countries.

Yasuyuki Aoshima is director and representative of UNESCO Office Beijing. He has an academic background in engineering sciences with a doctorate of engineering from the University of Tokyo in 1977. From 1970 to 1975 he worked as a teaching and research assistant at the Swiss Federal Institute of Technology in Lausanne. From January 1976 to February 1982 he worked as an engineer at the Nippon Kokan (steel making, ship-building, and construction company), Tokyo. He joined UNESCO in February 1982 as senior purchasing officer until 1988 when he became the chief of budgetary control and monitoring at the Section of Fellowships and Equipment Division. In 1993 he became chief of the Fellowship Section of UNESCO for a few months before being transferred to the Natural Science Sector as senior programme specialist at the Engineering and Technology Division responsible for the University-Industry-Science-Partnership programme assisting also the World Solar Programme. From October 1997 to September 2001 he worked as a senior programme specialist in the Science, Technology and Informatics division in the Jakarta field office. In September 2001 he was transferred to the Beijing Office as the director and representative of UNESCO to the People's Republic of China, Mongolia, the Republic of Korea, Japan, and the Democratic People's Republic of Korea.

Peter Arzberger is currently Director, Life Sciences Initiatives at the University of California, San Diego, and Director of the National Biomedical Computation Resource. Dr. Arzberger received his Ph.D. in mathematics from Purdue University. In 1988 he moved to the National Science Foundation (NSF) and was Program Director for Statistics and Probability, and later the first Program Director for the Computational Biology Activities program. He also served as Deputy High Performance Computing and Communications Coordinator at NSF. In 1995 he moved to the San Diego Supercomputer Center as Executive Director, and helped lead the development of the National Partnership for Advanced Computational Infrastructure and served as its first Executive Director. His interests focus on the interface between computing, information technology, and mathematics with the broad spectrum of biomedical research and biology from the molecular through to ecological to biodiversity scales. In addition he is interested in policy issues relating to international data sharing. He is the principal investigator on the NSF award to support the Pacific Rim Application and Grid Middleware Assembly.

Roberta Balstad is senior research scientist at Columbia University and director of the Center for International Earth Science Information Network. Dr. Balstad has published extensively on science policy, information technology and scientific research, remote sensing applications and policy, and the role of the social sciences in understanding global environmental change. Dr. Balstad received her Ph.D. from the University of Minnesota in 1974. She was appointed senior fellow at Oxford University in 1991-1992 and a guest scholar at the Woodrow Wilson International Center for Scholars in 1994. She was previously the director of the Division of Social and Economic Sciences at the NSF, the founder and first executive director of the Consortium of Social Science Associations, and president/CEO of CIESIN prior to its joining Columbia University. She has lectured widely, both in the United States and abroad. From 1992 to 1994, she was vice president of the International Social Science Council and has also served as chair of the NRC Steering Committee on Space Applications and Commercialization, the NATO Advisory Panel on Advanced Scientific Workshops/Advanced Research Institutes, the AAAS Committee on Science, Engineering and Public Policy, and the Advisory Committee of the Luxembourg Income Study. She currently serves on the Board of Directors of the OpenGIS Consortium and the Advisory Board of the Stellenbosch Institute of Advanced Studies (South Africa).

Ted Bergstrom holds the Raznick Chair in Economics at the University of California, Santa Barbara (UCSB). A native of Minnesota, he has an undergraduate degree in mathematics from Carleton College, and a Ph.D. in economics from Stanford University. His first employment was in the economics department at Washington University in St. Louis, Missouri. In 1975 he moved to the University of Michigan, where he was a professor of economics until 1997, when he moved to UCSB. Professor Bergstrom's interests within economics range widely over pure and applied microeconomic theory. Topics which have been central to his recent research include biology and economics, theory of public economics, and the economics of scholarly information.

Dora Ann Lange Canhos is the Project Director of the Centro de Referencia em Informacio Ambiental (CRIA; Reference Center on Environmental Information, www.cria.org.br). She has been working with databases and online information systems since 1985. She has been involved with biodiversity information networks since 1992, as a member of BIN21 (Biodiversity Information Network—Agenda 21) and responsible for its Web site, and serving as technical coordinator of the project BINbr (Biodiversity Information Network—Brazil) for the Ministry of the Environment from May 1997 to April 2001. She is also member of the Clearing-House Mechanism Informal Advisory Committee (http://www.biodiv.org/) and of the Liaison Working Group of the Biosafety Clearing-

House (http://bch.biodiv.org/), both of the Convention on Biological Diversity. Currently at CRIA, Dr. Canhos is a member of the developing team of the speciesLink network (http://splink.cria.org.br/).

