Report on Data Activities in Canada
1996

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

The following report on data activities in Canada was presented to the 20th General Assembly of CODATA at Tsukuba, Japan in September 1996. To obtain further details on individual items or to submit information on other Canadian data activities for inclusion in the next report (September 1998) please contact:

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

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

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

Le rapport ci-joint, qui fait état des activités du Canada en matière de données, a été présenté à la 20e 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:

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

Téléphone : (613) 993-3291
Télécopieur : (613) 952-8246
Internet: codata@nrc.ca

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

20th General Assembly of CODATA, Tsukuba, October 1996

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.

I. Biological Sciences

A. Data Banks with Public Access Via Internet:

1. Indices of Available Fungal Cultures

Produced by the Nova Scotia 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.
(http://aceis.agr.ca/icar/docs/90005436.html) ***NSIS***

2. Fungal Metabolites

Also produced by the Nova Scotia 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. ***NSIS***

3. Organelle Genome Database (GOBASE)

This is a unique, interdisciplinary project between two Canadian universities, Montreal and Dalhousie which differs in its concept from existing project-specific databases, e.g., flybase and ACeDB. This project is supported by the Canadian Genome Analysis and Technology Program (CGAT). The goal of the project is to create a comparative organelle genome database, integrating data from a wide range of sources, which will be a model for many of the other genome projects underway, for which comparative data are not yet available. GOBASE is being developed using SYBASE RDMS and WEB/GENERA software. The database will be network accessible and will permit submission of confidential data and password-protected access, at the same time allowing free access to most of its information by the general scientific community. The first Internet-accessible GOBASE release is scheduled for July 1996.
(http://megasun.bch.umontreal.ca/gobase)

4. 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)

5. Elegans Genetic Toolkit

The Genetic Toolkit Project is funded by a grant from the NIH National Center for Research Resources (NCRR) 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)

6. Cosmid Transgenics

Transgenic strains available from the Department of Medical Genetics (University of British Columbia) were constructed in association with the C. elegans Genome Sequencing labs in St. Louis, Missouri and at The Sanger Center, Hinxton, UK. Funding for this work was made available by a grant from the Canadian Genome Analysis and Technology Program (CGAT) to Ann Rose and David Baillie. As of March 27, 1996 the total number of cosmids available as transgenic strains is 129. In all cases, cosmids micro-injection was performed on N2 hermaphrodites. All cosmids were co-injected with the plasmid pCes1943 which contains a semi-dominant allele of the rol-6 gene.
(http://darwin.mbb.sfu.ca/imbb/dbaillie/cosmid.html)

7. Physical and Transcription Maps of Human Chromosome 22q13.3

Heather McDermid and Kenneth Roy (Biological Sciences, University of Alberta) propose to search for genes at chromosome band 22q13.3, which is predicted to be gene-rich and may contain over 300 genes, including those for several genetic diseases that map to this region. This project is funded by the Canadian Genome Analysis and Technology Program (CGAT). In collaboration with the U.S. Human Genome Center for Chromosome 22, a cosmid contig of 22q13.3 will be assembled using minimally overlapping yeast artificial chromosomes (YACs) as the starting material. The cloned DNA fragments identified will provide the source material for

isolating genes by exon amplification and cDNA selection. All genes cloned will be at least partially sequenced for expressed sequence tag (EST) production. The ESTs and their precise map location will be made available to the scientific community through the Genome Data Base.

8. Integrated Map of Human Chromosome 7

Lap-Chee Tsui and Stephen Scherer are with the Department of Genetics, The Hospital for Sick Children, Toronto, and the Department of Molecular and Medical Genetics, University of Toronto. Lap-Chee Tsui and Johanna Rommens are the co-investigators of a CGAT grant entitled "Large-scale physical gene mapping of human chromosome 7Q". The physical map of the long arm of chromosome 7 now contains the map positions for 1,200 YAC clones. A systematic search has resulted in the recovery of over 300 unique cDNA clones for 7q21- q22. The Toronto group works on the incorporation of genetic markers, such as those from Genethon and the Cooperative Human Linkage Center (CHLC), on the physical map. A major contribution has been in the area of contig assembly. Part of the YAC mapping work was accomplished in collaboration with Karl-Heinz Grzeschik's laboratory. Another collaboration was established with Helen Donis-Keller's laboratory to generate a fine map of the tip of the long arm.

