

CODATA

Newsletter 82

Committee on Data for Science and Technology

MONTREAL
CANADA
29 SEPT. - 3 OCT. 2002

CODATA 2002
18th INTERNATIONAL CONFERENCE

CODATA Announces its 18th International Conference

“Frontiers of Scientific and Technical Data”

October 2001

Highlights

- 1 CODATA 2002
- 2 (3) CODATA 2002
Call for Papers
- 3 International Scientific
Program Committee
- 4 Conference Location
CODATA Prize 2002
- 5 (6) CODATA Electronic
Journal
- 7 (8) General News

<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 interdisciplinary basis, CODATA seeks to improve the quality, reliability, information and knowledge processing, management and accessibility of data of importance to science and technology.

The 18th International CODATA Conference, *Frontiers of Scientific and Technical Data*, takes place 29 September - 3 October 2002 at the Hotel Delta Centre Ville, downtown Montreal. This four-day Conference is hosted by the Canadian and US National Committees for CODATA.

CODATA 2002 continues CODATA's 36-year tradition of serving international science by holding an open and exciting conference on the latest advances in scientific and technical data.

CODATA 2002 features six keynote plenary lectures addressing cross cutting themes, focussing on issues such as:

- Preservation and Archiving of Scientific and Technical Data
- Interoperability and Data Intervention
- Emerging Tools and Techniques for Data Handling
- Legal Issues in Using and Sharing Scientific and Technical Data
- Information Economics for Scientific and Technical Data
- Ethics in the Creation and Use of Scientific and Technical Data

CODATA 2002 involves twenty-four invited speakers from different scientific disciplines identifying and expanding on the themes presented during the plenary sessions. In addition, there are six parallel sessions, with opportunities for 160 - 190 contributed papers.

CODATA 2002 involves poster sessions, technical demonstrations and cross-disciplinary workshops focussing on basic data issues.

CODATA 2002 is designed to excite the scientific community with new ideas in data science and expose data specialists to the complexity and variety of scientific data needs among their colleagues in other disciplines. To achieve this objective an International Scientific Program Committee is established to ensure the high quality and interdisciplinary focus of **CODATA 2002**.

I invite you to read through this Newsletter to find out more about the Call for Papers and general Conference information. This information can also be consulted on our website <http://www.codata.org> which will be continually updated.

On behalf of the International Scientific Program Committee I look forward to your participation at this event in the picturesque city of Montreal!

Dr. Gordon Wood,
Co-Chair of the International Scientific Program Committee, 2002



CALL FOR PAPERS

CODATA 2002 features six keynote plenary speakers, twenty-four invited speakers and six parallel sessions with 160 - 190 contributed papers.

Contributed papers are currently been solicited in the following areas:

Physical Science Data

- Chemistry
- Physics
- Crystallography
- Astronomy
- Materials
- Metrology
- Energy
- Geothermy
- Gas Hydrates
- Analytical

Earth and Environmental Data

- Geology
- Geography
- Geochemistry
- Climate/Global Change
- Water and Ocean
- Land Use/Cover
- Urban/Agriculture
- Natural Disasters
- Remote Sensing

Social Science Data

- Demography/Statistics
- Psychological
- Political Science
- Behavioral
- Culture Collections
- History
- Archeology
- Anthropology

Data Science

- Errors
- Data Quality
- Statistics
- Ontologies
- Semantics
- Taxonomies
- Analysis
- Modelling

Technical Demonstrations

Biological Science Data

- Biochemistry
- Molecular Biology
- Microbiological
- Genomics
- Biodiversity/Environment
- Bioinformatics
- Structural Biology

Medical and Health Data

- Pharmacology
- Immunology
- Epidemiology
- Health Services
- Nutritional/Food

Informatics and Technology

- Archiving
- Data Management
- Interoperability
- Metadata
- Data Tools (GIS, Browsers, Multimedia)
- Data Mining
- Visualization
- Virtual Laboratories
- International Data Networks
- Multi-media Databases

Data Policy

- Ethics
- Legal Aspects
- Acces to Government Data
- Information Economics

Large Data Projects

- Astronomy
- Chemistry
- Genomic
- High-energy Physics
- Earth Remote Sensing
- Social Sciences

Multi-disciplinary Data Projects

DEADLINE FOR SUBMISSIONS



CODATA 2002

International Scientific Program Committee

SUBMISSION FORMAT

Abstracts should include:

- Title of proposed paper
- Author(s) name, title, address, e-mail, phone, and fax
- Classification of the paper by one or more of the designated topics listed Call for Papers on the left (For example, a single specific paper might logically be grouped with other presentations under Microbiological or under Archiving. Talks on multi-disciplinary topics are encouraged.)
- 200 to 400 word abstract in English
- Preference for oral presentation, poster presentation or technical demonstration

Notes:

1. Most oral presentations will be in parallel sessions.
 2. Posters will go on public display Monday morning and come down Wednesday evening; authors should be available to explain them at the Monday evening reception.
 3. Technical demonstration sessions involve typically six tables around the perimeter of a room. The six presenters will be given a maximum of 5 minutes each at the beginning of the session to describe their demonstration orally. The remainder of the session will be for simultaneous walk around demonstrations. Technical demonstrations must be self-contained on one's laptop and viewable either on the laptop screen or via one's own projector. Internet access will NOT be provided.
- Preference is for abstracts submitted by e-mail in non-formatted text. Submissions should be contained in an e-mail message, NOT as an attached file. In the subject line of the message, place **CODATA 2002 - <your last name>**

All submissions should be sent to:
codata@dial.oleane.com
(with a copy to **gordon.wood@nrc.ca**).

Co-Chair:

Dr. Gordon Wood, Canada

Professor Harlan Onsrud, USA

Members:

Dr. Marcelle Gaune-Escard, France

Physical Science Data

Dr. Takashi Kunisawa, Japan

Biological Science Data

Dr. Chuang Liu, China

Earth and Environmental Data

Dr. Elliot Siegel, USA

Medical and Health Data

Dr. David Johnson, USA

Social Science Data

Mr. Glen Newton, Canada

Informatics and Technology

Dr. Paul F. Uhler, USA

Data Policy

Professor Heston Phillips, South Africa

Technical Demonstrations

Dr. John R. Rodgers, Canada

Data 2050

Dr. Aleksandar Jovanovic, Germany

Engineering Data

28 FEBRUARY 2002

CONFERENCE LOCATION

The 18th International CODATA Conference takes place at the Delta CentreVille Hotel, located in the heart of the city of Montreal. Considered the “Paris of the North”, Montreal is the second largest French-speaking city in the world but welcomes English everywhere. It is a fascinating city, a multicultural medley of European ambience and North American style. Situated on an island in the middle of the majestic St. Lawrence River, Montreal is a beautiful blend of modern and historic buildings, tree lined boulevards and cobblestone streets. More detailed information on Montreal and all it has to offer can be found on:

<http://www.tourism-montreal.org>

The Delta Centre Ville is a 711 guestroom hotel with excellent conference facilities. The building is located in the heart of downtown Montreal. A direct passageway links Delta Centre-Ville to the subway and the underground city. It is only steps to Old Montreal, the Cité Multimedia and is a short walking distance to the Molson Centre and Place Bonaventure. The hotel is approximately 22km from Dorval International Airport and is accessible by taxi, limousine or bus. Bus departs from the airport every 20 minutes and arrives at the downtown bus terminal, which is only five minutes walk from the hotel. The bus ticket costs approximately 9 CAN\$ (subject to change). More details on the hotel and the facilities it has can be found on:

<http://www.deltahotels.com>

The Hotel Reservation Form for the Conference is available on the web **<http://www.codata.org>**. A limited number of rooms have been reserved at a special rate. To take advantage of this rate people are advised to reserve as early as possible.



Revolving rooftop restaurant of Hotel Delta Centre Ville.

CODATA PRIZE 2002

The Call for Nominations for the CODATA Prize 2002 was launched in September 2001.

This award acknowledges outstanding achievement in the world of scientific and technical data (S&T) and it recognises work carried out in fields of importance to S&T data, such as data management, evaluation, dissemination, access issues, international cooperation, knowledge discovery, archiving or related subjects.

The award was first launched at the 17th International CODATA Conference in Italy in October 2000.

Its first recipient was Dr. Barry Taylor of the National Institute of Standards and Technology (United States), for his major contributions to the advancement of our understanding of the physical world through critically evaluated values of the fundamental physical constants.

The next award will be presented at the 18th International CODATA Conference September 29-3 October 2002 in Montreal, Canada. The recipient of the award will be presented with a specially commissioned piece of sculpture and will also receive financial support to travel to Canada to deliver an awards lecture at the Conference.

The Nomination Form can be found on the website **<http://www.codata.org>**

Original Nomination Forms must arrive at the CODATA Secretariat,
51 Bld de Montmorency,
75016 Paris,
no later than Friday 1 March 2002.