Vishwas Chavan works at the National Chemical Laboratory (NCL), Pune, India, in the field of Biodiversity Informatics developing tools and standards to improve infrastructure and capacity building in collection, collation, analysis, prediction, and dissemination of knowledge. Mr. Chavan has been using information and communication technologies to create collaboratory environment, for biologists and ecologists ensuring that rate of exchange and sharing of information on biotic resources improves. One significant and impact-making activity that Vishwas is leading is development of the Electronic Catalogue of Known Indian Fauna (IndFauna), which aims to collates baseline information about 90,000 known faunal species in India. Prior to joining NCL, Vishwas was employed at the Centre for Cellular and Molecular Biology Hyderabad facilitating the activities of the Bioinformatics Centre and preparing a master plan for bioinformatics activities. He has published more than 60 research and review papers, and developed a dozen databases and program packages that have been employed in the service of the scientific community. Vishwas received a Fulbright Professional Fellowship in Information Sciences & Technology (2000-2001), during which he was closely associated with the U.S. National Biological Information Infrastructure. He is also a member of the CODATA Preservation of and Access to Scientific and Technical Data in Developing Countries, and co-chair of the Information Management Committee of Global Invasive Species Program. Vishwas received a B.S. in zoology from N. Wadia College, after which he joined the School of Environmental Sciences, University of Pune, and received an M.S. degree in environmental sciences in June 1990.

Jun Chen is President of the National Geomatics Center of China (NGCC). After obtaining his M.Sc. at Wuhan Technical University of Surveying and Mapping (WTUSM) in 1983, he worked as a lecturer, associate professor, and professor at WTUSM from 1983-1995. He was the executive director of the National Laboratory for Information Engineering in Surveying and Mapping from 1990-1995. He joined NGCC in the end of 1995. From 1999 to 2003, he was President of the Chinese Association for Geographical Information Systems, the corresponding member of the International Eurasian Academy of Sciences.

Panqin Chen is Deputy Director of the Bureau of Resources and Environmental Sciences and Technology at the Chinese Academy of Sciences. He is vice-chair of the Science Committee of the Chinese Ecosystem Research Network.

Robert Chen is CIESIN's Deputy Director and a Senior Research Scientist. He manages the Socioeconomic Data and Applications Center, a data center in NASA's Earth Observing System Data and Information System. He is currently Secretary-General of CODATA of the International Council for Science (ICSU). He is an ex officio member on both the U.S. National Committee for CODATA of the U.S. National Research Council and the Task Group on Data and Scenario Support for Impacts and Climate Analysis of the Intergovernmental Panel on Climate Change (IPCC). Dr. Chen serves as CIESIN's Technical Representative to the Open Geospatial Consortium and participates in the U.S. Federal Geographic Data Committee Historical Data Committee. He is also a member of the Scientific Advisory Council of the Meadowlands Environmental Research Institute of the New Jersey Meadowlands Commission. Dr. Chen has also coordinated CIESIN's spatial analysis and mapping support to the Millennium Development Project led by EI Director Jeffrey Sachs and oversees other projects on poverty mapping, sustainability indicators, and public health applications of Earth science data. He received his Ph.D. in geography from the University of North Carolina at Chapel Hill and holds Masters and Bachelors degrees from the Massachusetts Institute of Technology.

Jinpei Cheng is Vice Minister of the Ministry of Science and Technology of China and vice-chairman of the Central Committee of the China Zhi Gong Party. He attended Nankai University and Northwestern University in the United States, where he obtained a doctorate. Cheng previously served as an academician at the Chinese Academy of Sciences and the Third World Academy of Science. He serves as vice-minister of the Ministry of Science and Technology.

Michael T. Clegg received his B.S. and Ph.D. degrees in agricultural genetics and genetics, respectively, at the University of California, Davis. In 1972 he joined the faculty of Brown University, moving from there to the University of Georgia in 1976. In 1984, he assumed his present position as Professor of Genetics at the University of California, Riverside. He also served as Dean of the College of Natural and Agricultural Sciences from 1994 to

2000 and he is founding Director of the Genomics Institute and at the University of California, Riverside. Professor Clegg's research specialty is population genetics and molecular evolution. His early work in population genetics focused on the dynamical behavior of linked systems of genes in plant and *Drosophila* populations. During this period, he also contributed to the theoretical study of multilocus systems employing computer simulations together with the analysis of mathematical models. Later he helped pioneer the comparative analysis of cholorplast DNA variation as a tool for the reconstruction of plant phylogenies. His current work is concerned with the comparative genomics of plant gene families, the molecular evolution of genes in the flavonoid biosynthetic pathway, the use of coalescent models to study crop plant domestication, and the application of molecular markers to avocado improvement. Professor Clegg was elected to membership in the U.S. National Academy of Sciences in 1990 and he was elected a fellow of the American Academy of Arts and Sciences in 1992. He was elected foreign secretary of the U.S. National Academy of Sciences in 2002. He has also served as president of the American Genetic Association (1987) and the International Society for Molecular Biology & Evolution (2002), and chair of the Section on Agriculture, Food and Natural Resources of the American Association for the Advancement of Science (2003).

