

Report on Data Activities in Canada 1998

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

The following report on data activities in Canada was presented to the 21st General Assembly of CODATA at New Delhi, India in November 1998. To obtain further details on individual items or to submit information on other Canadian data activities for inclusion in the next report (September 2000 please contact:

Le rapport ci-joint, qui fait état des activités du Canada en matière de données, a été présenté à la 20^e assemblée générale de CODATA, à Tsukuba, Japon, en septembre 1996. 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 1998), veuillez communiquer avec:

Secretariat, CNC/CODATA
CISTI, Building M-55, Rm 249
National Research Council
Montreal Road
Ottawa, Ontario K1A 0S2

Secrétariat CNC/CODATA
ICIST, Édifice M-55, bureau 249
Conseil national de recherches
Chemin Montréal
Ottawa (Ontario) K1A 0S2

Telephone: (613) 993-3291
Fax: (613) 952-8246
Internet: codata@nrc.ca

Téléphone : (613) 993-3291
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.)

21st General Assembly of CODATA, New Delhi, November 1998

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 may be obtained either from the URLs or email addresses appearing in conjunction with most items or from the rapporteurs listed in Section XI.

I. **Biological Sciences** (M. Korab-Laskowska)

A. *Data Banks with Public Access Via the Internet:*

1. **Organelle Genome Database (GOBASE)**

GOBASE is a taxonomically broad organelle genome database that organizes and integrates diverse data related to organelles. The current version focuses on the mitochondrial subset of data.

<http://megasun.bch.umontreal.ca/gobase/gobase.html>

2. **Protist Image Database (PID)**

PID is part of the Molecular Evolution and Organelle Genomics program at the University of Montreal. PID provides images and online information on the morphology, taxonomy and phylogenetic relationships of protists. The PID Web page contains links to wide range of resources in protistology and related fields such as: microbiology, mycology, phycology and protozoology.

<http://megasun/protists/protists.html>

3. **Elegans Genetic Toolkit**

The Genetic Toolkit Project is funded by a grant from the NIH National Center for Research Resources (NCR) to the laboratories of Ann Rose, David Baillie and Don Riddle (University of British Columbia, Simon Fraser University and the University of Missouri respectively). The goal of the project is to provide genetic 'tools' to facilitate the cloning of genes and analysis of their function. The first stage has been the generation and characterization of chromosomal rearrangements (balancers) which are being used to isolate and maintain mutant strains. Current updates about balancers are available from the Web site. The project is now entering stage two, which is to provide overlapping deficiencies that will be aligned to both the genetic and the physical maps.

<http://genekit.medgen.ubc.ca/gb.html>

4. **Canadian Collection of Fungal Cultures**

The Canadian Collection of Fungal Cultures(CCFC) currently holds 10,500 strains of fungal

cultures representing about 2,500 species. The collection originated as an amalgamation of individual research collections and now serves as the primary repository for fungal cultures in the Agriculture and Agri-Food Canada research branch and accepts patent strains. It functions as a gene bank for this microbial resource and provides pure cultures to scientists in agriculture, forestry, medicine, private industry and biotechnology. Many species held in the collection are unique, and a number are new to science.

<http://res.agr.ca/brd/ccc/>

5. Directory of Canadian Culture Collections

Information was collected on the numbers of collections, diversity, availability, funding and methods of preservation used. Three types of collections emerged. A few collections were large in terms of taxa and isolates held. Others contained few species but represented important national or international collections of characterized strains. Most of these collections received institutional support for facilities and operations. Those remaining could be characterized as working collections of individual researchers. These were maintained with program budgets or from academic research grants.

<http://res.agr.ca/brd/ccc/ccfdir/ccfdire.html>

6. VectorDB

This database contains annotations and sequence information for many vectors commonly used in molecular biology. Information for more than 2600 vectors is available with search facilities. Vectors which are also in GenBank have direct links to that database via NCBI's Entrez browser.

<http://vectordb.atcg.com/>

7. Cystic Fibrosis Mutation Database

The information contained in this database is compiled with information collected by the Cystic Fibrosis Genetic Analysis Consortium.

<http://www.genet.sickkids.on.ca/cftr/>

8. The PAHdb

PAHdb contains data on PAH gene and alleles. It was developed by a Curatorial team for the PAH Mutation Analysis Consortium.

<http://www.debelle.mcgill.ca/pahdb/>

9. The Androgen receptor mutations database WWW Server

This resource contains a database of Androgen Receptor gene mutation, mutation maps and links to the references and the related EMBL site.