I look forward to your enthusiastic participation in the search for the best candidate for our Prize

Professor Akira Tsugita
Chair, CODATA Prize Committee 2002



New CODATA Electronic Journal on Scientific and Technical Data

Background

At the 46th Executive CODATA meeting, held in February 2001, a decision was made to publish a new quality journal on Scientific Data and Databases. To reach scientists and engineers in all countries, it will be an electronic journal, published on the internet, and for now, free to all.

The decision to start the journal was taken after a great deal of discussion commencing at the 43rd Executive Committee meeting in May '98. It was also discussed at the 21st General Assembly in India '98 and at the 22nd General Assembly in Italy 2000. Since its establishment in 1966, CODATA has commissioned several books; it has also hosted a biannual conference, numerous seminars, colloquia and meetings the proceedings of which have been published in a series of books and reports, many edited by the previous Executive Director, Phyllis Glaeser. These have included over 5000 papers, all of high quality, and most of them have not been published elsewhere.

Although these CODATA books are available in many major libraries, they are not found in every library and are therefore not available to a large number of scientists even in developed countries. Although a study of the content of the CODATA books, conference proceedings and reports shows that the quality of the papers is high, they do not have the same academic standing as a refereed journal in spite of the excellence of the papers. The new electronic journal should solve both of these problems, making the papers universally available but with high quality ensured by rigorous refereeing.

History

The new journal will be the first journal published by CODATA and the decision reflects the growth and maturity of the discipline since CODATA was founded over 30 years ago. At that time most data consisted of a limited number of facts and quantities, mostly found in text books, handbooks or journals. However the first databanks (the concept of "database" was not yet being used) of scientific data were being compiled with the help of early computers and the study of the problems encountered in the computerization of these data collections led to the creation of CODATA. Today, over 30 years later, the tremendous growth of com-

puter technology is reflected in today's PCs, which can store and process Gigabytes of data. Perhaps a bigger revolution is the development of the internet which can make data instantly available across the globe and will make our new journal also accessible in seconds to all scientists in all countries. These concepts could hardly have been conceived by those with the vision to start CODATA at its foundation.

Changing nature of data

Clearly the nature and scope of much scientific and engineering data, and in consequence of much of scientific research has changed. Measurement technologies have improved in quality and quantity with measurement times reduced by orders of magnitude. Virtually every area of science: astronomy, chemistry, geosciences, physics, biology and medicine is becoming model based, dependent on large bodies of data held in scientific databases. In the same way modern engineering design depends on large scientific databases on the properties of materials.

All these advances point to an ever-increasing need for more data, better access to them and higher data quality, stored in large interrelated databases. Already thousands of databases deliver a rich variety of scientific information including numeric data, graphs, charts, images, maps, videos, expert systems and full text of research articles, abstracts and indexing information. Almost all of science and engineering in the coming millennium will depend on the correct and imaginative design and implementation of these scientific databases, all available on the internet.

Scope of the journal

All this activity has naturally begun to generate a lively body of literature, growing rapidly in volume, detail and breadth. But it is still in its infancy. The study of these databases is itself a large subject covering the data themselves, the databases and their application. Data science includes the study of the capture of data, their analysis, metadata, fast retrieval of data in a large database or across the internet, archiving of data so that they are available for future generations, exchange of data and quality of data. Database research includes interfacing data, data mining to find unexpected data and relationships, and visualization of data in two and three dimensions including movement. It also includes data and database management including intellectual property rights and other legal issues. Adding intelligence, still in its infancy, will be a growing subject with long term benefits that are hard to predict but may be most significant of all. So it is clear that scientific databases will have a key role in all science and engineering in the coming millennium. It follows that this vital subject will need at least one quality journal to aid the study of all aspects of scientific data and databases. These considerations lead to the decision to create the journal by CODATA; its scope is defined in the list below.

continued from page 5

Data

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

Database

Database planning, design, maintenance; archiving;
Interfacing databases to the internet;
to other systems, to data products; interoperability;
Database standards; compatibility; federated databases;
Data mining, data science;
Human-computer interfaces; visualization in databases;
Use of database packages, commercial issues;
distributed databases;
Legal issues, intellectual property rights; data policy;
data access;
Financial management, pricing, charging, marketing,
selling; e-commerce;

Applications

Industrial applications, industrial requirements;
Adding intelligence to data systems, data modelling;
Novel applications; case studies;
interdisciplinary systems

The journal's focus will be on the above data management and delivery issues that confront builders and suppliers of scientific and engineering databases, rather than on the contents of specific databases. The journal is designed to foster communication on these issues and to convey new insights and progress to the scientific data community, regardless of discipline.