Helen J. Doyle is the director of Development and Strategic Alliances for the Public Library of Science. As former director of the Science Program at the David and Lucile Packard Foundation, Helen J. Doyle developed and managed grant-making programs in basic and interdisciplinary academic research, higher education and diversity, science and technology for international development, and science education. Helen spent nearly ten years in New York City, where she majored in biochemistry at Barnard College and received her Ph.D. in biological sciences from Columbia University for her work on *Drosophila* developmental genetics. She then spent two years as a postdoctoral fellow at the Max Planck Institute for Developmental Biology in Tubingen, Germany. Returning to her native California, Helen continued her research on early development and cell communication with J. Michael Bishop at the University of California, San Francisco (UCSF). While a postdoc at UCSF, Helen developed a strong interest in public understanding of science and science education reform issues. She joined UCSF's Science & Health Education Partnership as an academic coordinator, working with San Francisco public schools and the university to improve science, math, and health education. She has also taught at various institutions, including UC Berkeley, Mills College, and the California Academy of Sciences.

James Edwards is the executive secretary of the Global Biodiversity Information Facility (GBIF), an intergovernmental organization devoted to making biodiversity data freely and openly available via the Internet. He is also the director of the GBIF Secretariat in Copenhagen, Denmark. He received his B.S. (1967) and Ph.D. (1976) degrees from the University of California at Berkeley. His research interests are the systematics and functional morphology of amphibians and fishes, and biodiversity informatics. From 1974-1976, Dr. Edwards was an instructor in the Biology Department at Queens College of the City University of New York, and from 1976-1982 he was an assistant and associate professor in the Zoology Department at Michigan State University. In 1982, he took a position in the Directorate for Biological Sciences at the NSF, which funds the vast majority of non-medical biological research at U.S. colleges and universities. While at the NSF, he served successively as program director for several programs (Systematic Biology, Biological Research Resources, Field Stations and Marine Laboratories, and Biotic Surveys and Inventories), as deputy division director for Biotic Systems and Resources, and as deputy assistant director for Biological Sciences. In the latter capacity, he was the second-in-command of a yearly budget of approximately $500 million. Dr. Edwards served on several federal task forces, and was the chair of an interagency steering committee on biological and ecological informatics. He also chaired a working group on Biological Informatics of the Megascience Forum of the Organisation for Economic Co-operation and Development, which in 1999 recommended the formation of the GBIF. Dr. Edwards then chaired the Interim Steering Committee that developed the Memorandum of Understanding for the organization and recruited the requisite number of governmental members and funding to allow it to come into existence in March 2001. Currently, he is on a five-year leave of absence from NSF in order to serve as the executive secretary of GBIF.

Julie M. Esanu is a program officer for the Office of International Scientific and Technical Information Programs (ISTIP) at the U.S. National Academies. Her emphasis is policy and management issues related to digital scientific and technical data and information, primarily through the support of the U.S. National Committee for CODATA. Julie is the co-editor of two recent and related National Academies reports, including

Open Access and the Public Domain in Digital Data for Science: Proceedings of a Symposium (National Academies Press [NAP], forthcoming) and *The Role of Scientific and Technical Data and Information in the Public Domain: Proceedings of a Symposium* (NAP, 2003). She has provided program and research support to other National Academies' projects examining the role of remote-sensing research and applications; reviewing C4I planning for the Department of Defense; assessing the research programs at the Army Research Laboratory; and examining the allocation of federal research and development funds. Julie received her B.A. in political science and international affairs from George Washington University.

Zukang Feng received a Ph.D. in biophysics at Shanghai Institute of Biochemistry, Academia Sinica, China, in 1991. In 1994, he was awarded the Lise Meitner Postdoctoral Research Fellowship from the Austrian Science Foundation and completed postdoctoral research at the University of Salzburg, Austria. In 1996, Dr. Feng joined the Nucleic Acid Database (NDB) project at the laboratory of Professor Helen Berman of Rutgers University. In this position he developed the first package of programs that validate nucleic acid structural data and automate data processing systems. In 1998 he joined the Protein Data Bank (PDB) project. Dr. Feng is presently the leader of software development for both PDB and NDB projects and is responsible for the curation and annotation of the data that are received and distributed worldwide.

Xiaofeng Fu is Professor and Director of the Information Division of the Administrative Center for China's Agenda 21, which is a part of the Ministry of Science and Technology. He also serves as Deputy Team Leader of the National Science Data Sharing Project Workshop, Deputy Director of the Science Data Sharing Project Office, and Member of the CODATA Preservation of and Access to Scientific and Technical Data in Developing Countries. Dr. Fu is also in charge of the construction and administration of the China Sustainable Development Information Network. Previously, Dr. Fu worked as Assistant Research Fellow for the Institute of Geographic Sciences & Natural Resources Research of the Chinese Academy of Sciences. He received a Doctor of Science from Nanjing University in 1997 and also holds a postdoctorate degree.