<http://www.mcgill.ca/androgendb>

B. Organizations or Systems Providing Access to the International Data Banks

10. Molecular Biology

Base4 was selected by Canada's National Research Council (NRC) to provide its computer database service for biological researchers. The Molecular Biology Database Service (MBDS), established in 1988 by NRC's Canada Institute for Scientific and Technical Information (CISTI), has been upgraded extensively by Base4 and made available as part of their Genome Mine service. In addition, Base4 offers a number of premium service packages.

<http://www.base4.com/>

11. CIAR Program in Evolutionary Biology (CIAR-PEB)

The Canadian Institute for Advanced Research (CIAR) supports a network of researchers across Canada as well as in other countries. The goal of the Program in Evolutionary Biology (CIAR-PEB) is to use the comparative database of genome sequences, to which this project will contribute, for developing concepts of genome, cell and population evolution, and for constructing algorithms for molecular structure/function analysis which may be later applied to problems in biotechnology, microbial diversity and genetic/genome technology. The CIAR-PEB Home Page contains information about its programs and activities as well as provides links to world wide Molecular Evolution and Computational Biology resources.

<http://megasun.bch.umontreal.ca/ciar/>

C. Main Sequencing Projects, Which Make Their Data Available to the Public

12. Sulfolobus Solfataricus Genome Data

The Sulfolobus solfataricus genome-sequencing project is a collaboration among seven laboratories worldwide: three Canadian (W. Ford Doolittle, Dalhousie University; Robert Charlebois, University of Ottawa; Mark Ragan, NRC-IMB) and four European (Roger Garrett, University of Copenhagen; John van der Oost, Wageningen Agricultural University; Michel Duguet, Universite Paris-Sud; Ib Groot Clausen, Novo Nordisk, Copenhagen). The project was initiated in mid-1993 with primary support from the Canadian Genome Analysis and Technology (CGAT) program along with contributions by the Canadian Institute for Advanced Research, the National Research Council of Canada, and the Medical Research Council of Canada. With the demise of CGAT, the project is continuing on European Union BIOTECH funding.

The interaction among laboratories depends almost fully on Internet connectivity. Data are moved from automated sequencing equipment in four laboratories (Copenhagen, Wageningen, Orsay and Halifax) directly into the IMB UNIX environment for processing. Primer calculation, database searching and annotation are automated using software designed and programmed at IMB under the supervision of Dr Christoph W. Sensen (primers), or in collaboration between Dr Sensen and Terry Gaasterland (Argonne National Laboratory and University of Chicago; from 1998 at The Rockefeller University). Computational facilities are those of NRC's Canadian

Bioinformatics Resource (CBR-RBC). Processed data are distributed among the laboratories through secure network facilities. As of mid-1998 the project is more than 80% complete. (mark.ragan@nrc.ca)

13. Organelle Genome Megasequencing Program (OGMP)

The OGMP is an interdisciplinary collaboration of seven Canadian research groups from Eastern Canada, each of which is interested in molecular evolution, mainly focusing on mitochondria, plastids and bacteria. This collaborative project, supported by the Canadian Genome Analysis and Technology Program (CGAT), concentrates on organelle phylogeny and includes the establishment of a centralized sequencing facility (the Megasequencing Unit) that serves as the major research hub. The "Megasequencing Unit" is located at the University of Montreal. The OGMP bioinformatics division is responsible for the data handling and analysis. The sequences of mitochondrial genomes from the "Megasequencing Unit" will be made available to the scientific community through GenBank and GOBASE.

<http://megasun.bch.umontreal.ca/ogmproj.html>

14. Fungal Mitochondrial Genome Project (FMGP)

FMGP, a project of B. F. Lang's research group (Department of Biochemistry, University of Montreal), is supported by the Medical Research Council of Canada (MRC). The goal of the FMGP is to sequence complete mitochondrial genomes from all major fungal lineages, to resolve the fungal branch of the 'tree of life' and to investigate mitochondrial gene expression, introns and mobile elements. The webpages of the FMGP include extensive information on subjects such as general organismal information, gene map, complete sequence, phylogeny, etc.

<http://megasun.bch.umontreal.ca/People/lang/FMGP/FMGP.html>

II. Chemistry (P. Mezey)

A. LOGKOW - Databank on Octanol-Water Partition Coefficients.

Dr. James Sangster, of Sangster Research Laboratories, Montréal, Québec, has maintained and upgraded a databank on octanol-water partition coefficients of a large set of molecules, important in a variety of chemical and biochemical fields, including human health. In the study of biochemical activities of potential drug molecules as well as environmental toxicants, these data are essential in making comparisons and potential predictions.

