International Council for Science (ICSU)

CODATA Newsletter 91

Committee on Data for Science and Technology

The Information Commons for e-Science

Creating the Information Commons for *e*-Science: Toward Institutional Policies and Guidelines for Action

by Paul A. David and Paul F. Uhlir

CODATA, along with several other partner organizations, has organized an international workshop at the UNESCO Headquarters in Paris, France on 1-2 September 2005, the purpose of which is to highlight and analyze the variety of experiments that have already been undertaken to enhance the effectiveness of scientific activity in the current transition phase from the print to digital communications media.

July 2005

Highlights

http://www.codata.org

The Committee on Data for Science and Technology (CODATA) was established in 1966 by the International Council for Science (ICSU). Working on an inter-disciplinary basis, CODATA seeks to improve the quality, reliability, management and accessibility of data of importance in all fields of science and technology.

Paul A. David

Paul F. Uhlir

The international scope of digital networks and research collaborations make it both necessary and desirable to seek institutional policies and guidelines for action that will contribute to creating the "information commons" for global e-Science. The workshop aims to promote greater understanding of the variety of successful mechanisms that enhance the availability of public information resources for modern scientific research collaborations. It also seeks to facilitate the development of coordinated principles and guidelines for the rational management of publicly-funded data and information in today's rapidly progressing digitally networked research environments.

The Information Commons workshop will build on the body of practical experience and the empirical studies carried out by the participating organizations and other research and information policy institutions. Moreover, collaboration in this initiative by the major international science policy and scientific information policy organizations-CODATA, ICSTI, INASP, ICSU, UNESCO, TWAS, the OECD and the U.S. National Academies -- has provided an unprecedented opportunity to work towards the formulation of a common, international set of principles and guidelines for public access to scientific data and information.

From a scientific perspective, 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.

(continued on page 2)

Information Commons *(continued from page 1)*

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 cost and time to produce and disseminate additional copies of information in digital form to near zero.

e-Science has been at the forefront of many new paradigms of digitally networked information creation and dissemination activities. Scientific research communities have led efforts to develop open-source software, public-domain data archives and federated data networks, open access journals, community-based open peer review, collaborative research Web sites, collaboratories for virtual experiments, virtual observatories, and Grid-based computing, among other tools for the conduct of distributed research collaborations. These initiatives 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 information produced through public investments in research. Taken together, they are part of the emerging broader movement in support of both formal and informal peer production and dissemination of information in a globally distributed, volunteer, and open networked environment. Such activities are based on principles that reflect the cooperative ethos that traditionally has imbued much of academic and government (civilian) research agencies; their norms and governance mechanisms may be characterized as those of "public scientific information commons," rather than of a market system based upon proprietary data and information.

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.

Open access to the research literature produced from public funding also is a major issue that has received considerable scrutiny in the past few years worldwide. There are now over 1600 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 now are 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 with the needs of developing countries expressly considered, while new open access journals are being established within developing countries themselves.

The adoption of many promising new open access initiatives from the bottom up, coupled with the recent introduction of some new top-down legislative proposals, makes it a particularly appropriate time for a comprehensive review and stock-taking as what has been learned. But here too, there is a lack of strategic planning and concerted implementation of policies by

(continued on page 3)

Information Commons *(continued from page 2)*

government and academic scientific communities. At the same time, the global trend toward the commoditization of public research outputs -- including both 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 participating researchers and their host institutions. 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. It undoubtedly is important to identify those respects in which public policies in the developed and developing countries alike reflect 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. In public policy guidelines and practical implementations alike, progress is likely to be made through searching for workable solutions to “questions of balance” rather than debating “rights” and “wrongs.”

The Information Commons workshop 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, together with the collaborating organizations, have identified the workshop objectives, established an agenda for the meeting, and identified expert speakers and invitees. The workshop will examine the issues in the context of research in the OECD countries and in the developing world. A draft set of principles and guidelines developed in advance of the workshop will be discussed at the workshop, using successful models as exemplars. Three teams of rapporteurs will synthesize the presentations and discussions in the case study sessions, and their summaries will be offered as the basis for discussion in the Workshop's concluding plenary session. The CODATA website will publish material from the individual presentations, as well as background documents and a post-event report on the proceedings by the Co-Chairs of the Workshop.

For more information and to register, please visit the Information Commons workshop website at http://www.codataweb.org/UNESCOmtg.

