Overview of CODATA Activities Related to WSIS

In December 2003, more than 172 countries participated in the first phase of the World Summit on the Information Society (WSIS). These participants recognized that science has a central role in the development of the information society and that many of the building blocks of the information society are the results of scientific and technical advances made possible by the sharing of research results[1]. The role of e-science in the information society was also recognized in the subsequent Agenda for Action endorsed at the Summit[2].

An underlying theme of the Agenda for Action is the importance of access to information and knowledge. Indeed, many of the principles and actions identified in Phase One of the WSIS emphasize the importance of access to research data and the scientific literature generated from publicly-funded research. For example, Article C3 focuses specifically on "Access to Information and Knowledge" and C3 10 (a) promotes the development of policy guidelines for the *"development and promotion of public domain information as an important international instrument promoting public access to information."*

After a concerted effort, the scientific community was very successful in placing these issues on the WSIS agenda and in having them recognized by the participating governments. It now has a little over a year to respond to the challenges set out in the Agenda for Action.[3] Many initiatives are already under way in the United States, the European Union, and in developing countries that directly address the issues raised at WSIS.

During this interim period, CODATA, in collaboration with other scientific organizations, has already commenced activities in preparation for the next phase in Tunis. Most notably it held a special session within the framework of its 19th International Conference on November 10 in Berlin. A report on this can be found in Annex One.

It will also carry out the following activities to help implement key items from the Agenda for Action:

Organise an international workshop in September 2005 *"Moving Towards an Information Commons for e-Science and Innovation: Identifying, Strategic Institutional and Policy Guidelines"* Information on this workshop can be found in Annex Two of this document.

---

[1] Article 7 of the Declaration of Principles of the World Summit on the Information Society, December 2003, Geneva. http://www.itu.int/wsis/documents/doc_multi.asp?lang=en&id=1161|1160

[2] **22. E-science**

a) Promote affordable and reliable high-speed Internet connection for all universities and research institutions to support their critical role in information and knowledge production, education and training, and to support the establishment of partnerships, cooperation and networking between these institutions.

b) Promote electronic publishing, differential pricing and open access initiatives to make scientific information affordable and accessible in all countries on an equitable basis.

c) Promote the use of peer-to-peer technology to share scientific knowledge and pre-prints and reprints written by scientific authors who have waived their right to payment.

d) Promote the long-term systematic and efficient collection, dissemination and preservation of essential scientific digital data, for example, population and meteorological data in all countries.

e) Promote principles and metadata standards to facilitate cooperation and effective use of collected scientific information and data as appropriate to conduct scientific research.

[3] The second phase of the Summit takes place in Tunisia in November 2005.

- Highlight many of the initiatives that are currently under way within the scientific community that relate to the Agenda Action items, using different tools and platforms to bring this inventory together.

- Commence preparations for participation at the scientific event in Tunis in 2005.

ANNEX ONE

19th International CODATA Conference
Report on the Special Session on the World Summit on the Information Society
10 November 2004
Berlin, Germany

The Committee on Data for Science and Technology, CODATA convened a Special Session on the World Summit on the Information Society (WSIS), which was attended by over 90 people. The complete participant list can be found in Annex A. This Session took place on 10 November 2004, within the framework of the 19th International CODATA Conference: *"Information Society: New Horizons for Science."* The agenda and background documents to this Session can be found on the CODATA website. The Session was organized by CODATA in collaboration with its supporting partners:

- U.S. National Academies;
- French Academy of Sciences;
- The Academy of Sciences for the Developing World (TWAS);
- The International Council for Scientific and Technical Information (ICSTI);
- The International Network for the Availability of Scientific Publications (INASP); and
- The International Union of Psychological Science (IUPsyS)

The Session aimed to:
➢ Motivate discussion among the scientific community on the important role of scientists within the information society;
➢ Highlight initiatives currently unde way within the scientific community that complement the science-related Agenda for Action items from the first phase of WSIS, which took place in Geneva in December 2003. The relevant sections can be found marked in red in the following links: Science in the Declaration of Principles and Science in the Agenda for Action; and
➢ Commence discussions on how the scientific community can constructively contribute to the second phase of WSIS in Tunis 2005.

The first part of the Session was chaired by Dr. Krishan Lal, who has since been elected CODATA Vice-President.

His Excellency Adama Samassekou, President of the WSIS Preparatory Committee for the Geneva Phase, was the keynote speaker. His presentation focused on open access to scientific data and information.