(james.sangster@mail.polymtl.ca)

B. Canadian Domestic Substances & Non-Domestic Substances Databank.

TerraBase Inc. has released its Canadian Domestic Substances & Non-Domestic Substances List (DSL&NDSL) on CD-ROM. The DSL&NDSL covers over 66,400 substances scheduled under the

Canadian Environmental Protection Act (CEPA). The Toxicity Data & QSAR Database, an integral part of the TerraTox / TerraFit Software Suite, contains records on more than 8000 compounds, many of which are known or suspected carcinogens, mutagens, pesticides, endocrine disruptors. It also contains at present more than 5000 references of recent scientific publications on quantitative structure-activity relationships (QSARs) and closely related subjects (including data sources).

(klaus.kaiser@cciw.ca)

C. Molecular Toxicology Databank, Environment Canada.

More info: Dr. Mark Lewis, Commercial Chemicals Evaluation Branch, Environment Canada, Place Vincent Massey, 14th floor, 351 St. Joseph Blvd., Hull, Quebec K1A 0H3, Canada, tel: 819-953-7199; fax: 819-953-4936; e-mail: mark.lewis@ec.gc.ca

D. PAH (polyaromatic hydrocarbon) Aquatic Toxicity Databank.

A databank on the photochemical activities and aquatic toxicity of polyaromatic hydrocarbons, as well as their photooxidized products is maintained and further developed by Prof. Bruce Greenberg and Prof. G. Dixon, University of Waterloo, Ontario. Data on the chemical properties and toxicities recorded in this database are expected to serve both academia and the chemical industry, providing tools for toxicological risk assessment and environmental action.

(Prof. Bruce Greenberg, Prof. G. Dixon, Department of Biology, University of Waterloo, Ontario.)

E. Pesticide, Herbicide, Metal Contaminants, Synergistic Toxicity in Soil Database.

A database of pesticide and herbicide activities in the presence of metal contaminants, affecting the soil - plant root system interface is maintained and further developed by Prof. Huang, University of Saskatchewan. This database is explored in similarity studies, in order to enhance the predictability of adverse effects of new pesticides and herbicides entering the market and for suggestions of potential modifications.

(Prof. P. Ming Huang, Dept. of Soil Science, University of Saskatchewan, Saskatoon, SK Canada)

F. Hemoglobin Binding Affinity Constants Database.

A database on the hemoglobin binding affinity constants of a large series of organic molecules has been maintained and further developed by Prof. Krishnan, Université de Montréal. This database is already being applied for the study of some of the adverse effects of toxic substances.

(Prof. Kannan Krishnan, Dép. Médecine du Travail et d'Hygiène du Milieu, Faculté de Médecine, Université de Montréal, Québec.)

G. Cadmium and Zinc Uptake by Grain Varieties Databank.

A database on the toxicity of various metals, including Cadmium and Zinc, with special emphasis on their uptake by grain varieties, is being maintained and upgraded by the research groups of Prof. Beverly Hale, University of Guelph, Ontario, and Prof. Francine Denizeau, Dép. Chimie, Université du Québec à Montréal, Québec.

H. Functional Group Electron Density Databank for Carcinogenic Carbonyl Compounds.

A functional group electron density database of carcinogenic carbonyl compounds involved in vehicle exhausts is being developed by Dr. Serge Lamy, Health Canada and Dr. Mezey, University of Saskatchewan.

(mezey@sask.usask.ca)

I. Halogenated Organic Molecules Electron Density Databank

A molecular shape database for a series of halogenated organic molecules is maintained and upgraded by Prof. Mezey, University of Saskatchewan. The earlier polycyclic aromatic hydrocarbon (PAH) shape database is continuously updated. These shape databases have new applications in the pharmaceutical industry, in new lead search, in toxicological risk assessment within the framework of CNTC (Canadian Network of Toxicology Centres) Quantitative Risk Assessment project, and in pesticide research.

(mezey@sask.usask.ca)

III. Crystallography (G. Wood)

A. NRC Metals Crystallographic Database (CRYSTMET)

Under an exclusive license from NRC, Toth Information Systems not only maintained and updated the database but has developed two new CD-ROM-based products called CRYSTMET and CRYSTMET-P. The CD version of CRYSTMET features a new search system and a suite of analysis tools; CRYSTMET-P is similar to basic CRYSTMET but contains, as well, calculated powder patterns. Both products will be released October 1, 1998.

(<http://www.nrc.ca/programs/toth/>)

B. Cambridge Structural Database (CSD)

The CSD is distributed in Canada by Dr. George Ferguson at the University of Guelph. There are now nineteen Canadian university sites with their own CSD system. The CSD CD-ROMs are distributed to the sites in mid-April and mid-October each year. Access to the CSD is then available to the group covered by the relevant site-license at each university.