Featured Speakers include:

Opening Remarks

Paul David, Professor of Economics, Oxford University and Stanford University

Koiichi Matsuura, Director General, UNESCO

Shuichi Iwata, President CODATA

Keynotes

Sir John Sulston, 2002 Nobel Laureate in Physiology and Medicine, Former Director, Wellcome Trust Sanger Institute

Subbiah Arunachalam, Fellow, MS Swaminathan Research Foundation

Presentations

Tony Hey, Corporate VP Technical Computing, Microsoft Corporation (former Director of the UK e-Science Core Programme)

Danny Quah, Professor of Economics, London School of Economics

Elizabeth Longworth, Director, Information Society Division, UNESCO

Michael Spence, Head of Division of Social Sciences, Oxford University Law Board (former Head of the Oxford Law Board)

John Dryden, Deputy Director for Science, Technology and Industry, OECD

Lawrence Lessig, Professor of Law, Stanford Law School

Clemente Forrero Pineda, Professor of Economics, Universities of Andes and Rosario in Bogotá, Colombia

Zhang Xian’en, Director General, Division Basic Research Ministry of Science & Technology, China

Closing Comments

Paul Uhlir, Director, Office of International Scientific and Technical Information Programs, U.S. National Academies

Additionally, a key part of the proceedings will be three parallel "breakout" sessions in which nine case studies will be presented:

Atmospheric Chemistry and Physics (ACP) Journal
Ulrich Pöschl, Chief Executive Editor, ACP

Global Biodiversity Information Facility (GBIF)
James Edwards, Executive Secretary, GBIF

Science Commons
John Wilbanks, Executive Director of Science Commons

Indian Institute for Science
H.P. Khincha, Professor and Divisional Chairman of Electrical Sciences, Division of IISc

CGIAR Nairobi, International Livestock Research Institute
Patti Kristjanson, Senior Agricultural Economist
Global Project Leader: Poverty, Sustainable Livelihoods & Livestock

Open Archive INRIA Activity in e-Science
Anne-Marie Vercoustre, Senior Researcher, INRIA

National Library of Medicine/National Center for Biotechnology (NLM/NCBI)
Elliot Siegel, Associate Director for Health Information Programs Development, National Library of Medicine/NIH

International Seismic Data Network
Igor Chernobay, Chief, Services, Review and Training Section, International Data Centre, CTBTO Preparatory Commission

OpenCourseWare, MIT
Steve Carson, OCW Senior Strategist

In Memoriam
Jacques Emile Dubois
1920-2005

Jacques-Emile Dubois, Emeritus Professor at the University Paris VII, died of cancer in Paris on the 2 April 2005 in his 85th year. Man of action, immensely talented scientist and humanist, attentive and open to others, Jacques-Emile Dubois will leave the image of a scientist at the service of his country, who forced respect by his competence, his passion in action, his determination and his dynamism. Physicist and chemist by training, he promoted multi-disciplinarity and excelled in numerous scientific fields to which his rich life bears witness; an example for many.

Born on the 13th April 1920 in Lille, Ph.D. in Physical Sciences in 1947 in Grenoble, Jacques-Emile Dubois began his career as scientific adviser to the Cultural attaché of the French Embassy in London from 1948 to 1950. Very active in the French Resistance in Grenoble, his university career began as Reader in Chemistry at the University of Grenoble, then Professor of Physical Chemistry and Petrochemistry at the University of the Saar (1949-57), he was Founding Director of the Trilingual Chemistry Institute of that University, then Dean of the Science Faculty there (1953-57). Back in France in 1957, he had the Chair of Physical Organic Chemistry at the Science Faculty in Paris, later the University Paris VII, until 1988 when he continued his activity as Emeritus Professor.

In 1962, he was called to the Ministry of National Education as Scientific Adviser in the Cabinet of Christian Fouchet (1962-63), then became Deputy Director of Higher Education (1963-65), before taking on the position of Director of Research for Defence (DRME) (1965-77).

He assumed many national and international responsibilities, both in public service and in civil society: President of the French Physical Chemistry Society (1974-76), Director of the Biology Division of the Curie Institute (1977-80), member of the Board and of the Directorate of the National Science Research Centre (CNRS) (1963-75) and of the Office of Scientific and Technical Research (ORSTOM) (1963-75), Member of the Scientific Council of the Atomic Energy Commissariat (CEA) (1971-77), Board member of the Palace of Discovery Museum (1961-77), Member of the High Council of Universities (1980-82), Vice-President of the National Centre of Chemical Information (1973-90), Scientific Director of the General Electricity Company (CGE) (1979-1983), Founding Director in 1958 of the Physical Organic Chemical Laboratory that, in 1977, became the Institute of Topology and Systems Dynamics (ITODYS), President of the Interdivisional Committee of Machine Documentation of IUPAC (1969-77), Vice President of CODATA (1981-88), President of CODATA/ICSU (1994-98), then of CODATA France (2000).