9. 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/)

10. 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/ccfcdir/)

11. CANSIS-Canadian Soil Information System

CanSIS uses a GIS called ARC/INFO to maintain land resource data in the NSDB. These data describe the location of soil types in Canada, and include characteristics that are relevant to a soil's biological productivity, as well as landscape attributes such as slope, local surface form and the presence of rock outcrops.
(http://res.agr.ca/CANSIS/_overview.html)

12. Canadian Poisonous Plants Information System

This information system provides details taken from literature references for over 250 plants that can poison livestock, pets and humans in Canada. Its interactive search engine provides information on plant names, distribution, toxic plant parts, toxic chemicals, and symptoms of poisoning.
(http://res.agr.ca/brd/poisonpl/)

13. Beetles of Canada and Alaska

This database provides the current nomenclature of all the beetles known to occur in Canada and Alaska. Distributions are rendered by maps indicating the presence or absence of the taxa by geographic subregions.

The purpose of this database is to provide the correct names of the beetles occurring in Canada and Alaska and give a brief survey of the distribution of the taxa in the area covered. The ending of all species-group names has been checked to conform with the gender of their respective genera. The distributional records are based on published records and on specimens in the Canadian National Collection. For some families, records from other collections have been included. Doubtful records from literature that apparently are based on misidentifications or mislabeled specimens are rejected.
(http:/res.agr.ca/brd/beetles/english/html/bhome_e.html)

14. Diptera Types in the Canadian National Collection of Insects - Part 4, Tachinidae

This work is the fourth, and last, in a series of catalogues on the Diptera types in the Canadian National Collection of Insects (CNC). The first, published in 1991, dealt with the Nematocera. The second, published in 1993, documented the types of brachyceran Diptera exclusive of the Schizophora. The third, currently in production, will cover the Schizophora exclusive of the Tachinidae. A brief history of the Diptera collection in the CNC is given in Part 1 of this series.
(http://res.agr.ca/brd/tachinid/tacheng.html)

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

1. Molecular Biology

The CAN/SND Molecular Biology Database System (MBDS) is a sub-system of CAN/SND, the Canadian Scientific Numeric Database Service operated by the National Research Council Canada (NRCC) through the Canada Institute for Scientific and Technical Information (CISTI). Via the MBDS, CAN/SND provides ready access to the world's major protein and nucleic acid sequence databases, including GenBank, EMBL, NRL-3d, PIR, ProSite, Swiss-Prot, and others. Similarly, CAN/SND provides access to software packages available in the field which enable users to search, analyze, manipulate and display protein and nucleic acid sequences. Among these are GCG, FASTA, BLAST, ATLAS, Entrez, PHYLIP. CAN/SND's databases may be accessed in several different ways depending upon one's local resources and needs.
(http://www.cisti.nrc.ca/cisti/cansnd/lifesci.html) ***CAN/SND***

2. 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/ciar/)

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

1. Sulfolobus Solfataricus Genome Data

The Sulfolobus solfataricus genome sequencing project is a collaboration of W. Ford Doolittle, Robert Charlebois (U. of Ottawa), Mark Ragan (NRC-IMB) and Christoph Sensen (NRC-IMB). This is the only all-Canadian whole-organism genome-sequencing project and the first off the mark of several archaeal genome projects worldwide. It was initiated in mid-1993 with primary support from the Canadian Genome Analysis and Technology (CGAT) Program and contributions by the Canadian Institute for Advanced Research (CIAR), the National Research Council of Canada (NRC), and the Medical Research Council of Canada (MRC). The interaction among the three laboratories relies heavily on the Internet. Data are moved directly into a UNIX environment at IMB for initial processing and database analysis. Extensive development of the Sulfolobus computational environment at IMB by Christoph W. Sensen and Terry E. Dalton now allows to

work at a single computer with seamless access to UNIX XWindows, MS Windows, WWW tools, and remote terminal applications. Processed data are distributed among the laboratories through secure network facilities and analyses are stored in a single high-security location. (http://www.imb.nrc.ca/imb/sulfolob/indcon_e.html)