(george@angus.chembio.uoguelph.ca)

IV. Geoscience (R. Berman)

A. GSC Canadian Geoscience Publications database

The Canadian Geoscience Publications Directory (<http://ntserv.gis.nrcan.gc.ca>) is being developed by the Geological Survey of Canada (GSC) to provide "one-window" internet access to all Canadian Geoscience publications. The goal is to allow the user to perform both text based and

interactive spatial queries. Currently the directory consists of GSC map publications and geoscience publications from Newfoundland. Other GSC publication types as well as data from other provinces will soon be available.

B. Natural Resources Canada bibliographic database

GEOSCAN is NRCAN's bibliographic database of GSC publications. With over 40,000 records, it contains bibliographic, geographic and subject control for all publications of the Geological Survey of Canada and many of its contributions to outside publications. Clients may access the GEOSCAN database through the Earth Sciences Information Centre home page at

http://www.nrcan.gc.ca/ess/esic/esic_e.html (English) or

http://www.nrcan.gc.ca/ess/esic/esic_f.html (French).

C. The National Mapping Program (NATMAP)

This is a major geoscience initiative, conceived in 1988 by the GSC, aimed at increasing the level of geoscientific mapping in Canada through multi-institutional and multi-disciplinary projects. The databases associated with each project consist of bedrock geology maps, surficial geology maps, topographic base maps, as well as geophysical, remotely sensed, mineral deposit, and geochronologic data. Some NATMAP data sets can be viewed for the Slave Province

(<http://gis.nrcan.gc.ca/natmap/slave/slave.html>) and Shield Margin

(<http://gis.nrcan.gc.ca/natmap/shield.html>).

D. Geothermodynamic Database

An internally consistent database of mineral end-members and solution properties is maintained by the Continental Geoscience Division (CGD) of the GSC as part of the TWQ thermobarometry software program (<http://www.gis.nrcan.gc.ca/twq.html>). The most recent version of this database incorporates new experimental measurements on Ti and Al solubility in biotite and Al in orthopyroxene.

E. Computer Aided Field Mapping Software

The GSC's CGD has continued development of the PC-based Fieldlog software package designed to assist in computer-aided field mapping and rapid construction of geological databases. Fieldlog v3.0 (<http://gis.nrcan.gc.ca/fieldlog/Fieldlog.html>) incorporates a number of very useful features including plotting of geological symbols, stereonet, rose diagrams; sophisticated database queries; customizable coordinate systems; projections and transformations between coordinate systems; integration with GPS and mobile recording software.

F. North Baffin/Melville Peninsula database

A digital 1:500,000 scale seamless integrated geoscience compilation map of the North Baffin/Melville Peninsula region is being created at the GSC. This digital geoscience knowledge base (<http://gis.nrcan.gc.ca/baffin>) will include all previous regional scale bedrock and surficial

geological mapping, geophysical data, mineral occurrences, rock geochemistry, and a geochronology database. A user-friendly map browsing tool enables spatial queries as an aid to visualizing and extracting specific geoscience information.

G. The Airborne Geophysics Section of the GSC's Mineral Resources Division (MRD)

MRD maintains the National Airborne Gamma-Ray Spectrometry Database, consisting of about 1.5 million line-km of multivariable airborne gamma ray spectrometry data (K, U, Th, Exposure, Mag & VLF), mostly in the Canadian Shield areas. Survey line spacings are 25 km, 5km, 1 km, and <250-500 m. Data can be accessed via: natgam@gsc.nrcan.gc.ca or

<http://www.geophys.gsc.nrcan.gc.ca/airgs>

H. GSC's MRD Geochemical Reconnaissance Database

The National Geochemical Reconnaissance Database (<http://gds.agg.gsc.nrcan.gc.ca>; email: mmccurdy@nrcan.gc.ca) contains field and analytical data of stream and lake sediment and water samples. Data have been gathered from surveys conducted since 1974 for selected areas in Canada. More than 200 surveys have been carried out, representing almost 200,000 samples covering about 2,100,000 square kilometres. The surveys were carried out using consistent sampling and analytical procedures.

I. GSC Atlantic Geoscience Division databases

The GSC Atlantic Division is the principal repository for marine geophysical data, sediment grab and core samples, and rock and paleontological collections resulting from government/industry collaboration in Canada's offshore areas. The databases (<http://agcwww.bio.ns.ca/pubprod/pubprod.html>) consist of 1.3 million line kilometres of seismic reflection and refraction data, 8000 km of deep seismic reflection data, 1.7 million offshore gravity observations; 3.5 million offshore shipborne magnetic observations, 10000 core stations from 900+ cruises, 250,000 seafloor sediment and rock samples, 800 onshore maritime borehole records, a biostratigraphic, geographic, and taxonomic database containing 825,000 records, and BASIN, a database of geological and engineering information on 300+ offshore petroleum and exploration wells.