Huadong Guo is Deputy Secretary-General of the Chinese Academy of Sciences. He graduated from Nanjing University in 1977, received a master's degree from the Graduate School of the Chinese Academy of Sciences in 1981, and studied at Oregon State University from 1984 to 1985. Guo spent his formative research years with the Institute of Remote Sensing Application at the Chinese Academy of Sciences. He served as a research fellow at the institute as well as an executive deputy director. In 1988, he ascended to the institute's directorship. Guo later became director of the key lab of remote-sensing information science at the academy. A member of the National 863 Program Information Acquisition and Process Technology Theme Expert Team, Guo has served as executive deputy editor-in-chief of the *Remote Sensing Journal* and is a member of the editorial board of six Chinese and foreign magazines. He is the director-general of the environment remote-sensing division of the China Geography Society. Beginning in the late 1970s, Guo has studied remote-sensing information science, notably radar-to-ground observations. He has supervised more than 20 domestic and international research projects and published more than 140 theses. Guo is the winner of numerous first, second, and third prizes of scientific and technological progress from the academy as well as a prominent contribution prize about the National 863 Program.

Jianguo Han is Director of the International Cooperation Bureau of the National Natural Science Foundation of China.

Qiheng Hu graduated from the Moscow Institute for Chemistry Mechanics with a specialization in Industrial Automation in 1959 and obtained a Ph.D. in 1963. From 1980 to 1982, she visited Case Western Reserve University of the United States as a research professor. From 1983 to 1989 she was Director of the Institute of Automation in the Chinese Academy of Sciences; in 1986 she led the construction of the 1st National Laboratory on Pattern Recognition in China. From 1988 to 1996 she served as Vice-President of the Chinese Academy of Sciences. She is a past president of the China Computer Federation and the China Association for Automation. She was elected Member of the Chinese Academy of Engineering in 1994. She currently serves as Vice-President of the China Association for Science and Technology, President of the Internet Society of China, and a member of the Working Group for Ethics in Research Training of UNESCO. She is enthusiastic in promoting international exchanges and collaborations for the Chinese scientific community. Since 1994 she has devoted great efforts for the introduction and development of the Internet into China. She is a member of the State Advisory Committee on Information, and has been active in ICSU activities on the Ethics for Science from 1996 to 2002.

Dingcheng Huang is a Professor of the Chinese Academy of Sciences, President of the China Engineering Geology Commission, and Director of the Expert Committee of the China Scientific Data Sharing Program.

Funan Huang graduated from The First Military Medical University in 1990 with a major in medicine, and obtained his Doctor's Degree from Chinese PLA General Hospital and Graduate Medical School in 1998 with a specialization in geriatric neurology and neurobiology. From 2000 to 2003, he worked at the Parkinson Research Unit in Thomas Jefferson University, Philadelphia. Dr. Huang's main clinic and research work was in the fields of clinic neurodiagnostics and neurodegenerative diseases and cerebral vascular diseases. Dr. Huang has published 20 papers in nationwide and international academic journals.

Tieqing Huang works in the Department of Natural Resources and Environment Management of the Chinese Academy of Sciences.

Shuichi Iwata is professor of data science and environmental engineering at the University of Tokyo and president of CODATA, a term he holds until 2006. Dr. Iwata received his doctorate in nuclear engineering from the University of Tokyo in 1975. He has served in various capacities at the University of Tokyo, including lecturer, associate professor, and head of the Metallurgical Division (1978-1981) of the Engineering Research Institute; associate professor of nuclear fuels and materials (1981-1991); professor of materials design (1991-1992), Department of Nuclear Engineering; professor of design science (1992-2002), director (1997-2000), and professor of life cycle engineering (2002-2003) at the RACE (Research into Artifacts, Center for Engineering) Center; and professor of Design Science of Materials (2003-2004), Quantum Engineering and Systems Science, School of Engineering. He served as a guest researcher in FIZ-Karlsruhe, Germany, from October 1985 to October 1986. His work includes researches on design science of materials and engineering products, nuclear fuels and materials and materials databases. He has also served as project leader and coordinator in the fields of materials databases and materials design. Dr. Iwata serves as chairman of JSPS 122 Committee and member of SCJ Liaison Committee. He is a member of the Academic Societies for the Japan

Institute of Metals, the Iron and Steel Institute of Japan, Japan Society of Energy and Resources, the Physical Society of Japan, Information Processing Society of Japan, and the Atomic Energy Society of Japan. He received the Promotion of Science and Technology Information Award from JST in 1998, a Paper Award from the Japan Institute of Metals in 1999, and the GIW Best Paper Award in 2003.

Heather Joseph is President and Chief Operating Officer of BioOne, a Web-based aggregation of research in the biological, ecological, and environmental sciences. In this position, Ms. Joseph leads the nonprofit startup enterprise's business, operational, administrative, and strategic development. Ms. Joseph has worked on a variety of innovative electronic publishing projects during the past decade. As Director of Publishing for the American Society for Cell Biology (ASCB), she managed the growth of *Molecular Biology of the Cell*, the first journal to partner with the National Institutes of Health's PubMed Central initiative. She also created a system to peer review and publish multimedia content in the journal. Before ASCB, Joseph held publishing positions with the Society for Neuroscience, where she managed the transition of *The Journal of Neuroscience* from print to Web publication, and at Elsevier Science. She began her work in scientific publishing at the American Astronomical Society, where she collaborated in the creation of one of the first fully electronic journals, *The Electronic Astrophysical Journal Letters*, with funding from the NSF. Ms. Joseph also participates in several professional societies, currently serving as a member of the Board of Directors for the Society for Scholarly Publishing, and as a member of the Program Committee for the upcoming meeting of the Council of Science Editors.