NRC-IMB

2. 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/ogmpproj.html)

3. 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/People/lang/FMGP/FMGP.html)

4. Human Chromosome 14 Expressed Sequence Map

An expressed sequence map of 14q32 project is funded by a grant from the Canadian Genome Analysis and Technology (CGAT) to the CO, Diane Wilson at Medical Genetics H.S.C. (Toronto). The goal of the project is to identify and map all of the expressed sequences, or genes, in the terminal 10 percent of chromosome 14, using yeast artificial chromosomes (YACs) for the region as a starting resource. The region is believed to include genes for colon cancer, neuroblastoma, seizures, eye disease and heart malformations. The database of chromosome 14 mapping and sequence information will be developed. The data will be presented to the user in graphic form. Newly generated data will be added to the database and accessible to all chromosome 14 investigators.

D. Projects Developing Analytical Tools for the Scientific Community

1. Models and Algorithms for Genomic Evolution

Sankoff David (Mathematics and Statistics, University of Montreal) has a project funded by CGAT which focuses on evolutionary mechanisms which operate at the genomic level without affecting the composition of individual genes (e.g. insertion/deletions of genes or segments of chromosomes, transposition, inversion, reciprocal translation, duplication, fusion etc.). The aim of this project is to elaborate a common mathematical framework for investigating these processes, while performing genomic comparisons in inferring evolutionary divergence. The project will include construction of the first database of paralogous genes and chromosome segments in the human and mouse genomes which will be used to study questions about genome organization and evolution.

2. Computational Issues in Alignment and Sequencing

Derick Wood's group (Computer Science, University of Western Ontario) works on tools for bioinformatic sequence analysis; the project is funded by the CGAT.

II. Chemistry

A. Data on Properties of Nanocomposites, Magnetic Nanostructures and Polymers

Prof. Françoise M. Winnik, Department of Chemistry, McMaster University, Hamilton, Ont., is developing a database on the properties of nanocomposites, magnetic nanostructures, and polymers involved in surfactant interactions, as well as amphiphilic polymers. These data are expected to be useful in a variety of applications, including surface chemistry, studies of reactivities in organized media and models of cytoskeletons.

B. Databank on n-octanol: Water Partition Coefficients

Dr. James Sangster, Sangster Research Laboratory, Montréal, Québec, has maintained and upgraded a databank on n-octanol:water partition coefficients of a large set of molecules, important in many chemical and biochemical fields. These data are essential in making comparisons and potential predictions of biochemical activities of potential drug molecules as well as environmental toxicants. ***SRL***

C. Polyaromatic Hydrocarbons Database

Prof. Bruce Greenberg and Prof. G. Dixon, University of Waterloo, Ont., are developing a databank on the photochemical activities and aquatic toxicity of polyaromatic hydrocarbons, as well as their photooxidized products. 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.

D. Database Relating Pesticide / Herbicide Activities to Metal Contaminants

Prof. P. Ming Huang, Department of Soil Science, University of Saskatchewan, Saskatoon, is developing a database of pesticide and herbicide activities in the presence of metal contaminants, affecting the soil - plant root system interface. Using similarity search techniques, the database is expected to enhance the predictability of adverse effects of new pesticides and herbicides entering the market.

E. Hemoglobin Binding Affinity Database

Prof. Kannan Krishnan, Dép. Médecine du Travail et d'Hygiène du Milieu, Faculté de Médecine, Université de Montréal, Québec, has developed a database on the hemoglobin binding affinity constants of a large series of organic molecules. This database is already being applied for the study of some of the adverse effects of toxic substances.