J. The Atlantic Coastal Zone Database Directory

Found at (<http://www.ndi.nf.ca/ndi/aczisc/index.html>), the Directory lists and describes 608 databases of relevance to the integrated management and sustainable development of the coastal zone of Atlantic Canada. The database descriptions include such details as contact persons, availability, formats, geo-referencing and scales, etc. The databases are maintained by governments (federal, provincial, municipal/regional), academic/research institutions, the private sector, non-governmental organizations and non-profit organizations.

K. British Columbia Ministry of Energy & Mines mineral occurrence database:

MINFILE (<http://www.ei.gov.bc.ca/geosmin/minfile/minfile.htm>) contains geological, location and economic information on over 11,900 metallic, industrial mineral and coal mines, deposits and occurrences in British Columbia. The MINFILE database can be searched, reported on, and updated using the MINFILE/pc software program. The program, data and user documentation can be downloaded from this site.

V. Geophysical (R. Coles)

A. Geomagnetic Data

The National Geomagnetism Program of the Geological Survey of Canada (GSC) maintains the archive of Canadian magnetic observatory data, describing the variations with time in the Earth's magnetic field at points across Canada. This archive of about 6 Gigabytes contains high-resolution digital data from 13 observatories for the past 22 years plus historical data back to the time of the International Geophysical Year and earlier. The most recent two years of data are maintained on-line. The database is accessed by researchers and others from all parts of the world. An automatic DRM (data request manager) using electronic mail is in operation, and custom requests can be handled using Internet ftp. Descriptive material on data acquisition, data availability, and conditions of access can be found at the website: www.geolab.nrcan.gc.ca. Data older than eight days can be viewed in graphical form on the Web without restriction. Viewing of newer data requires special authorization, to which special conditions apply.

B. Seismological Data

The National Earthquake Hazards Program of the Geological Survey of Canada maintains the archive of Canadian seismological data from the Canadian seismograph network. The archive contains a large number of older analogue seismograph records dating back to the early 1900s. The modern data archive contains over a Terabyte of digital time series data from 1980 onward. In addition, the archive contains digital data beginning in about 1966 from the Yellowknife seismic array, used in nuclear explosion detection studies. The archive also holds first-level derived data in the form of earthquake epicentre locations for Canada. An automatic DRM (data request manager) using electronic mail is in operation, and is heavily used. Direct links exist with the International Data Centre for Seismology in Washington DC. The WWW site (www.seismo.nrcan.gc.ca) provides derived data such as epicentres, current information on recent earthquakes, a catalogue of data availability, and plots of waveform data for selected events.

C. Aeromagnetic Data

The Regional Geophysics group of the Geological Survey of Canada maintains the National Database for Aeromagnetic Data for Canada. Data date back to 1947, with early analogue maps converted to digital form. The database contains holdings for about 80% of Canada at a regional scale, amounting to about 7 Gigabytes. Data are available in many forms: as point values, gridded

sets, plots at any scale, data in any format, on any media type, electronic file transfer. An on-line ordering system is in operation, and full details of the data and services can be found at the website: (gdcinfo.agg.nrcan.gc.ca).

D. Gravity Data

The Atlantic Division of the Geological Survey of Canada is the principal repository for marine acoustic, magnetic, seismic and gravity survey data resulting from government/industry collaboration in Canada's offshore areas. Holdings include 1.3 million line kilometres of seismic reflection and refraction profiles, 8000 km of deep seismic reflection data, 1.7 million offshore gravity observations and 3.5 million offshore magnetic observations. Catalogues and data accessibility information are presented on the GSC Atlantic website under GSCA databases at: agcwww.bio.ns.ca/data_collections.html.

E. Radiometric data

The Mineral Resources Division of the Geological Survey of Canada operates the National Database for Radiometric Survey Data. Coverage includes about 2 million square kilometres at 5 km line spacing in the Canadian Shield, and many more areas in other parts of Canada at closer line spacing. Data descriptions and index maps are available on a website at: www.geophys.gsc.nrcan.gc.ca/hp_e.htm

VI. Geospatial (R. Tomlinson)

A. Geospatial Data Access

In cooperation with other agencies, Natural Resources Canada and the Department of National Defence are developing key technical components of the Canadian Geospatial Data Infrastructure (CGDI) that will enable Canadians to access vast quantities of geospatial data on the information highway. A CGDI project, CEONet, is developing a comprehensive clearinghouse for suppliers and users of geospatial data and services. To date more than 4900 international data holdings are registered in CEONet of which approximately 250 are Canadian. The Canadian data holdings include many of the major national data holdings, including the National Topographic Data Base, national remote sensing archives, National Geochemical Reconnaissance Data Base and many others. CEONet provides capabilities for browsing and searching these data holdings and provides a mechanism for directly searching the detailed inventories, e.g., unique map sheets.