This was followed by presentations by:

- Carthage Smith of ICSU, the International Council for Science;
- Axel Plathe of UNESCO;
- Bertrand de la Chapelle of WSIS online, a community platform to connect all actors addressing WSIS issues*;*

- Peter Schröder on "Cherishing the Memory of Science: Towards International Guidelines for Access to Research Data from Public Funding;" and
- Wesley Shrum on the science event planned during the Tunis phase in November 2005.

These presentations were followed by a Roundtable Discussion with CODATA, its supporting partner organizations, and the invited speakers. The supporting partner organizations were represented as follows:

Gerard Siclet: Académie des Sciences, France
Paul Uhlir: US National Academies
Carol Priestley: International Network for the Availability of Scientific Publications (INASP)
Daniel Schaffer: The Academy of Sciences for the Developing World, TWAS
Michel Sabourin: The International Union of Psychological Science (IUPsyS)
Kathleen Cass: CODATA
This part of the Session was chaired by Dr. Robert Chen, who has since been elected the Secretary General of CODATA.

Participants focused on four major questions related to the scientific community's role in WSIS.

1. What do you see as the major challenges facing your institution regarding scientific data management and access?

* 20,000 petabytes (gross estimate) of data are being produced annually. The problem is not just of access, but of usability of such amounts as well.
* Access and connectivity are essential, but only as first steps. We also need new techniques for knowledge discovery – techniques that depend on an ability to integrate knowledge at different scales.
* New forms of dissemination are potentially useful in helping policy makers and the general public to understand scientific issues.
* Funding is a common problem: How to persuade governments to finance data management as well as the basic science?
* There is a lack of resources (and interest) in digitising heritage data (e.g. astronomical photographic plates).
* There remains a mismatch in the collection of environmental data between what are being gathered and what are actually of most use, particularly in the developing world.
* The International Mathematical Union is working on the goal of digitising all mathematical publications to produce a complete digital library of mathematics.
* Geodiversity needs to be emphasised.
* WSIS should emphasise the need for common data standards.
* Personnel in the developing world need to become more involved. There are issues of language and training, and, specifically, a lack of awareness of the need for archiving data.
* INASP emphasized the need for improved access as a first step, and can provide many examples of how the benefits of increased bandwidth to developing institutions are very quickly realised.

* The Third World Academy of Sciences acknowledges the need for archiving, but their priority is rapid access to the latest information.

2. What issues and accomplishments do you think are the most important to highlight at WSIS in Tunis?

* Need to discriminate among different types (and quality) of data.
* Want to see more new horizons for science arising from WSIS, and a proper respect for, and understanding of, the role of science within the broader Information Society.

* The need to see "better" science coming out of WSIS, and a culture change. Data should be taken seriously; the science is not finished until the associated data have been publicly posted.
* Scientists are beginning to work towards the design of data representation languages that have common structures and vocabularies, to encourage interoperability between related domains.

3. What activities are currently taking place within your institution that directly or indirectly relate to the implementation of the Agenda for Action that has come out of the first phase of WSIS?

* The International Polar Year of 2007-2008 (marking the 50th anniversary of the International Geophysical Year) demonstrates the role of science in promoting international cooperation.
* The World Data Centres offer another good example.
* The forthcoming "Electronic Geophysical Year" will contribute towards the new horizon of taking data and information seriously.
* A project is under way to create a 1:1,000,000 digital map of the entire world, with eight layers critical to sustainable development. The best input so far has come from the developing world.
* The OAI-PMH transport mechanism for metadata in the provision of open access is a noteworthy achievement.

4/5. What specific outcome do you think CODATA and its partners should aim to achieve in Tunis and are there specific actions, in addition to the ones highlighted above, that CODATA and its partners should take in preparation for WSIS in Tunis in 2005?

* Renewed efforts towards the provision of electricity and power globally - no data if no power!
* Much of science is based on relationships, and initiatives promoting interpersonal contacts should be encouraged.
* ICT developments may lead to an entirely different structure for science in the future. CODATA should paint the picture of what science will be like in 15 years time.
* The exercise of producing an inventory of specific activities is very important, but should not end with the Tunis summit.

* Scientists need to engage more with policy makers on issues of relevance. Internet governance is one such area.
* WSIS is an opportunity to emphasise the non-monetary value of sharing knowledge. This is understood intrinsically within scientific culture, but may need to be spelled out to the world at large.
* Intellectual property rights must be managed sensitively in cooperation with WIPO.
* Open access to data and equitable access to publications remain specific goals that should emerge from the WSIS summit in Tunis.