F. Toxicity of Metals Database

Prof. Beverly Hale, University of Guelph, Ont., and Prof. Francine Denizeau, Dép. Chimie, Université du Québec à Montréal, Québec, and coworkers, are developing a database on the toxicity of metals, including Cadmium and Zinc, with special emphasis on their uptake by grain varieties.

G. Shape Databases for Drug Design Strategies

Prof. Paul G. Mezey, Dept. Chemistry, University of Saskatchewan, Saskatoon, has extended the molecular shape database (MEDLA) to a series of haloalkanes, alcohols, aldehydes, ketons, and esters. The polyaromatic hydrocarbon (PAH) shape database has also been substantially updated. These shape databases have already been applied by the phamaceutical industry in calibrating various drug design strategies in new lead search, and
additional applications have been completed in toxicological risk assessment within the framework of CNTC (Canadian Network of Toxicology Centers) Quantitative Risk Assessment project.

UOS

III. Crystallography

A. NRC Metals Crystallographic Database (CRYSTMET)

Work continued on updating the database making it exhaustive in coverage to 1913 and containing about 60,000 entries. The database may be licensed for private or multiple use and it is also available online via the CAN/SND and STN services.

B. 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.

C. 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.

D. 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

CAN/SND

IV. Geoscience

A. Geothermodynamic Database

A revised, internally consistent thermodynamic database of end-members and solid solutions has been completed for important high temperature metamorphic minerals: olivine, orthopyroxene, garnet, cordierite, clinopyroxene, biotite and ilmenite. This database is utilized by the TWQ thermobarometry software package which has been improved and can be obtained by anonymous ftp from emr1.emr.ca (gsc/berman directory). In experimental measurements, equilibrium Al_2O_3 contents were defined for Fe-, Mg- and Fe-Mg biotites in equilibrium with sillimanite, quartz, sanidine and water.
(http://gold.gsc.emr.ca.) ***GSC***

B. Canadian Mineral Occurrence and National Mineral Collection Databases

Two mineralogy databases are maintained by the Mineral Resources Division (MRD) of the Geological Survey of Canada (GSC). The Canadian Mineral Occurrence File (MOFILE) presently has about 27,000 occurrences listed from about 5000 localities. The National Mineral Collection catalogue has about 19,000 records on individually catalogued specimens (world-wide) in the collection. WWW access is planned in the near future.

GSC

C. National Geochemical Reconnaissance Database

MRD also maintains the National Geochemical Reconnaissance database that contains 4.2 million determinations for 164,000 stream and lake sediment samples representing 2.2 millions km^2 of Canada. It is being moved to MS-SQL in preparation for permitting access, selective retrieval and purchase of data over the Internet. The data are of interest to both geoscientists and others concerned with the chemistry of the environment.

GSC

D. National Mapping Program

The GSC's National Mapping Program (NATMAP) is producing multidisciplinary, digital databases (GIS, geochemistry, geophysics, geology). These databases can be obtained through the GSC's Continental Geoscience Division (CGD) for projects near completion (Shield Margin, Slave Province). ***GSC***

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 point source databases. This software maintains a relational database, especially tailored for geological data and spatial operations, and provides linkages to CAD or GIS software for visualizing and graphically adding or editing the field data. The latest version operates in DOS or WINDOWS and incorporates user-defined glossaries of geological terms throughout the data entry process.
(http://gold.gsc.emr.ca) ***GSC***

F. Canadian Geodetic Information System (CGIS)

The Geodetic Survey Division of the GSC is establishing the Canadian Geodetic Information System (CGIS) as the integration of existing national gravity and survey control (horizontal and vertical) databases. The CGIS is a UNIX/Oracle database of approximately 5 GB with public access through an electronic bulletin board service linked to the World Wide Web. (http://www.geod.nrcan.gc.ca) ***GSC***

G. National Coal Inventory Database

The Calgary office of the Geological Survey maintains a National Coal Inventory data base with primary geological, analytical and environmental data collected from sources within the private sector and provincial agencies. Some 60,000 boreholes covering 70% of Canada's thermal coal resources have been digitally captured. They also provide interpretive GIS based spatial datasets, derived from 3D modelling of the primary data, which address economic, geological and environmental issues / constraints associated with coal deposits.