While most of these data are still delivered in an offline fashion, many are now available to be downloaded directly over the network. An important example is the Canada Land Inventory data that consists of six different themes for the southern third of Canada.

The CEONet information can be found at <http://cgdi.gc.ca/ceonet>. Other direct Email addresses for geospatial data holdings in Canada are as follows:

Aeronautical charts and publications:	Email: aero@nrcan.gc.ca
Cadastral Survey products:	Email: lst@NRCan.gc.ca
Geodetic Survey products:	Email: information@geod.NRCAn.gc.ca
Remote Sensing imagery products:	Email: orderdesk@ccrs.nrcan.gc.ca
National Atlas products:	Email: atlasinfo@ccrs.NRCAn.gc.ca
Digital Topographic products:	Email: bndt@CCG.NRCAn.gc.ca
Topographic Map products:	Email: topo.maps@NRCan.gc.ca
Geographical Names products:	Email: geonames@NRCan.gc.ca
Aerial Photography products:	Email: NAPL@GeoCan.NRCAn.gc.ca

B. Geospatial Framework Data

Framework data provides a foundation of reliable geospatial data to enable data integration for value-added data, application development or detailed data collection. As one of the themes of the CGDI, the framework data will provide precise reference locations as the basis for other CGDI data. This will allow geospatial data users to georeference data from various sources and at various scales. The foundation framework will be primarily composed of a Data Alignment Layer, which will be freely accessible by mid-1998. Natural Resources Canada has been mandated to produce and implement the CGDI Data Alignment Layer on the Internet in accordance with CGDI recommendations.

The CGDI Data Alignment layer (CDAL) contains easy-to-identify points at various map scales and allows for georeference of digital data sets (vector and raster) from different sources. The geographic coordinates used to define the positions of CDAL points are degrees and decimal degrees of latitude and longitude. The unit of measurement used to store the positions of CDAL points in Cartesian coordinates (X,Y) is the metre. The unit of measurement used to store angular values is the degree (integer). The resolution of CDAL points is one metre. Each CDAL point has at least one set of characteristics that defines its entry in the CDAL. All CDAL features are points of one of three types: line intersection, centre of mass or point feature. Data Alignment layer information can be found at <http://www.ctis.nrcan.gc.ca/~cdal>.

C. Geospatial Data Standards

The area of geospatial data standards has seen a significant level of activity in Canada over the past two years. Canada has actively participated in the International Standards Organization (ISO) TC211 standards on Geographic Information/Geomatics and the other international

geospatial standards developments in ISO JTC1 SC32 WG4 SQL/MM – Spatial, the ISO TC204 Road Transportation, the International Hydrographic Organization (IHO) and the NATO based Digital Geographic Information Working Group (DGIWG). In addition, several Canadian companies have participated in the Open GIS Consortium.

The IHO, the DGIWG and the TC204 Road Transportation committees have developed product-based standards that describe the data products used in particular application areas. The IHO S-57 standard is narrowly focused on safe navigation at sea and is unique in that it incorporates dynamic updating to maintain the currency of charts. The DGIWG is more broadly focused, covering a range of military data. These standards have been implemented in Canada and data are available. There has been a significant effort by Canada to harmonize these standards. The Road Transportation standard is nearing completion and there is a commercial commitment to implement it. The OGC represents an industrial consortium including the major GIS vendors. It has issued several specifications, through a proposals process, that address the interfaces to GIS equipment. The ISO JTC1 SC32 WG4 SQL/MM – part 3 Spatial work addresses the manner in which spatial data are stored and accessed in larger databases.

ISO TC211 is developing a suite of common component standards that can be used as building blocks to construct the other standards. TC211 ISO 15046 standards will be implemented as profiles and standards such as the DIGEST standard can be described as an assemblage of these more basic building blocks. Harmonization efforts have already produced significant alignments, for example, the OGC, SQL/MM – Spatial, DGIWG and IHO have agreed to common data model sets for computational geometry under the umbrella of TC211.

The level of Canadian participation in these international committees is high. The activity comes under the auspices of the Canadian General Standards Board – Committee on Geomatics (CGSB-COG) through the Standards Council of Canada (SCC). The CGSB-COG structure mirrors that of ISO TC211.

The COG Technical Committee – D. V. Hume of Indian and Northern Affairs;
WG1 – Framework and Reference Model – Dr. K. Fadaie, Canadian Centre for Remote Sensing;
WG2 – Geospatial Models and Operations – (vacant);
WG3 – Geospatial Data Administration – P. Charlesworth, Natural Resources Canada;
WG4 – Geospatial Services – P. Morin, Department of National Defence;
WG5 – Functional Standards – C. D. O'Brien, IDON Technologies Inc.;
CAC – Canadian Advisory Committee – D. McKellar, Department of National Defence.

