

Report on Data Activities in Canada 1994

Prepared by
the Canadian National Committee for CODATA (CNC/CODATA)

The following report on data activities in Canada was presented to the 19th General Assembly of CODATA at Chambéry in September 1994. To obtain further details on individual items or to submit information on other Canadian data activities for inclusion in the next report (September 1996) please contact:

Le rapport ci-joint, qui fait état des activités du Canada en matière de données, a été présenté à la 19^e assemblée générale de CODATA, à Chambéry, en septembre 1994. Pour obtenir de plus amples renseignements sur des points particuliers ou pour soumettre de l'information sur d'autres activités canadiennes sur les données aux fins d'insertion dans le prochain rapport (septembre 1996), veuillez communiquer avec:

Secretariat, CNC/CODATA
CISTI, Building M-55, Rm 275
National Research Council
Montreal Road
Ottawa, Ontario K1A 0R6

Secrétariat CNC/CODATA
ICIST, Édifice M-55, bureau 275
Conseil national de recherches
Chemin Montréal
Ottawa (Ontario) K1A 0R6

Telephone: (613) 993-3294
Fax: (613) 952-8246
Internet: codata@nrc.ca

Téléphone : (613) 993-3294
Télécopieur : (613) 952-8246
Internet: codata@nrc.ca

(For a copy of the report in French, please contact the Secretariat.)

(Pour obtenir la version française du rapport, veuillez communiquer avec le Secrétariat.)

19th General Assembly of CODATA, Chambéry, September 1994

Report on Data Activities in Canada

Activities in Canada, as known to the Canadian National Committee for CODATA (CNC/CODATA), are reported below in the categories shown. Further information on some items may be obtained from the contacts cited at the right margin and identified in the Appendix.

Biological Sciences

1. Molecular Biology

Since 1989 the Molecular Biology Database System has been available for online use on the Canadian Scientific Numeric Database Service (CAN/SND) operated by the National Research Council of Canada (NRC). Recently moved to a powerful new UNIX platform, the system is available via packet-switched networks and the Internet worldwide. Some of the features offered are:

2. Databases

- EMBL - European Molecular Biology Data Library
- GenBank - Genetic Sequence Data Bank
- SWISS-PROT - Swiss Protein Sequence Database
- NRL-3D - Naval Research Laboratory
- CRYSTPRO - Brookhaven Protein Data Bank
- OMIM - Online Mendelian Inheritance in Man
- GD - BHuman Genome Database - Johns Hopkins
- FLYBASE - Drosophila

3. Search & Analysis Systems

- GCG - University of Wisconsin Genetics Computer Group
- STADEN - Staden Package
- Phylip - Phylogenetic Analysis Package
- ReadSeq - Sequence File Format Conversion Program
- Prosearch - Prosite Database Pattern Search Program
- LINKAGE - Lathrop and Lalouel
- PIR - Protein Identification Resource
- FASTA - Lipman-Pearson Search Algorithm

4. CODATA Hybridoma Data Bank

The latest release is searchable online on the CAN/SND service. Searches may be made on various indexes (author, source, reactant, distributor, etc.) and boolean operations may be used to refine and enrich the query.

5. *Sulfolobus* Genome Data

A group of workers at the University of Ottawa (Dr. R. Charlebois, Department of Biology), Dalhousie University (Dr. F. Doolittle, Department of Biochemistry) and the NRC Institute for Marine Biosciences (IMB, Dr. M. Ragan), under a grant from CGAT (Canadian Genome and Analysis Technology Program) have begun work on the genome of the *sulfolobus* bacterium. The IMB is responsible for most of the data handling and analysis.

6. Indices of Available Fungal Cultures

Produced by the Nova Scotian Institute of Science, these indices are lists of cultures available from culture collections and include the following details of each culture: binomial name, accession number, substrate, place of origin of the fungus as well as details of its maintenance and toxicity. Cultures covered include at least seven Canadian collections, with an aggregate of about 14,000 cultures, as well as those available from the International Mycological Institute in the UK with about 9200 cultures.