GSC

H. Geomagnetic Data

The Geomagnetism Program of the Geological Survey of Canada (GSC) maintains and updates the archive of Canadian magnetic observatory digital data. This archive of about 5 GB contains high-resolution data from 13 observatories for the past 20 years plus historical data back to the time of the International Geophysical Year and earlier. It is accessed by researchers and others from all parts of the world. The Internet has greatly increased the speed with which data can be provided to users and an automatic DRM (data request manager) using electronic mail is in operation. The recent opening of a WWW site is resulting in a significant increase in the frequency of requests. The presence on the WWW is also resolving a contentious issue with the World Data Center system, since linkages between the Canadian geomagnetism website and the World Data Center website are being established that will protect the interests of Canada. (http://gdcinfo.agg.emr.ca/geomag) ***GSC***

I. Seismological Data

The Seismology Program of the GSC maintains and updates the archive of Canadian seismological data from the Canadian seismograph network. The archive contains a large volume of older analog seismograph records dating back to the early 1900s and microfilm, all stored in environmentally controlled conditions. The modern data archive contains digital data from 1980, amounting to more than a Terabyte of time series data. In addition the archive contains digital data from about 1966 from the Yellowknife seismic array, used in nuclear explosion detection studies, and also first-level derived data in the form of earthquake epicentre locations for Canada. An automatic DRM using electronic mail is in operation and is heavily used. Direct links exist with the International Data Center for Seismology in Washington DC. A WWW site provides derived data such as epicentres, current information on recent earthquakes, and a catalogue of data availability.
(http://www.seismo.emr.ca) ***GSC***

J. Aeromagnetic Data

The Regional Geophysics group of the GSC operates the National Database for Aeromagnetic Data for Canada. Data range back to 1947, with the early analogue maps having been converted to digital form. The coverage is of 80% of Canada at a regional scale. Data holdings amount to about 6 GB. Data are available in many forms: as point value, gridded sets, plots at any scale, in any format and on any media type, including electronic mail and the Internet. An online ordering system (Auto DRM) is under test. A WWW site is in operation, containing a full catalogue of data and services. Interactive use of data through the WWW is under development.
(http://gdcinfo.agg.emr.ca) ***GSC***

K. Gravity Data

The Regional Geophysics group of GSC also operates the National Database for Gravity Data for Canada, in conjunction with the Geodetic Survey of Geomatics Canada. Data extend back to the mid-1940s and are mostly at regional scale (about 10 km data point spacing). Data holdings amount to about 2 GB. Data availability and related services are as described above for the aeromagnetic database. Both aeromagnetic and gravity data are now used frequently as "layers" within GIS systems.
(http://gdcinfo.agg.emr.ca) ***GSC***

L. Radiometric data

The Mineral Resources Division of the GSC 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 requests are filled in the form of gridded data, or preferably in the form of line-data. Increasing demands for data are coming from the exploration industry. The database is currently being converted from a minicomputer to a PC system. It is not yet accessible via Internet and the WWW, though that is an intention for the future. Index maps of data are available.

GSC

One of the common issues for all of these geophysical database operations is that of data ownership and copyright. In the rapidly evolving Internet the traditional ownership and copyright provisions are being undercut, with no clear replacement mechanism to protect the interests of the institutes that have expended large amounts of money to collect the data. While recognizing the importance of making data available to as wide a user base as possible, we have to protect the sources of data in these times of reduced funding levels.