Menas Kafatos is dean of the School of Computational Sciences at George Mason University. The School is an interdisciplinary academic unit that provides graduate, state-of-the-art education and training in the biological and physical sciences, with an emphasis on the computational and data analysis techniques and methodologies. He also directs the Center for Earth Observing and Space Research, a center that focuses on both Earth and space sciences research and applications. Kafatos has broad interests in astrophysics, cosmology, Earth systems science, data systems, foundations of quantum theory, and neuroscience. Author or editor of 12 books, and more than 180 articles, he is the recipient of sizable grants in a variety of areas. He is the principal investigator of the VAccess/Mid-Atlantic Geospatial Information Consortium funded by NASA, which provides access to remote-sensing and other data and information products to a variety of state and local agencies.

Jun Li works for the Academy of Macroeconomic Research of the Chinese National Development and Reform Commission, spending most of his time working for the office for National Geospatial Information Coordination Committee, which focuses on an e-government basin database for natural resources and geospatial basic information database. Li previously worked for the resources and environment data center that is housed in the Institute of Geographic Sciences and Natural Resources Research of the Chinese Academy of Sciences. From 1998 to 2000, he designed and built the Geographical Information System for Landslide Research for landslide risk assessment and landslide information exchange. He received his Ph.D. in cartography and geographical information sciences from the Institute of Geography of the Chinese Academy of Sciences.

Shunbao Liao is an Associate Professor in the Global Change Information and Research Center of the Institute of Geographic Sciences and Natural Resources Research of the Chinese Academy of Sciences. Dr. Liao has a Ph.D. in the application of remote sensing and GIS. Dr. Liao's interests include the analysis and design of information systems and the development and sharing of geoscientific data.

Anne M. Linn is a senior program officer with the Board on Earth Sciences and Resources of the National Academies. She has been with the board since 1993, directing the USA World Data Center Coordination Office and staffing a wide variety of geophysical and data policy studies. In addition, she is the secretary of ICSU's Panel on World Data Centers, and a member of the ICSU Ad Hoc Committee on Data. Prior to joining the staff of the National Academies, Dr. Linn was a visiting scientist at the Carnegie Institution of Washington and a postdoctoral geochemist at the University of California, Berkeley. She received a Ph.D. in geology from the University of California, Los Angeles.

Chuang Liu is Professor and Director of the Global Change Information and Research Center in the Institute of Geography and Natural Resources at the Chinese Academy of Sciences. She currently serves on the Chinese CODATA as Co-Chair of the Task Group on Preservation of and Access to

Scientific and Technical Data in Developing Countries and Committee of Earth Observation Satellites (CEOS) as the User Co-Chair of the Working Group of Information Systems and Services. She is Chair of the Spatial Data Committee of Chinese Associate of Geographical Information Systems and Secretariat General of the Working Group of Remote Sensing and Data Information Systems, China National Committee for IGBP. Dr. Liu is active in research on the broad issues regarding strategy, technology, and capacity building of open access to environmental data. She is Chair of the Expert Group on International Cooperation for the National Facilities and Information Infrastructure for Science and Technology Program, Member of the Working Group of China Scientific Data Sharing Program, and ICSU Priority Area Assessment Panel on Scientific Data and Information. Dr. Liu received her Ph.D. in geography from Peking University in 1989. She was Visiting Professor at the University of British Columbia, Canada, from 1992-1993, Post-Doctor Fellow and then Information Scientist, as well as the China Project Leader in CIESIN, USA from 1994-1998. She also served UNDP/FAO and Asia Development Bank as the GIS Consultant and Technical Assistant.

Depei Liu is Vice President of the Chinese Academy of Engineering, President of the Chinese Academy of Medical Sciences, and President of the Peking Union Medical College. Dr. Liu holds a Ph.D. in molecular biology and has studied at the Institute of Basic Medical Sciences of the Chinese Academy of Medical Sciences, the Peking Union Medical College, and the University of California, San Francisco. Dr. Liu's research interests include medical molecular biology; gene regulation of expression, gene transfer, and gene therapy; and transgenic animals and disease models. Dr. Liu has published more than 100 academic papers. Dr. Liu is a member of the Chinese Society of Biochemistry and Molecular Biology, the American Society of Hematology, the Society of Chinese Bio-scientists in America, the Chinese Academy of Engineering, and the Chinese Medical Association.