It is planned by the Federal Government Inter Agency Committee on Geomatics, that the TC211 components and registered data product profiles will form the standards for the Canadian Geospatial Data Infrastructure.

VII. Environment (B. Malone)

A. Databases for Environmental Analysis (Cat. No. 16-506XCB)

This unique CD-ROM-based reference guide to environmental data available throughout Canada documents some 1200 databases held by 127 departments of the federal, provincial and territorial governments. Covering subjects from digital base maps through acid rain to biodiversity, it is indispensable for those looking for information on linkages between human activities and the environment. Each of the databases is described in terms of twenty variables including a summary description, responsible organization, contact persons, list of parameters, geographic coverage and update frequency.

(trephele@statcan.ca)

B. Specimen Banking

Specimen banks, a collection of archived biological tissues or environmental materials and associated data, are a special form of database.

Environmental monitoring is one of the fundamental tools for effective management of anthropogenic stresses on the biosphere at local, national and international levels. In order to be effective, the monitoring must be flexible enough to cope with changes in the nature of the problems addressed, the technologies used and the state of knowledge. It must be able to describe the state of the environment, evaluate environmental threats, analyse anthropogenic effects on the environment and provide the basis for establishing and monitoring our responses to those stresses. Specimen Banking is one of the most cost-effective ways of achieving those results.

The main function of a Specimen Bank is to collect and maintain material in such a way as to allow for:

- ?? real-time monitoring of environmental parameters of current interest
- ?? retrospective monitoring to investigate newly discovered parameters of interest, to validate new or modified analytical methods and to verify earlier results
- ?? ecotoxicological research

VIII. Materials Properties Data (G. Wood)

A. Ageing of concrete structures in a nuclear environment

Before proceeding with the initiation of an International Concrete Ageing Database the matter was passed by IAEA in 1995 for an assessment to OECD/NEA in Paris as reported in the 1996 *Data Activities* report.

Several meetings were held by NEA and the database was considered to be a priority. At NEA's May 1997 meeting Canada (C. Seni) presented a report on how the database could be initiated. However, concern was expressed by Country Member States that the required cooperation from the Nuclear Utilities in various countries may not be obtained and that IAEA would be in a better position to obtain this cooperation. During this period and with participation from C. Seni, NEA has also organized two international workshops, concrete ageing related, which will become later a source for populating the concrete ageing database.

Over the same period and in parallel with IAEA's activity on concrete ageing issues, C. Seni and D. Naus (USA) organized within RILEM (International Reunion of Laboratories for Testing of Structures and Materials), a Technical Committee to address ageing issues of concrete nuclear structures including the concrete ageing database. Its activity was coordinated with NEA's activity, by C. Seni. RILEM's newly created Technical Committee TC-160 MLN (Methodology for Life prediction of concrete structures in Nuclear power plants) defined a structure of the Concrete Ageing Database which will be taken over by IAEA when launching the program. The participants to this task were C. Seni, B. Oland (USA) and M. Johnston (UK).

According to IAEA's planning, the initiation of the Concrete Containment Database is scheduled for the year 2000 when a specialists' meeting will be called to decide how to proceed.
(cseini@globalserve.net)

B. Computer Integrated Material Database (CIMDATA)

The Industrial Materials Institute of the National Research Council of Canada has developed an integrated system of knowledge and factual databases covering plastics, non ferrous metal alloys, cast iron alloys and tooling materials. More than 100 physical and engineering properties and over six hundred moulding and casting materials are encompassed.

The knowledge-based components of the system dealing with process control, for example, provide answers or suggestions to problems that might arise on the production floor concerning a given process. Thus, for instance, an engineer with an injection mould that is not working correctly may query the system and be led to a solution based on the expert knowledge captured therein.

Complementing these components are the factual databases which assist a user during the simulation of casting and moulding forming processes and in the selection and comparison of

constant or variable material properties. Here, the user may choose from a wide number of properties and parameters for the materials of interest and be guided in making the optimum choice.

(georges.salloum@nrc.ca)

IX. Physics - Astrophysics (D. Durand)

The Canadian Astronomy Data Centre (CADC) continued to be the focus of data activities in astronomy. The CADC developed an innovative approach to provision of archival data from the Hubble Space Telescope which involves storing the raw data on CD-ROMs in a 500 platter jukebox and performing automatic calibration of the data when it is requested from the archive. This takes advantage of the latest calibration software and reference files. The archive can be searched via a World Wide Web interface which provides ubiquitous access to all computing platforms.

The CADC is moving the Canada-France-Hawaii Telescope (CFHT) Archive to CD-ROM as well to provide online access to the data. The CADC is also archiving data from Canada's two other major facilities, the James Clerk Maxwell Telescope in Hawaii and the Dominion Radio Astrophysical Observatory in Penticton, B.C.