M. Standards

In the area of geoscience data standards, Canadian initiatives have encouraged the formation of a new International Standards Organisation Technical Committee entitled ISO/TC211 Geographic Information/Geomatics. The Technical Committee has five working groups:

WG 1 Framework and Reference Model (convener United States)
WG 2 Geospatial Models and Operations (convener Australia)
WG 3 Geospatial Data Administration (convener United Kingdom)
WG 4 Geospatial Service (convener Norway)
WG 5 Profiles and Functional Standards (convener Canada)

Many countries have initiated efforts over the past ten to fifteen years to produce national geoscience data standards, especially as concerns the transfer of digital geographic information between users with different computer systems and environments. It was therefore only a matter of time before this important field should have been taken up by the ISO. The long-term objective from the Canadian perspective is to move away from Canadian-developed standards to participate in and adopt ISO standards. To further this objective, Canada has established a Canadian Advisory Committee to ISO/TC211. This activity is coordinated by the Canadian Committee on Geomatics, which is the official committee representing the Standards Council of Canada to ISO on these matters. Canadian participation in the above efforts is very strong. The Chairman of Working Group 5, Profiles and Functional Standards, is David McKellar, Department of National Defence. The Canadian Advisory Committee has established subcommittees to mirror the activities of the ISO/TC211 structure. The following have been nominated to chair these subcommittees:

Geographic Information Standards	Yves-Luc Hudon, Gouvernement du Quebec
Geospatial Data Modelling	Tim Evangelatos, Fisheries & Oceans Canada
Geospatial Data Administration	Valerie Hume, Department of Indian Affairs and Northern Development
Geospatial Services	Dave Coleman, University of New Brunswick
Functional Standards	Pierre Beaulieu, Department of National Defence

The schedule of work of ISO/TC211 is extremely ambitious. A positive outcome would be the winnowing of existing standards to identify the strong contenders. Some aspects of the work on temporal data standards may be premature at this time, but are included for reasons of completeness.

CCG

The related Canadian MERCATOR initiative embodies three major thrusts: geospatial information standards, data warehousing, and related software tools development. Across Canada there are a number of similar activities on-going that share common elements such as exploiting new RDBMS technology. A key strategic relationship between the Department of National Defence, the Directorate of Geomatics, the Canadian Hydrographic Services and the B.C. Ministry of Environment, Lands and Parks has been initiated. An initial MERCATOR Project has been defined and an announcement made by CANARIE. The MERCATOR Project will use HH Code technology developed cooperatively between the Canadian Hydrographic

Service and Oracle Canada. The MERCATOR Project will implement, for example, the various data models from SAIF, DIGEST and S-57 into Oracle spatial data option, formerly Oracle-7 multi-dimension. The result will be a proof of concept to implement a massive geospatial database online -- the MERCATOR Warehouse. Existing development activity has contributed significantly, such as the Chartnet initiative led by the Canadian Hydrographic Service, and in British Columbia where Land Data BC is being constructed to provide online timely access to a variety of information products using the World Wide Web.

To ensure that the MERCATOR standard becomes a reality, it will need to incorporate and build on existing standards such as DIGEST, S-57, and on modules of SAIF. This resultant body of work will go a long way to influence strongly the direction of standards development within ISO TC211.

N. Access

Access to geospatial data in Canada is being facilitated by the Inter Agency Committee on Geomatics (IACG).

A Canadian metadata standard for geospatial data has been established and adopted by the Canadian General Standards Board. This standard is generally compatible with the U.S. standard for digital geospatial metadata but has the benefit of being condensed and less onerous to use, which will encourage its more widespread adoption.

The database collected for the report on Current Status and Trends in Federal Digital Geographic Data in Canada, which is a complete description of geospatial datasets in federal government departments in Canada in 1991, will be made digitally available, and there are plans to update this in 1996.

The Canadian Geographical Information Systems Sourcebook, previously available in hardcopy, will also be made available in digital form.

It is anticipated that these new services will be available in the Spring of 1996 on the Internet. (http://www.geocan.nrcan.gc.ca/iacg/) ***GEOCAN***

V. Environment

A. Great Lakes Information Network

This online database about the Great Lakes and their basins (8 US states and Ontario) is maintained by the Great Lakes Commission. Currently it contains: library of legislative and policy

developments; business, industry and socio-economic statistics; toxic air emissions; human health and lake levels data.
It is available via gopher (gopher.great-lakes.net at port 2200) or via the WWW. (http://www.great-lakes.net:2200/0/glinhome.html) ***GLC***

B. Computerized mining libraries

Several databases of mineral resource and land use information have been developed by provincial geological surveys.