The Session identified five action items for CODATA:

Near-term Actions (Before February 2005)

1. CODATA will distribute a short report on the Session to all participants, together with list of participants.
2. CODATA will contact participants requesting a description of the activities that are currently been carried out by their respective scientific communities that directly/indirectly focus on the science-related issues from the Agenda for Action that came from Geneva. These activities will be highlighted in the inventory of activities that CODATA and its partners are bringing together.

Long-term Actions:

3. CODATA in collaboration with its partners will compile an inventory of activities referred to in item 2.
4. CODATA in collaboration with other international organizations is organizing an International Workshop in September 2005 at UNESCO on "Building Information Commons for e-Science: Strategic Institutional and Policy Approaches."
5. CODATA in collaboration with other members of the scientific community plans to organize a number of sessions within the framework of the Scientific Event in Tunis in 2005.

Annex Two

Moving Towards an Information Commons for e-Science and Innovation: Identifying Strategic Institutional and Policy Guidelines

An International Workshop
Organized by
The Committee on Data for Science and Technology (CODATA)
The International Council for Scientific and Technical Information (ICSTI)
The International Network for the Availability of Scientific Publications (INASP)
The International Council for Science (ICSU)
The United Nations Educational, Scientific, and Cultural Organization (UNESCO)
The Academy of Sciences for Developing Countries (TWAS)
and
The U.S. National Academies

SUMMARY
This proposed workshop will develop strategic institutional and policy approaches at the national and international levels for broadening the "information commons" for public science and innovation, pursuant to the following Statement of Task:

1. Review the opportunities for public science and innovation on the emerging e-infrastructure at the international levels.
2. Review government and university mechanisms in the United States, in the European Union, and in selected developing countries for managing government funded scientific information in the digitally networked research environment, and identify examples that could serve as models for the broader research community.
3. Identify and analyze the institutional, economic, policy, and legal benefits and drawbacks in providing public access to and unrestricted use of government-funded scientific information.
4. Based on the findings under tasks 1-3 above, develop strategic institutional and policy principles and guidelines to enable the scientific community to more effectively manage government-funded scientific information, and for broad consideration by government science agencies, universities, and the research community.

CONTEXT
From a scientific perspective, the importance of access to data and information has never been as important as it is now. The rapid advances in digital technologies and networks over the past two decades have significantly altered and improved the ways that data and information can be produced, disseminated, managed, and used, both in science and in many other spheres of human endeavor. This progress in the emerging e-infrastructure has enabled scientists to perform quantitatively and qualitatively new functions to: collect and create unprecedented and ever-increasing amounts and types of raw data about all natural objects and phenomena; collapse the space and time in which data and information can be made available; facilitate entirely new forms of distributed research collaboration and information production; and integrate and transform the data resources into unlimited configurations of information, knowledge, and discovery. Perhaps most important in this context is that the internet has reduced the marginal cost of dissemination of every additional copy of any information item to near zero.

* Indicates organizations that have been invited to participate, but pending official confirmation.
The development and early institutionalization of many new paradigms of virtual knowledge-based communities and related information activities have occurred in public science. Examples include the

open-source software movement, open public-domain data archives and federated data networks, open access journals, community-based open peer review, collaborative research Web sites, collaboratories for virtual experiments, and virtual observatories, among others. Taken together, these emerging capabilities represent aspects of a broader trend toward both formal and informal peer production of information in a highly distributed, volunteer, and open networked environment. Such activities are based on principles that may be more accurately characterized as a "public scientific information commons," rather than as proprietary information, and that reflect the cooperative ethic that imbues much of public science. These developments have given rise to unprecedented opportunities for accelerating the progress of science and innovation and creating wealth based on the more efficient exploitation of data and non-proprietary information produced through public investments in research.

With regard to access regimes for public research data, the situation is mixed, with some countries and disciplines providing more open and comprehensive access than others. At the international level, there have been a number of notable efforts to institute open access policies, including the 1991 "Bromley Principles" on the full and open exchange of global change research data, the 1996 Bermuda Principles on the Release of Human Genome Sequence Data, the 1997 ICSU-CODATA Principles for Dissemination of Scientific Data, and the 2004 OECD Ministerial Declaration on access to data from publicly-funded research, among others. The practical effectiveness of these initiatives in altering access conditions, however, has not been systematically evaluated, and the adherence of governments and publicly funded research institutions to such principles remains far from universal.