7. Fungal Metabolites

Also produced by the Nova Scotian Institute of Science, this database, with coverage from 1789 to 1993, includes the binomial names of the producing organisms, the name (trivial or systematic) of the metabolite, its molecular formula and a literature reference giving details of the method of isolation of the metabolite

Chemistry

1. *MEDLA* Molecular Shape Database

The Molecular Modeling Group in the Department of Chemistry, University of Saskatchewan has developed a substantial molecular shape database of standard and distorted sets of molecular electron density fragments. The database is designed to be used with the Molecular Electron Density Assembler (*MEDLA*) method for building molecular electron densities of small and large inorganic and organic molecules along with various polymers including polypeptides and proteins.

2. Database for Non-Carcinogenic Toxicity of Poly-Aromatic (PAH) Molecules

Leadership is being given by the Universities of Waterloo, Montreal and Saskatchewan in collecting data on toxic effects, other than those related to cancer, of PAH molecules. Examples of such effects are those of photo-oxidized products of PAHs on plants, fish and other species.

Crystallography

1. NRC Metals Crystallographic Database (CRYSTMET)

Work was completed on adding retrospective entries to the database making it exhaustive in coverage to 1913 and containing over 52,000 entries. The database may be licensed for private or multiple use and it is also available online via the CAN/SND and STN services.

2. Inorganic Crystal Structure Database (ICSD)

Through an exchange agreement between NRC and the FIZ Energie, Physik, Mathematik (Karlsruhe) the ICSD continued to be made available online on the CAN/SND system and CRYSTMET continued to be made available online on STN.

3. NIST Crystal Data File (CRYSTDAT)

Under an umbrella arrangement between the two organizations, NRC and NIST continued to collaborate on the production and enhancement of the Crystal Data File known as CRYSTDAT on the CAN/SND system. This collaboration has produced software tools to address some of the research needs of materials science, particularly in the areas of materials design and identification. Crystal Data now contains over 180,000 entries.

4. Brookhaven Protein Data Bank

NRC continued to be one of the many sites offering network access to this important data collection.

5. Online Access

The CAN/SND system continued to offer public, international online access to the complete suite of crystallographic databases both via the Internet and the X.25 packet-switched networks. The databases available online are:

- CRYSTDAT - NIST Crystal Data File
- CRYSTIN Inorganic Crystal Structure Database
- CRYSTMET - NRC Metals Crystallographic Database
- CRYSTOR - Cambridge Structural Database
- CRYSTPRO - Brookhaven Protein Data Bank

Geoscience

1. Standards

In the area of geoscience standards, the Canadian General Standards Board (CGSB) Committee on Geomatics has adopted both the Spatial Archive and Interchange Format (SAIF) and the Digital Geographic Information Exchange Standard (DIGEST) as National Standards of Canada.

The Surveys and Resource Mapping Branch, British Columbia Ministry of Environment, Lands and Parks, submitted a series of four papers on behalf of Canada for consideration by the International Standards Organization (ISO) Database Language Multimedia Working Group. Included in this submission was a framework for the development of Part 3 of the Spatial Query Language/Multi-Media (SQL/MM) based directly on the SAIF standard. This proposal was accepted and will be the basis for future ISO work regarding spatial/temporal data management in SQL/MM.

The Committee on Geomatics working group on feature cataloguing has taken the FACC (Feature Attribute Coding Catalog) as a starting point for features and attribute coding and has harmonized the National Digital Topographic Database and the provincial topographic database objects through a one-to-one relationship. Work is currently being done to standardize directory information describing geo-referenced data sets.

Future development in the area of standards is being driven by the concept of Open GIS using virtual data models for access.

2. Database Access

Within the Canadian government, the GIS Division of the Surveys, Mapping and Remote Sensing Sector has developed a federal multidatabase management system that provides interoperability between different GIS. Known as the Delta-X, the implementation is based on the assumption that each underlying DBMS is based on a client-server architecture and each client workstation is connected to a network that is configured with access to at least one DBMS server. A client, besides being able to query and access a server database, is typically configured as a geographic information system's workstation. Furthermore, as a front-end to the Delta-X, a spatial data browser facilitates the access to metadata of various databases, e.g., information on specific datasets, ownership, geographic coverage, format, availability, etc.