Ontario:
A computerized mining library, known as the Earth Resources and Land Information System (ERLIS), is accessible for a fee through the computer centres of the Ministry of Northern Development and Mines in Ontario. ERLIS contains 120000 maps along with mining databases and 1.5 million pages of mining related documents (480 Gbytes); it includes GIS, database and document processing software for accessing the information.

British Columbia:
Mineral Data BC is an integrated, networked desktop mapping and data retrieval system for which a full description can be accessed from: http://www.empr.gov.bc.ca.

C. Occupational Health and Safety

The Canadian Centre for Occupational Health and Safety (CCOHS) CDrom databases (MSDS, chemical profiles, noise level measurements, legislative information, research references) have now been put on the internet along with their BBS system.
(http://www.ccohs.ca) ***CCOHS***

D. Databases for Environmental Analysis: Provincial and Territorial Governments (Cat. No. 11-529E)

This publication is a joint product of Statistics Canada and the Canadian Council of Ministers of the Environment. It presents an inventory of Provincial and Territorial databases useful for environmental reporting. With over 800 databases listed, each one has a detailed description which lists: purpose, contact information and much more. As a special feature, a diskette for micro computer users is also included. It contains all the information in the report and allows users to search the database on a variety of fields, including "Key words".

E. National Accounts: Natural Resource and Environmental Accounts: The Greening of Canada's National Accounts

Statistics Canada is developing a set of natural resource and environmental accounts that will serve as satellite accounts to the System of National Accounts. There are four distinct components:

The natural resource stock accounts record the known size and composition of Canada's natural resource assets as they evolve over time, in both physical and monetary terms. The links between the value of Canada's natural resources and traditional national accounts provide a tool to measure national wealth and sustainable development.

The natural resource use accounts record in physical terms, when and how non-produced goods and services are brought into the economic sphere and used in production and consumption activities, and highlight the role of selected produced goods that are important in analyses of certain environmental issues.

The waste output accounts record in physical terms, the types and quantities of waste products that are generated in the economy and relate these to the flow of output.

The environmental protection accounts identify current and capital expenditures, by business, government and households, that are intended to conserve or protect natural resources and the environment. Supply side information provides a perspective on the economic benefits of spending on the environment. **STATS CAN**

F. Databases for Environmental Analysis (Cat. No. 11-5320XDE)

1. Description

Databases for Environmental Analysis is a collection of descriptions documenting the characteristics and contents of over one thousand federal, provincial and territorial government databases that are useful for the analysis of environmental conditions and trends. The inventory is a tool to assist researchers investigating the relationship between human activities and environmental conditions and trends. We recognised the need for such an inventory from our own experience in developing statistical reports on human activities and the state of the environment.

The inventory guides users to sources of data and it offers information that assists the person in obtaining data from the database. It also contributes to a timely and comprehensive picture of environmental databases held by the federal, provincial and territorial governments in Canada and assists in the identification of data gaps in existing environmental information.

A typical entry in the inventory contains a summary description, the name of someone to contact for more information, and details about the database contents and coverage. Click here to see an example of a typical page.

2. Obtaining the Inventory

The full inventory is available from Statistics Canada on diskette in Folio VIEWS=AE format ($75 in Canada; $90USD in the USA; $105USD elsewhere). The Folio VIEWS=AE software to read and search the inventory is included on the diskettes. A complete list of the titles of all the databases in the Inventory, organised by jurisdiction, is also available.

The project is a combined effort involving Statistics Canada, Environment Canada, the Canadian Council of Ministers of the Environment, and all the Federal and Provincial Government agencies that the collection represents.