Public information regimes for scientific data produced in developing countries have remained among the least open. In addition to the economic and organizational limitations on the capabilities of the government apparatus for gathering and distributing such data, and the political restrictions placed upon disclosure of information regarding social and economic conditions, access to scientific data and information has been inhibited by researchers' and research institutions' suspicions that free and open information exchanges, like free trade, will turn out not to be "fair" trade. The marked asymmetries between rich-country and poor-country partners in the division of intellectual property rights from new discoveries and inventions have certainly contributed to undermine the ethos of scientific cooperation in some fields, notably the life sciences.

Moreover, it is important to recognize that public policies in the developed and developing countries alike are shaped by legitimate countervailing interests that place limits on openness and cooperation. These include: national security concerns (including grey areas such as "dual use" or "sensitive but unclassified" information), the interests of private-sector parties in the legal protections that have been accorded to their intellectual property rights, and the practice of allowing publicly funded researchers limited periods of exclusive use of their data prior to the publication of their research findings.

Open access to the research literature produced from public funding is also a major issue that has received considerable scrutiny in the past few years worldwide. There are now over 1300 scholarly journals provided under open access conditions on the Internet, including some notable initiatives such as the Public Library of Science and BioMed Central. Policy principles on open access to publicly funded journals were issued in both the United States and Europe in 2003 through the "Bethesda Principles" and the "Berlin Declaration." In 2004, many professional society journal publishers produced the "DC Principles," which also recognized the imperative of broad access to the scholarly literature produced from publicly funded research. Commercial journal publishers are now allowing more open access to the articles they publish as well. Most recently, the U.S. House of Representatives and the House of Commons in the United Kingdom proposed legislation that would enhance public access to scientific literature produced from publicly funded research.

Institutional repositories also have been established for pre-prints and e-prints of journal articles (e.g., the Cornell arXiv, originally established for high-energy physics and now expanded to include other

areas of physics, mathematics, computer science, and computational biology), for individual research articles and other information resources (e.g., the Social Science Research Network, the MIT D-Space initiative), and for university educational material (e.g., MIT's OpenCourseWare)..Public access initiatives in developed countries frequently are being designed either with the needs of developing countries expressly considered, while new open access journals are being established within developing countries themselves. Here too there is a lack of strategic planning or implementation of policies by the government and academic scientific communities, although the adoption of many promising new initiatives from the bottom up, coupled with the recent introduction of some new top-down legislative proposals, make it a particularly appropriate time for a comprehensive review.

At the same time, however, an opposing trend has been gaining momentum worldwide. It is characterized by the progressive privatization of government research functions and data collection activities, and by the commercialization of research results by universities. This trend in the commoditization of public research outputs, including of the underlying data and information resources, is being reinforced by the creation of new legal rights and protectionist mechanisms that are largely extrinsic to the scientific enterprise, but increasingly adopted by it. These include greatly enhanced copyright protection of digital information; new ways to control access to and use of digital data by contractual restrictions that are technologically enforced; and the enactment of proposals for novel intellectual property rights to protect non-copyrightable collections of data. If these pressures continue unabated, they could result in the disruption of established scientific research practices and in the loss of new research opportunities that digital networks and related technologies make possible, as suggested above. The benefits and drawbacks of privatization and commercialization of data and information activities in public-sector science, as well as the application of IP and related restrictions to such activities, need to be more clearly understood and better managed.

Furthermore, within the technical community at least, there appears to be a general expectation that solving the *engineering* problems associated with the advanced hardware and software systems of the emerging E-infrastructure will automatically unleash new scientific capabilities. Advances in engineering themselves, however, will not be enough to realize the societal gains in knowledge creation that are being made feasible. The promise of a great reduction in the marginal social costs of information dissemination and reuse will be achieved only as a result of combined social, legal *and* technical transformations. In comparison with the pace of engineering advances, far greater uncertainties remain inherent in the institutional and policy realms that govern the management of information resulting from public research, and to achieving optimum levels of access to and use of such information.

The workshop will highlight and analyze the variety of experiments that have already been undertaken in this transition phase from the print to digital media, point to what is working and what is not, and recommend how these emerging institutional and policy approaches can be connected in a strategic framework to further develop an information commons for the record of science. The challenge is to understand better and to apply more broadly those mechanisms and approaches that have been shown to work, and to ameliorate policies that impose unnecessary social costs. Moreover, because of the international scope of digital networks and research collaborations, strategic international approaches are both necessary and desirable.