Electronic publishing

If the journal had been started any time up to a few years ago it would have been a traditional paper journal that would have been costly to publish and would have been expensive to buy. Only the major libraries and a few scientists would have been able to afford it. Today we have the option to bypass the conventional costs of publishing on paper and to publish on the internet. To make it freely available to all it will be necessary to keep the costs low; so authors will be required to compose their own papers according to the format of the journal which is explained on the journal web site (<http://journals.eecs.qub.ac.uk/codata/prospectus.html>). This is not difficult to do with modern word processing systems and is the normal procedure for most conference proceedings today and many journals. There are no staff employed by CODATA to compose papers; so authors are asked to submit papers by email, not by post, in the format defined by the website, i.e. in the same form as they will be published.



Professor Jack Smith.

The first few papers have already been received in this format without any difficulties expressed by the authors. The papers have been sent electronically to the first referees since all papers will be refereed by at least two referees. The executive is in the process of appointing an international editorial board covering all disciplines covered by the journal. So the journal is about to become a reality; but it will only become as successful as the papers that are published in it. So the readers of this article are invited to submit relevant papers to the new journal, possibly important papers presented at previous CODATA conferences.

The first issue

Following the decision to start the journal I was asked to be the first Editor-in-Chief and a part-time Assistant Editor, Maria Husin was appointed. The executive of CODATA has begun the process of inviting scientists and engineers from a wide range of disciplines to join the international editorial board. The first few papers have been received and are already with referees. We now look forward to receiving your papers (which might be based on papers you have presented at previous CODATA conferences). Details on how to submit can be found on the website <http://journals.eecs.qub.ac.uk/codata/prospectus.html>. It is planned that the first issue will be early in 2002.

Professor Jack Smith,

Editor-in-Chief

General News

Call for New Task Groups 23rd CODATA General Assembly 2002

As we are now approximately within one year of the next CODATA General Assembly, 4-5 October 2002, Montreal, it is time to commence preparations for proposals for New Task Groups as well as applications for renewals of existing Task Groups. Applications for New Task Groups and Renewals should be submitted on the appropriate forms. These are available electronically on the CODATA website <http://www.codata.org>

Conventional forms are available from the CODATA Secretariat.

Please be aware that CODATA activities are increasingly judged by deliverables of explicit short term as well as long term benefits to information and knowledge management in science and technology. In the evaluation process of the quality of proposals these aspects are going to be important. Groups able to identify additional sources of support will strengthen their case. Please also note that Task Groups are encouraged to submit high quality manuscripts for possible publication in our new CODATA Journal. All submissions to the journal are subject to a refereeing process.

To facilitate refereeing and review process of the Task Groups, we ask your cooperation in ensuring that the completed form is received at the secretariat by 31 January 2002.

Professor Paul G. Mezey,
Secretary General, CODATA

Curation of Databases in Molecular Biology

CODATA was host to a European Science Foundation course in Functional Genomics: Curation of Databases in Molecular Biology, 11-14 October 2001. This course was part of the ESF Program "Integrated Approaches for Functional Genomics"

<http://www.functional.genomics.org.uk>

Participants included 9 invited speakers and 22 students. Two factors dominate current developments in bioinformatics: the great increase in the amount of raw data, and the requirement, for successful application of these data to research, of carefully and continuously curated databanks.

Annotation is a weak component of the enterprise. This is partly because much essential information in databanks is inferred rather than measured. The identification of genes in genome sequences is a prime example; for researchers analysing the protein sequences implicit in genomes, errors in gene assignment vitiate the high quality of the sequence data themselves.

In the formal lectures and the discussion sessions, the problems of annotation were presented, and some principles for solutions were suggested. The meeting was organized by A. Lesk, J. Garnier, and M. Helmer-Citterich. M.J. Taussig and A. Martin have overall responsibility for the ESF Program in Functional Genomics.

CODATA Working Group on Biological Collection Data Access

A joint CODATA and TDWG initiative

www.bgbm.org/TDWG/CODATA/

Report from the first workshop, "Building blocks for common access to biological collections" at the National Center for Ecological Analysis and Synthesis (NCEAS) in Santa Barbara, California, June 11 - 13, 2001.