In 1993, he was co-founder, with Admiral Lacoste, and Vice President, of the Centre of Scientific Defence Studies (CESD) of the University of Marne-la-Vallée.

Jacques-Emile Dubois was a pioneer of informatics applied to chemistry, in particular by the use of mathematical tools for coding organic compounds; he used the graph theory to propose a coherent ordering system. In 1965, he invented the DARC Topological System (Documentation and Automated Research of Correlations), a structural search system of chemical formulas, with organic compounds coded via matrices and the vectorial representation of Topological-property-reactivity correlations. Among other things, his System is used for similarity searches in chemical structures throughout the organic compound data bases, by the pharmaceutical industry to search for new molecules and to predict their properties, and in data bases for patent research of chemical formulas (Markush DARC system, first commercialised by INPI-Questel).

He is the author of numerous scientific publications (over a 1000 articles and lectures given in International Symposia) in fast kinetics, thermodynamics and chemical informatics; he also edited a number of scientific books, primarily in the field of scientific and technological information.

Jacques-Emile Dubois received many scientific awards: Ramsay Fellow at University College London (1949), laureate of the French Academy of Sciences, of the French and Belgian Chemical Societies, of the Grand Prize of the City of Paris (1975) for the development of the DARC System for Chemical Informatics, Grand Prize of graphic animation for a film at the Festival of Angers (1985), Dr. Honoris Causa of the University of Regensburg, Germany. In 1992, he received the Skolnik Award of the American Chemical Society.

He was awarded the French Resistance Medal and is Commander of the Legion of Honour, the National Order of Merit, the Academic Palms and several foreign Orders. He also received the Aeronautics Medal.

Jacques-Emile Dubois devoted his life to serving an ideal, that of science in all its aspects. He was one of those renowned scientific explorers on the frontiers of several areas, who helped develop the information society of tomorrow.

Call for Proposals to Oversee and Manage the CODATA *Data Science Journal*

CODATA is currently seeking applications from interested organizations or institutions to take on the responsibility of the oversight and management of the Journal. The successful applicant will be offered a contract for this work commencing on 1 January 2006 for a period up to and including 31 December 2007. The anticipated functions and services required from the successful applicant are

- Establishment of an editorial office responsible for processing submissions, coordinating peer review, and ensuring timely online publication.
- Selecting and appointing one or more Editors to manage the editorial process. The formal appointment of the Editor(s) shall be made in consultation with the CODATA Officers.
- Appointment of an Editorial Board to ensure that all publications are properly peer-reviewed and published in a timely manner. The Proposed Editorial Board will be reviewed by the CODATA Officers prior to formal appointment.
- Web design and maintenance for the Journal, including collection and reporting of user metrics.
- Travel for the Editor(s) and editorial office staff if necessary.
- Provision and maintenance of all computer equipment, software and internet requirements for the Journal.
- Independent backups for the publication.
- Indexing of the Journal contents both through its web site and with relevant scientific index and citation services.
- Implementation of a marketing strategy.

The Applicant may contract these services to third party(s); however the applicant is legally responsible to CODATA for all services provided under any sub-contracted agreements

The scope/content of the journal may include but is by no means limited to the following topics;

Data - Data capture, synthesis, analysis, metadata; Data structures; data storage, indexing, retrieval, archiving; Data exchange and sharing; Data display and manipulation; Data dissemination strategies; printing, CD-ROM, internet; Data quality;

Database - Database planning, design, maintenance; archiving; Interfacing databases to other systems, to data products; Database standards; federated databases; Data mining, data science; Visualisation in databases; Use of database packages, commercial issues; Legal issues, intellectual property rights; Financial management, pricing, charging, marketing, selling; e-commerce;

Applications - Industrial applications, industrial requirements; Adding intelligence to data systems; Novel applications; case studies;

Applicants are asked to complete the two-part Application Form (Application plus Budget request) which can be submitted electronically. All applications must be sent in electronic format to the CODATA Secretariat with a copy to the CODATA Secretary General, Dr. Robert Chen. (This happens automatically when you use the online submission application.)

For more information and to apply, please visit http://www.codataweb.org/journalproposals/index.html

The deadline for receipt of applications is Friday 19th August 2005 and a decision shall be announced in September 2005 following an evaluation of all proposals by the CODATA Officers, in consultation with e-journal experts where appropriate.

Why CODATA sponsors the Data Science Journal

No existing electronic Data Science Journal - CODATA *Data Science Journal* is a unique publication that facilitates the development and sharing of new knowledge about all aspects of scientific and technical (S&T) data.