Access to sources of geographic data has improved within the federal government. The thematic map databases of most departments have been converted to digital format and are stored and structured using geographic information systems. Cross-indices between thematic databases are being developed. Data describing Forestry, Agriculture and Environment across Canada can now be used in an integrated way at national scales.

3. Digital Chart of the World (DCW)

The DCW is a huge vector base map of the world at 1:1000000 scale including cartographic, attribute and textual information. It comes on four CD-ROMs with extraction, display and query software (VPFVIEW). Developed for a multinational project involving the Canadian Directorate of Geographic Operations (DND), the U.S. Defense Mapping Agency and the equivalent groups in Britain and Australia, it includes data on 17 thematic layers including political boundaries, ocean coastlines, cities, transportation networks, drainage, land cover and elevation.

4. New Projects related to High-Speed Network Developments

The Canadian Network for the Advancement of Research, Industry and Education (CANARIE) is a government supported not-for-profit corporation dedicated to the promotion and advancement of networking and networking technologies in Canada. Phase 1 of the CANARIE implementation plan (June 1993 to March 1995) includes upgrading the National Research and Development and Education Network, establishing a national high-speed testbed network and initiating product and service developments that would utilize that network. Geoscience projects that have been approved under this program by Canadian industry include "Chartnet" and the "CARIS wide-area data browser".

Chartnet will be an integrated suite of software systems and processes for the collaborative production, maintenance and distribution of electronic charts in a high-speed wide area network (WAN) environment. Electronic chart products are derived from very large databases of spatial/temporal hydrographic source data. The lead contractor for this project is Nautical Data International Inc.

The **CARIS wide-area data browser** will involve the development and testing of spatial data and delivery software for broad-band wide area networks. This will improve the collection, management and distribution of geographically related information in electronic form. This project will examine the effects of broad-band communication services on the delivery of digital property mapping and image data to end users. The lead contractor for this project is Universal Systems Ltd., in partnership with telephone companies and universities in eastern Canada.

Environment

Global Change

1. GCNet (Online Global Change Information)

GCNet (formerly the Global Change Network) was developed at the Canada Centre for Remote Sensing (CCRS) to serve as a single point of contact for global change researchers, scientists and

users of remote sensing information. It is a free online system that directs users to pertinent international datasets and other up-to-date information. The following information is now available through GCNet:

Directory Service

Users can access a centralized directory of scientific data sets which identifies Canadian and International data sets pertinent to global change research. Directory access is done through the Master Directory which is part of the International Directory Network (IDN). IDN nodes are currently located in Canada, France, Germany, Italy, Japan, Russia, the United Kingdom and the United States.

Data Centre Links

Permits users to link to other data centres world-wide and access their inventories, bulletin boards or information networks. This feature is part of the IDN directory service.

CCRS Image Inventory

This option invokes the CCRS Query program, which permits searches of the LANDSAT, MOS, NOAA and SPOT satellites' raw image inventories. A products catalogue of NOAA geocoded and composite products processed on the GEOCOMP system is also available. Results can be viewed on screen, mailed via Internet, NSI/DECNet or sent by surface mail. An ERS-1 image inventory will also be available soon.

CCRS Bulletin Board

The CCRS Bulletin Board PlaNet contains detailed information about CCRS activities and profiles Canadian Companies, regional centres and educational institutions involved in remote sensing.

SMRSS Products and Services

Users can scan a complete list of products and services offered by the Surveys, Mapping and Remote Sensing Sector (SMRSS) of Energy, Mines and Resources Canada. This includes detailed information such as product descriptions, prices and order contacts in Canada for all digital products normally used in the geomatics field.

RESORS

This option provides users with information on how to get an account of the CCRS document retrieval service, RESORS. RESORS is a unique online bibliographic database that provides

rapid and precise access to information on the technologies and applications of remote sensing world wide.

2. Hydrologic Data

Hydrologic data (quality and quantity) are collected by various federal, provincial agencies and industries but the quality and quantity data are managed differently. All stream flow information is managed by the federal government using a relational database HYDAT; these data are published annually on CD-ROM.