Lulama Makhubela has more than 20 years of professional activism as ex-professor in Information Science in various universities, including the University of the Western Cape, South Africa. She has consulted in library and information services with a specific focus on collection development policies, information literacy and ICTs in teaching. She served as a board member in several national bodies and is currently a member of national committees in research and development, information, and data-related matters.

Dr. Makhubela is currently head of the Research and Development Directorate at the National Development Agency (NDA) a position that she took in October 2003. NDA is an agency of the South African Government tasked with eradicating poverty in poor communities. She is tasked with a challenge of developing the South African Poverty Data Center. Dr. Makhubela is a member of the CODATA Preservation of and Access to Scientific and Technical Data in Developing Countries.

Raymond A. McCord is an environmental information manager in the Environmental Sciences Division at Oak Ridge National Laboratory (ORNL). Dr. McCord has been a staff member at ORNL for 18 years. During this time, he managed the development and operation of three major environmental information systems supporting environmental research, restoration, compliance, and assessment. Dr. McCord was also responsible for establishing a geographic information system (GIS) center within the Division. Currently, Dr. McCord is manager of the data archive for the Atmospheric Radiation Measurements (ARM) Program. This archive contains more than 6,000,000 data files (~30 TB of storage) of information about meteorology, atmospheric physics, and cloud formation. The ARM Program is a major component of the Department of Energy's global climate change research program. Prior to working at ORNL, Dr. McCord was employed as an information analyst at Scinece Applications International Corporation (SAIC) in Oak Ridge. Dr. McCord received his Ph.D. in ecology from the University of Tennessee in 1980.

Dahe Qin is an internationally recognized scientist in Glacier (including Antarctic ice sheet) and ice core studies relating to paleoclimate and paleoenvironment. He has committed himself to cryosphere and climate change research since 1970 and has done field work in Antarctica and Tibet. He presently focuses on environmental and climatic variation; climatic and environmental evaluation; and glacier change relating to global change. Previously, he served as chief scientist of the projects supported by the Chinese National Natural Sciences Foundation, the Basic Research Project of the Ministry of Science and Technology, and the Key Innovation Projects of the Chinese Academy of Sciences, as well as of related bilateral cooperation programs between China and the United States, Australia, and France. He also oversees the scientific Assessment on Climate, Ecology and Environment of Western China. He was voted Academician of the Chinese Academy of Sciences in 2003 for his research on the cryosphere. Dr. Qin is currently the administrator of the China Meteorological Administration as well as Director of the China National Climate Committee, Permanent Representative of China with WMO, Co-chair of the Working Group I of IPCC, and Vice-Chair of the China National Committee of IGBP and International Commission of Science and Snow and Ice.

Jerome Reichman is Bunyan S. Womble Professor of Law at Duke Law School. He has written and lectured widely on diverse aspects of intellectual property law, including comparative and international intellectual property law and the connections between intellectual property and international trade law. His articles in this area have particularly addressed the problems that developing countries face in implementing the World Trade Organization's Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS Agreement). Other recent writings have focused on intellectual property rights in data; the appropriate contractual regime for online delivery of computer programs and other information goods; and on the use of liability rules to stimulate investment in innovation without impoverishing the public domain. His most recent articles are "The Globalization of Private Knowledge Goods and the Privatization of Global Public Goods" (co-authored with Keith Maskus), 7 *Journal of International Economic Law No. 2* (forthcoming, 2004), and "A Contractually Reconstructed Research Commons for Scientific Data in a Highly Protectionist Intellectual Property Environment" (co-authored with Paul Uhlir), 66 *Law and Contemporary Problems* 315-462 (2003). Professor Reichman serves as special advisor to the U.S. National Academy of Sciences, the National Academy of Engineering, and ICSU on the subject of legal protection for databases. He is an academic advisor to the American Committee for Interoperable Systems (ACIS); a consultant to the Technology Program of the United Nations Conference on Trade and Development (UNCTAD); and a member of the Board of Editors, *Journal of International Economic Law.*

Paul G. Richards was born in the United Kingdom and obtained a Bachelor's degree in mathematics at the University of Cambridge before coming to the United States in 1965, where he obtained his Ph.D. in seismology at the California Institute of Technology. Richards has been a professor at Columbia University since 1971, teaching and doing research in all aspects of seismology. He has co-authored a major seismology textbook (translated into Chinese), and for the last five years has specialized in practical methods to obtain accurate locations for the thousands of earthquakes that occur each year in East Asia. He has received fellowships from the Sloan, Guggenheim, and MacArthur Foundations, and is a member of the Council on Foreign Relations.

Steve Rossouw graduated from the Universities of Pretoria, Cape Town, and Stellenbosch with a DPhil in Information Science. Professor Rossouw was the executive director of the Information Center, Medical Research Council from 1975-1992. He is presently a professor of Information Studies at Cape Technikon, a position he has held since 1992.