Need to share Data Science results across disciplines - As new ideas are developed in data science and informatics, it is important for researchers, regardless of specialities, to learn about the latest advances.

A Journal gives identity to a discipline - Data Science is emerging as an important component of the information revolution. The electronic *Data Science Journal* provides an international focal point for this work. Experience has shown that a journal helps individual researchers to feel part of and nurture a scientific community.

Increased access to Data Science by member countries – Scientists in many countries find it difficult to access literature, because of high subscription rates and book prices, expensive mail and shipping, and weak libraries. The online *Data Science Journal* provides an important mechanism for free access to up-to-date work on S&T data.

In summary the mission of CODATA's *Data Science Journal* is to provide a high-quality, free, peer-reviewed journal on data science, reaching interested scientists throughout the scientific community.

The *Data Science Journal* is available online at: http://www.datasciencejournal.org/.

CODATA Sponsors a Scientific Markup Languages Workshop

In June 2005, CODATA , along with the National Science Foundation (NSF) and the National Science Digital Library Program, sponsored the ACM IEEE Joint Conference on Digital Libraries Workshop on International Scientific Data, Standards, and Digital Libraries

This workshop, organized by Laura Bartolo, Associate Professor , Kent State University, and John Rumble, Technical Director, Information International Associates and Past President of CODATA, helped to bridge the international scientific/technical data community with the international digital library community in order to form the cyberinfrastructure for scientific data standards and enabling technologies. The ultimate objective of the workshop was to educate standards and digital library developers about models for international S&T information standards appropriate for the cyber-infrastructure and their communities.

The agenda included the following:

Data standards for the International Virtual Observatory (Astronomy)
Robert Hanisch, Space Telescope Science Institute (USA)

Biodiversity data standards
Vivian Hutchinson, U.S. Geological Survey (USA)

Crystallographic Data Interchange Format
David Brown, McMaster University (Canada)

Exchanging technical product data
Sharon Kemmerer, National Institute of Standards and Technology (USA)

Implementing MatML toward sharing materials information on the net
Yoshio Monma, Kochi, University of Technology (Japan)

Geographic mark-up language
David Burggraf, Galdos Systems (Canada)

Standards for the publication of scientific data by World Data Centres and the National Library of Science and Technology in German
Jan Brase, University of Hannover (Germany)

Semantic Web & digital libraries as knowledge systems
Deborah McGuinness, Stanford University (USA)

Data-centric view in eScience information systems
Gregor Erbach, German Research Center for Artificial Intelligence (Germany)

Scientific mark-up languages for uplifting a digital library to knowledge archive
Toshihiro Ashino, Toyo University (Japan)

MathML and digital libraries
Timothy Cole, University of Illinois at UC (USA)

Understanding and using standards
Robbie Robson, EduWorks (USA)

Scientific discovery and large-scale databases
John Rumble, Information International Associates (USA)

For more information and to view the presentations, visit http://scimarkuplang.comm.nsdl.org/cgi-bin/wiki.pl?SeeAgenda.

Upcoming Events

ISGI CODATA Symposium, Berlin, Germany, 14-16 September 2005.
http://www.codata-germany.org/ISGI

Multimedia. Where Do We Go From Here? CODATA International Symposium, September 19-20, 2005, European Academy, Berlin, Germany
http://www.codata.org/announcements/Berlin_2005_CfPF.htm and http://www.codata-germany.org/MIST_2005/index.html
The call for papers has been extended to July 31st.

Ensuring Long-term Preservation and Adding Value to Scientific and Technical data (PV 2005), 21- 23 November 2005, Royal Society, Edinburgh, UK
http://www.ukoln.ac.uk/events/pv-2005/

International Conference on Preservation of Digital Objects (iPRES), 15-16 September 2005, Göttingen, Germany
http://www.langzeitarchivierung.de/ipres/

n-ABLE 2005, Saarbruecken, Germany, 26-28 September 2005
http://www.apnf.org/ocs/themes/nABLE05/front1.html

International Symposium on Geotechnical Hazards: Prevention, Mitigation and Engineering Response, Yogyakarta, Indonesia, 28 November – 1 December 2005
http://www.geotek.lipi.go.id/events/
(Click on this event on the left side)

36th COSPAR Scientific Assembly and Associated Events, 16-23 July 2006, Beijing, China
http://www.copernicus.org/COSPAR/COSPAR.html
Abstract Deadline: 17 February 2006

CODATA Newsletter Editor: Cheryl Levey

CODATA Secretariat
51, Boulevard de Montmorency, 75016 Paris, France
Tel: +33 1 45250496 / Fax: +33 1 42881466
Email: **codata@dial.oleane.com**
Website: **http://www.codata.org**