The workshop will build on the successful examples and on previous studies and activities at the participating organizations and at other research and information policy institutions. Moreover, collaboration on this initiative by the major international science policy and scientific information policy organizations—CODATA, ICSTI, INASP, ICSU, UNESCO, OECD, IAP, and TWAS—will provide an unprecedented opportunity to arrive at a common set of principles and guidelines for public access to scientific data and information internationally. The rationalization of policies and practices across nations, institutions, and disciplines will result in a much greater social and economic

impact from the investment in public research overall by enabling much greater access to and use of scientific data and information resources, and facilitating interdisciplinary and international cooperation in public research and education.

PLAN OF ACTION

This proposed workshop will develop strategic institutional and policy principles and guidelines for broadening the "information commons" for public science and innovation, pursuant to the Statement of Task as outlined in the summary above.

Preliminary Work Plan:
An international ad hoc organizing committee composed of approximately 15 members will be appointed to organize the workshop; this committee will be responsible for producing the resulting guidelines identified at the workshop. The areas of expertise required for the committee include: research policy; information law and policy; intellectual property; information economics; data management; information technology; scientific publishing; international relations; natural sciences; social sciences; and higher education policy. The committee will include well-informed proponents and skeptics of open access approaches. Several commissioned background papers on the issues identified in the task statement will be prepared by experts prior to the workshop, as will a draft set of principles and guidelines for public access to publicly funded scientific data and information.

The workshop, which will be held in early September 2005 at UNESCO headquarters in Paris, France, will bring together managers from science agencies, university administrators, researchers, data and information managers and publishers, and science and information policy experts to discuss and develop elements of the strategy. The organizing committee will identify the workshop objectives, establish an agenda for the meeting, and suggest experts and invitees. The workshop will examine the issues in the context of research in the OECD countries and in the developing world. The collaborating organizations will assist in identifying the workshop objectives, establishing an agenda for the meeting, and suggest expert invitees. The draft set of principles and guidelines will be discussed and adopted in whole or in part at the workshop.

The principles and guidelines from the workshop will be published subsequently by the collaborating organizations, and the workshop presentations will be made available on the CODATA Web site (see http://www.codata.org) and on the Web sites of the other collaborating organizations, as appropriate. The results of the workshop also will be used as part of a larger study being organized by the U.S. National Academies. Both the workshop and the study results will be discussed and disseminated broadly with relevant stakeholder groups, particularly from government science agencies, universities, and research organizations.

Annex Three

INVENTORY OF EXISTING ACTIVITIES FOCUSED ON ISSUES IDENTIFIED IN THE WSIS ACTION PLAN AND THE TOOLS TO CONDUCT IT

The Plan of Action adopted by the World Summit on the Information Society in December 2003 identified many issues relevant to the scientific community, particularly in its section on e-science.. These issues included: high speed internet connectivity for research facilities and universities; use of electronic publishing and open access policies to make scientific information affordable; the collection, dissemination and preservation of essential scientific data; and the promotion of metadata standards to facilitate cooperation.

Many actors in the scientific community have already undertaken actions in these fields and a better understanding of "who is doing what" is necessary to facilitate implementation of the WSIS Action Plan.

Outreach and Inventory Effort

CODATA and its partner organizations will therefore undertake in 2005 an important outreach effort within their respective networks to identify relevant ongoing programs. This inventory will be done in the perspective of the September 2005 workshop on: ***"Moving towards an Information Commons for e-science and Innovation"*** and the Tunis event itself.

To build this inventory, each member organization will be invited to describe its relevant activities and the people responsible for them, in order to raise awareness and facilitate networking.

To minimize the human resources required to conduct this stocktaking exercise, this effort would greatly benefit from an online application that would facilitate:
- gathering the large amount of data in an easy way for all partners; and
- displaying this information and making it available in an easy to search format

In particular, each activity contributing to the implementation of the WSIS Action Plan should be described on a specific page with the appropriate links to information available online and the page describing the organization undertaking it. Indexing along the various themes of the Action Plan will facilitate networking among actors addressing similar issues and facilitate highlighting best practices and/or particularly interesting operations.

In this context, various options are being studied with the objective of identifying the most cost-efficient approach, drawing lessons in particular from the online tools developed for the Geneva Summit itself. Based on the results of this inventory, CODATA and its partners will be able to highlight the most exemplary efforts and best practices.