The water quality data are generally managed by the agency that collected them. Environment Canada and all provinces have water quality databases. Under Federal/Provincial Agreements some data have been transferred between agencies. The ENVIRODAT database is used by Environment Canada to manage these types of data; some provinces have expressed interest in using the ENVIRODAT system.

With the reorganization in Environment Canada the hydrologic databases and climate databases are now managed by the Climate Information Branch, Atmospheric Environment Service. Over the next four years they are planning to integrate the data models and distribute the databases to improve access.

3. Databases for Environmental Analysis: Government of Canada

An inventory of over 370 Government of Canada Databases useful for environmental reporting. It lists the purpose, contact information and included variables for each database. The book includes a diskette copy with keywords for automated searching.

4. SEDTEC (Sediment Treatment Technologies Database) 2nd ed.

This database documents 210 treatment technologies worldwide for the treatment of organic and inorganic contaminants in soil, sludge and sediment. It includes established, pilot scale and demonstration technologies in eight categories (Alternate Heat, Biological, Chemical, Incineration, Extraction, Fixation/Stabilization, Other and Pre-/Post- Treatment). The data were submitted by the developers/vendors of the technologies.

5. Materials Properties Data - Ageing of concrete structures in a nuclear environment

Atomic Energy of Canada and Ontario Hydro are collaborating with the International Atomic Energy Agency (IAEA) in Vienna on the development of a database for nuclear concrete structures, in particular on the processes associated with ageing. The ageing of nuclear structures is of special interest because of its impact upon the safety and reliability of operation of nuclear facilities, including the nuclear power plant concrete containment designed to separate the reactor

and other systems from the outside environment. The proposed database represents the first time this aspect has been addressed in particular.

Data, which are being gathered via a world wide IAEA survey, will be screened and processed by an international panel of eight experts. When completed, the database will be accessible to all nuclear utilities around the world. Its use will help the industry to control and manage ageing, thereby reducing its effect upon nuclear structures, and also to design future stations with greater insight.

Physics

1. Astrophysics

The Canadian Astronomy Data Centre (CADC) was established in 1986 as one of three worldwide distribution centres for data from the Hubble Space Telescope. Since then, the CADC mandate has expanded to include the provision of online astronomical archives and data distribution facilities for data from both ground- and space-based sources to the Canadian astronomical research community. Through collaboration with such centres as the Space Telescope Science Institute, the Space Telescope - European Coordinating Facility (ST-ECF) and the European Southern Laboratories (ESO), it has installed a variety of accessing and archiving software packages, the most prominent ones being STARCAT and PREVIEW. Heavy use is made of the high-speed CA*net for access and data transfer. In addition CADC provides access to the Star Guide Catalogue, Calibration Database and SIMBAD and maintains an archive of data obtained from the Canada-France-Hawaii Telescope.

2. Thermodynamics - Facility for the Analysis of Chemical Thermodynamics (F*A*C*T)

F*A*C*T is a Canadian thermochemical database system which contains thermodynamic properties on over 4000 inorganic stoichiometric compounds (5000 phases) including aqueous and gaseous ions. The public system is accessed via X.25 networks with host computers at McGill University, École Polytechnique de Montréal and CISTI (NRC, Ottawa).

Canadian National Committee for CODATA

The Committee, which met annually during this biennium, experienced some changes in sponsorship, structure and membership. The Canada Institute for Scientific and Technical Information assumed responsibility for the Committee and established a secretariat to administer and fund its activities. Professor Hugh King succeeded Dr. John Rodgers as Chairman and two new members, Drs. Paul Mezey and Roger Tomlinson, replaced Drs. David Brown, Alan Beck and Andrew Zolnai whose terms had expired.

Distribution of the CODATA Newsletter to over 400 addresses in Canada continued with inserts of particular interest to the Canadian community being added to several of the issues. Practical support was given to the June 1994 W. B. Pearson International Symposium on the *Impact of Structures on Materials Science* and organizational input was given to a symposium on geophysical data to take place at the 1995 IUGG meeting in Boulder, Colorado. Liaison was initiated with the Data Information Systems Panel of the Canadian Global Change Program.