Peter Schröder studied sociology (mental health and epidemiology as minors) at the University of Amsterdam. He worked as a journalist and rock critic before joining the Ministry of Education and Science as policy advisor for educational support systems. After a stay at Utrecht University managing the multidisciplinary Urban Networks research program, he rejoined the Ministry's Directorate of Research and Science Policy as policy advisor on social sciences and information policies. Focusing on issues of access to data and information for research he acted as secretary of the Auditor General's committee advising on privacy protection in scientific research that led to the establishment of the Scientific Statistical Agency in The Netherlands. As coordinator for information policy he is involved in e-science programs for cultural heritage and humanities. Peter Schröder was co-chair of the CSTP/OECD group that published the report *Promoting Access to Public Research Data for Scientific, Economic and Social Development* (March 2003, see http://dataaccess.ucsd.edu and see also *Science*, Vol. 303, 1777-1778, 19 March 2004) and secretary to the OECD group that framed the draft *Guidelines on Access to Research Data from Public Funding* in the *Declaration* endorsed by OECD science ministers at their meeting on 30 January 2004 in Paris (see http://www.oecd.org/document/15/0,2340,en_2649_34487_25998799_1_1_1_1,00.html).

Belinda Seto joined the National Institute of Biomedical Imaging and Bioengineering (NIBIB) as its deputy director in December 2003. The NIBIB, a component of the National Institutes of Health (NIH), improves human health by leading the development and accelerating the application of biomedical technologies. The Institute is committed to integrating the physical and engineering sciences with the life sciences to advance basic research and medical care. In her position, Dr. Seto oversees NIBIB's programs, priorities, resources, policies, and research dissemination efforts. Prior to joining the NIBIB, Dr. Seto served as the acting deputy director for Extramural Research, NIH, as well as the director of the Office of Extramural Research (OER). OER serves as the focal point for policies and guidelines for extramural research grants administration. Prior to joining OER, Dr. Seto held positions in other components of the NIH, as well as in the Office of the Assistant Secretary for Health, Department of Health and Human Services (DHHS). She earned her Ph.D. in biochemistry at Purdue University in 1974. Following postdoctoral training in the National Heart, Lung and Blood Institute, Dr. Seto joined the Food and Drug Administration where she conducted research in virology for nearly 10 years. She is the recipient of numerous awards for her research, including the DHHS Secretary's Award for Exceptional Achievement, Inventor's Awards, NIH Director's awards, and she is listed in the American Men and Women of Science. Dr. Seto has served on numerous NIH and interagency committees, and is a member of several professional societies including the American Society for Biochemistry and Molecular Biology.

Pippa Smart has worked in academic publishing for over 17 years in both the commercial and noncommercial sectors. She has experience in print and electronic production, business and publishing development, and, in her current role for INASP, works in an advisory role with publishers in less developed countries. She is also the chair of the Professional Development Committee of the Association of Learned, Professional and Society Publishers, and sits on its Council.

Carthage Smith is deputy executive director of ICSU, which is an international non-governmental organization, whose membership includes national science bodies and international scientific unions. ICSU's mission is to strengthen international science for the benefit of society. It achieves this by: (1) planning and coordinating international science programs and (2) representing the science community in international policy fora. Carthage has been at ICSU since 2001 and his activities have included leading ICSU's representation at the World Summit on the Information Society (Geneva, December 2003). Prior to joining ICSU, he was head of the International Section at the Medical Research Council in the United Kingdom. Originally trained as a biochemist, his Ph.D. and research background are in neuroscience.

Khudulmur Sodov is the Director of the Information and Computer Center and Director of the National Remote Sensing Center of the Ministry of Nature and Environment in Mongolia. Khudulmur's educational background is mathematics and computer science. He has participated in a variety of projects on such topics as Mongolian biodiversity, natural resource management using GIS, land-cover mapping, grasslands, and GIS in cadastral mapping.

Chengquan Sun is director of the Scientific Information Center for Resources and the Environment of the Chinese Academy of Sciences. He is also a professor in the Department of Geography at Uni-Bonn as well as a professor at Lanzhou University. Dr. Sun is engaged in the analysis and management of information.

Xiaowei Tang is a professor at Zhejiang University. His research interests include nuclear and high-energy experiments, the application of nucleus technology, and neuroinformatics. Tang is a member of the Chinese Academy of Sciences.

Yiyuan Tang is a professor and director of the Institute of Neuroinformatics and Director of the Laboratory for Brain and Mind at the Dalian University of Technology. Tang also serves as the government representative of the OECD-GSF-NI Working Group, associate director of the National Communication Group of Neuroinformatics, guest professor in the Laboratory for Mental Health of the Chinese Academy of Sciences, and senior visiting professor at the Institute of Psychology & Institute of Biophysics. His research interests include neuroimaging, Chinese cognition, psychosomatic medicine, and body-mind interaction. He has published in the *Journal of Integrative Neuroscience*, *Neuroinformatics*, *NeuroImage*, *Neurocomputing*, *Human Brain Mapping*, the *Chinese Science Bulletin*, and the *Chinese Medical Journal*.