Group members include representatives from existing or emerging distributed query systems, portals to multiple biological collection data providers which enable a user to formulate a query against a simple generic concept of a collection unit, to broadcast the query to data providers and return to the user a single structured data set for viewing on-line or downloading. The working group's prime aim is to develop a content-oriented XML Schema for biological collection units (including simple field observations that did not produce voucher specimens) which serves as a reconciled set of element names and semantics for software designers, scientists and curators to use. The data specification is to be both comprehensive and general, including a broad array of concepts that might be available in a collection database, but to mandate only the bare minimum of elements required to make the specification functional. This track of the workshop ended with the creation of a rough, but mostly complete, hierarchical structure of data elements as the base for further development.

Taking advantage of the expertise in the group, a second focal point was the development of a common software architecture to support distributed queries across collection databases. Its specification will be cast as a "search" protocol <http://www.gils.net/search.html>, which combines elements of SOAP, ANSI Z39.50, and UDDI. Portals will broadcast queries to providers as XML documents, in which the elements of the query have been "marked up" as XML



IMPORTANT DATES IN THE CODATA CALENDAR

31 January 2002

Deadline for receipt of applications
for New CODATA Task and Renewal of
existing Task Groups

31 January 2002

Deadline for Call for Papers, CODATA
Workshop on Information Visualization
and Presentation,
Museum National d'Histoire Naturelle,
Paris

28 February 2002

Deadline for receipt of Nominations
for CODATA Executive Committee

28 February 2002

Deadline for Call for Papers, 18th
International CODATA Conference,
2002

1 March 2002

Deadline for receipt of Nominations
for CODATA Prize 2002

11-15 March 2002

Second International CODATA Africa
Workshop, Dakar, Senegal

15 April, 2002

Deadline for Call for Papers,
1st CODATA Brazilian Conference
on Data and Information in Science
and Technology,
Instituto Nacional de Tecnologia,
Rio de Janeiro

22-24 April 2002

CODATA Workshop on Information
Visualization and Presentation
Museum National d'Histoire Naturelle,
Paris

14-16 August 2002

1st CODATA Brazilian Conference
on Data and Information in Science
and Technology,
Instituto Nacional de Tecnologia,
Rio de Janeiro

29 September - 3 October 2002

18th International CODATA Conference

4-5 October 2002

23rd CODATA General Assembly

elements. Providers will convert the query into SQL or other native query language, as required, and pass it to the local database. The provider will then take the result set returned from the database, and return it to the portal as an XML document. The portal will then merge the documents returned from multiple providers into views or data sets of interest to the user. As a result, portals and providers will be de-coupled, and both portal and provider software can be built in modules, using an open-source model. Compliance with the data-related part of the schema will ensure compatibility. Please refer to the website for further details and current developments.

Dr. Walter Berendsohn,
Chair, Working Group-Biological Collection Data Access

International CODATA Workshop on Information and Infrastructure for Science and Technology

A CODATA sponsored Workshop on Information Infrastructure took place in Zurich 12-13 October 2001. It was organized as part of the International Symposium of Environmental Informatics at Zurich Technical University (ETH Zürich). Participants from 8 countries discussed infrastructure aspects of access, interoperability, use and archiving of data, as well as aspects of quality control, legal aspects and education.

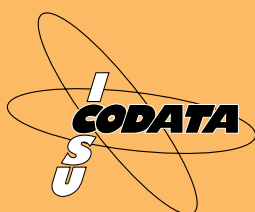
Six presentations produced important key discussions on these aspects. The complexity of the material covered highlighted very well the fact that infrastructure-reliable data, documentation, functions, services etc- is a key issue in present and future interdisciplinary contributions to decision-oriented information systems. The aspects (physical) of network infrastructure seemed to the participants to be sufficiently covered by specialised organizations. For CODATA, the aspects of Logical Information Infrastructure can be a field of interest because of the variety of disciplines involved, the demands on local, national, and international development, the broad need of awareness rising on the methods and techniques involved and for the expected benefits to the information society.

Workshop Chairman: Dr. Horst Kremers (Horst_Kremers@compuserve.com)

Some of the participants



From left to right: Abdoulaye GAYE (Member of the CODATA Executive Board), Kit BUURMAN, Reiner KRAUSE, Pravir CHAUDHRY, Reine KARLSSON, Tatyna SOBCHUK, Horst KREMERS.



Editor : Kathleen Cass
CODATA Secretariat
51, bld de Montmorency
75016 Paris

Tél. : 33 1 45 25 04 96
Fax : 33 1 42 88 14 66
e.mail : codata@dial.oleane.com
<http://www.codata.org>