Astronomers have also gained access to the two versions of the Digital Sky Survey through the CADC. From the more than 300 CDs currently available, WEB users may retrieve any part of the sky on their own computer via the Internet.

The CADC is now involved in designing and implementing TERAPIX, a data processing pipeline dedicated to the new camera to be attached to the CFHT in 2001. This camera, delivering a one-degree field of view, will produce images of around 800 MB each. Because of the large volume of data, 100 images per night for typically 100 nights per year, TERAPIX has to be able to process these images automatically.

(<http://cadcwww.hia.nrc.ca/>)

X Thermodynamics (J. Sangster)

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

F*A*C*T is a fully integrated Canadian thermochemical database system which couples proven software with self-consistent critically assessed thermodynamic data. It currently contains data on over 5000 chemical substances as well as solution databases representing over 70 non-ideal solutions (liquid alloys, slags, mattes and molten salts). F*A*C*T is accessible online from McGill University and is available also as a PC version.

(<http://www.crct.polymtl.ca>)

B. University Research Programs

Profs. C. B. Alcock and V. Itkin (University of Toronto) assess thermodynamic data of the elements (Debye temperature, $C_p(T)$, enthalpy, third law entropy and fusion properties). C_p data are described by several equations and recommended data are given.

Prof. A. E. Mather (University of Alberta) measures vapour-liquid equilibria and enthalpies of reaction and solution for acid gases in aqueous solution of polar organic solvents (application in gas purification). He has contributed to the IUPAC Solubility Data Series in compilation and assessment of data for CO_2 in water and non-aqueous systems, as well as for solids and liquids in supercritical CO_2 .

Prof. J. Lielmezs (University of British Columbia) compiles and evaluates data for the development of equations of state. They also serve for correlative predictive methods for heats of vaporization, surface tension and transport properties as well as ideal gas thermodynamic properties.

Prof. P. Englezos (University of British Columbia) measures gas hydrate phase equilibria involving methane, CO_2 , hydrocarbons and nitrogen. Measurements also include the solubility of calcium carbonate in the presence of adsorbed substances.

XI. Canadian National Committee for CODATA

The Committee continued to meet annually during this biennium under the sponsorship of The Canada Institute for Scientific and Technical Information (CISTI). Dr. Daniel Durand, replacing Dr. Denis Crabtree, and Dr. George Needler joined as new members; the term of Drs. Sangster was renewed for three years. Current membership, along with rapporteur responsibilities for this report, are shown in the following table:

Chairman	Rapporteur - Section	Email address
Prof. Paul Mezey	Chemistry	mezey@sask.usask.ca
Members		
Dr. Daniel Durand	Physics-Astrophysics	daniel.durand@hia.nrc.ca
Dr. Maria Korab-Laskowska	Biology	korablam@bch.umontreal.ca
Dr. George Needler	Oceanography	G_Needler@bionet.bio.dfo.ca
Dr. James Sangster	Thermodynamics	jsangster@mail.polymtl.ca
Dr. Roger F. Tomlinson	Geospatial	talgeo@magma.ca
Observers		
Dr. Robert Berman	Geoscience	berman@gsc.emr.ca

Dr. Andrée Bichon		andree.bichon@nrc.ca
Dr. Richard L. Coles	Geophysical	coles@geolab.emr.ca
Mr. Brian Malone	Environment	bmalone@synapse.net
Dr. John R. Rodgers	Crystallography	rodgers@snd.cisti.nrc.ca
Dr. Larry Speers		speersli@em.agr.ca
Dr. Barry M. Wood		barry.wood@nrc.ca
Secretariat		
Mrs. Roxanne Rochon (Secretary)		roxanne.rochon@nrc.ca
Dr. Gordon H. Wood (Exec. Secretary)	Materials	gordon.wood@nrc.ca

At its meeting in May 1997, the Committee initiated a pilot project to promote awareness in Canada of the need for data quality and data consistency. The objective of the pilot phase was to ascertain by interview the quality assurance methodology used by producers of a limited number of Canadian data sets. It is anticipated that at least another six months to a year will be required to complete the study.

The Committee continued its responsibility for distributing the CODATA Newsletter to over 400 addresses in Canada. CISTI, as the Secretariat for the Committee, has the distinction of hosting the main web site for CODATA which links to all the other CODATA activities world wide and includes electronic versions of the Newsletter, Handbook, various reports, etc.
(<http://www.cisti.nrc.ca/programs/codata/welcome.html>)

In addition, the NRC Herzberg Institute of Astrophysics in Victoria continued to host the web site for CNC/CODATA. Dr. Durand served as webmaster.
(<http://cadcwww.dao.nrc.ca/cnc-codata/>)