Paul F. Uhlir, J.D., is director of the ISTIP Office at the U.S. National Academies in Washington, D.C., where he has worked in various senior professional capacities since 1984. Paul's area of emphasis is on issues at the interface of science, technology, and law, with primary focus on digital information policy and management. He is currently organizing a series of projects on policy issues concerning open access to public scientific information at both the national and international levels, as well as a series of workshops on data management and policy issues in developing countries. He also has been the National Academies' coordinator on intellectual property-related activities. Paul has published 23 reports through the National Academies, and over 50 articles. Prior to joining the National Academies, he worked as a foreign affairs officer at the U.S. National Oceanic and Atmospheric Administration on international remote-sensing law and policy issues. Additional information about Paul's activities may be found at: http://www7.nationalacademies.org/biso/ISTIP.html.

Henda van der Berg is with the National Research Foundation in South Africa where she is responsible for the Content Resources Function (Information and Data) within the Information Strategy and Advice Unit. She qualified with honours in library and information science in 1976 from the University of South Africa. She has extensive experience in metadata descriptions and international standards from 1971 until 1991. From 1993, Ms. van der Berg worked in the field of database management and database development. She is a developer of the Nexus Database System Portal and databases. She is a member of the Steering Committee for the South African Data Centre for Oceanography and of an expert advisory group that served as a Reference Group for the Institutional Research Information System tasked with the Data Specification Project of the Southern African Research & Innovation Management Association. Her professional memberships include the Library and Information Association for South Africa and SA Online User Group. Ms. van der Berg has published a number of publications in the field of current research information systems, especially the analysis of data for the evaluation of user interfaces.

Peter N. Weiss began work with the Strategic Planning and Policy Office of the National Oceanographic and Atmospheric Administration, National Weather Service, in March 2000. His responsibilities include domestic and international data policy issues, with a view towards fostering a healthy public/private partnership. Mr. Weiss was a senior policy analyst/attorney in the Office of Information and Regulatory Affairs, Office of Management and Budget (OMB), since 1991. Mr. Weiss analyzed policy and legal issues involving information resources and information technology management, with particular emphasis on Electronic Data Interchange and electronic commerce. He is primary author of the information policy sections of OMB Circular No. A-130, "Management of Federal Information Resources," and was a member of the Administration's Electronic Commerce

Working Group. (See "A Framework for Global Electronic Commerce.") From 1990-1991, Mr. Weiss was deputy associate administrator for Procurement Law, Office Federal Procurement Policy. In this position, he analyzed legal and policy issues affecting the procurement process. Major projects included examination of legal and regulatory issues involving procurement automation, policies and FAR revisions to facilitate EDI, as well as ADP procurement legal and policy issues. From 1985 to 1990, Mr. Weiss was the assistant chief counsel for Procurement and Regulatory Policy, Office of Advocacy, U.S. Small Business Administration. From 1981 to 1985, Mr. Weiss was in private practice in Washington, D.C. Mr. Weiss holds a B.A. from Columbia University and a J.D. from the Catholic University of America, Columbus School of Law. Recent publications include "International Information Policy in Conflict: Open and Unrestricted Access versus Government Commercialization," in *Borders in Cyberspace*, Kahin and Nesson, eds., MIT Press 1997; "Borders in Cyberspace: Conflicting Public Sector Information Policies and their Economic Impacts," in Georg Aichholzer/Herbert Burkert (eds.); and "Public Sector Information in the Digital Age: Between Markets, Public Management and Citizens' Rights," Cheltenham: Edward Elgar Publishing (2004).

Raymond J. Willemann is a senior technical advisor for GEM Technology in Washington, D.C. From 1998 to 2003, Dr. Willemann was director of the International Seismological Centre (ISC), a nongovernmental organization that collects, reanalyzes, and redistributes parametric information on earthquakes from more than 100 seismic networks around the world. At the ISC, Dr. Willemann led a modernization effort that greatly increased the volume of data collected, improved the timeliness of data distribution, and offered new avenues for accessing data. Before joining the ISC, he was a senior scientist with SAIC, where he played a key role at the International Data Center for GSETT-3, an experiment in near-real-time international exchange of seismic data. Dr. Willemann is a member of the International Association for Seismoloyg and the Physics of the Earth's Interior (IASPEI) Commission on Seismic Observation and Interpretation and the IASPEI Resolutions Committee, which formulated a resolution on sharing information on seismic station locations in 2003. He earned his Ph.D. in geophysics from Cornell University in 1986. He is a fellow of the Royal Astronomical Society and a member of the American Geophysical Union and the Seismological Society of America.

John Willinsky is currently the Pacific Press Professor of Literacy and Technology and Distinguished University Scholar in the Department of Language and Literacy Education at the University of British Columbia (UBC). He is a fellow of the Royal Society of Canada and a member of the U.S. National Academy of Education, as well as author of a number of books, including *Technologies of Knowing* and *If Only We Knew: Increasing the Public Value of Social Science Research*. Examples of his recent work, including open-source software developed to improve the access and quality of research, are available at the Public Knowledge Project (http://pkp.ubc.ca), which he directs at UBC.