Collaboration with other cataloguing projects has been an important element of the work. Major sources that were drawn upon are the Atlantic Coastal Zone Database Directory, the TERRAMON Database held by Memorial University of Newfoundland, the 1991 Environment Canada Catalogue of Environmental Data in Atlantic Canada, REPEN produced by the St. Lawrence Centre of Environment Canada, the Electronic Inventory of Environmental Data for the Hudson Bay Region collected by Environment Canada (and now held by the Department of Indian Affairs and Northern Development), and the British Columbia Resources Inventory Committee.

The new Folio VIEWS=AE product is the third in a series. Statistics Canada previously published Databases for Environmental Analysis: Government of Canada in 1992 and Databases for Environmental Analysis: Provincial and Territorial Governments in 1994. Each publication contained an electronic version of the database on diskette. The new Folio VIEWS=AE product contains both the Federal and Provincial inventories including the latest updates of the database descriptions.

3. Forthcoming Products

(September 1996): Studies in National Accounting, No. 4: Accounting for Canada's Mineral Resources (Cat. No. 13-603E)

This publication will cover estimates of the physical quantity and monetary value of Canada's mineral reserves including oil, natural gas, coal and nine metals. It will also include a complete description of the methodology and the data sources used. An analytical section will show remaining life of reserves, marginal finding costs, etc.

(June 1997) Canadian Environmental and Resource Accounts

This publication and associated database will summarize the results of Statistics Canada's Natural Resource and Environmental Accounts Programme. It will include detailed tables on the physical stocks of resources (oil and gas, minerals, forests, land), their use and values.
Also included will be details on the generation of wastes and expenditures on environmental protection.

STATSCAN

VI. Materials Properties Data

A. 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.

154 Nuclear stations from around the world responded to the survey initiated by the International Atomic Energy Agency (IAEA) in Vienna. The data were gathered and analyzed by a panel of specialists from six countries, including Canada, which are involved in this Coordinated Research Program (CRP). The first results were published in a IAEA Working Material Document in 1995. Following further evaluation of the data, a second publication is scheduled for the end of 1996 or early 1997.

In the fall of 1995 a review of priorities took place and the design of the database structure was transferred by IAEA to OECD/NEA (Organization for Economic Cooperation and Development/Nuclear Energy Agency) where the work will continue.

AECL

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 molding 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 in the selection and comparison of 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. ***NRC-IMI***

VII. Physics - Astrophysics

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 will be moving the Canada-France-Hawaii Telescope Archive to CD-ROM as well to provide on-line access to the data. This year the CADC will also begin 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.
(http://cadcwww.dao.nrc.ca/) ***NRC-HIA***

VIII Thermodynamics

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) ***FACT***

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.

IX. 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. Paul Mezey succeeded Prof. Hugh King as Chairman and three new members, Drs. Denis Crabtree (Astronomy), Maria Korab-Laskowska (Biology) and James Sangster (Thermodynamics) joined the ranks. In addition, Drs. Robert Berman (Geothermodynamics), Larry Speers (Taxonomy) and Barry Wood (Physics) served as Observers by virtue of their membership on CODATA Task Groups. Completing the team as Observers were Drs. Richard Coles (Geomagnetism), John Rodgers (Crystallography) and Andrée Bichon (International Affairs) along with Mr. David Henderson (Royal Society of Canada, Global Change).

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. Through the auspices of the CAN/SND program at CISTI, a WEB site was established for CODATA which links to all the other CODATA activities world wide and includes an electronic version of the Newsletter. In addition, a WEB site was established for CNC/CODATA first, at the

Université de Montréal by Dr. Tim Littlejohn, and currently at the NRC Herzberg Institute in Victoria by Dr. Crabtree.
CODATA - (http://www.cisti.nrc.ca/programs/codata/welcome.html)
CNC/CODATA - (http://cadcwww.dao.nrc.ca/cnc-codata/)

To assist the USNC/CODATA International inquiry on issues related to the transborder flow of scientific data, CNC/CODATA undertook distribution and collection of the questionnaire forms (in both hard- and electronic-copy) in Canada. It is planned to analyze the Canadian responses to try to discern any peculiarly Canadian problems that may exist in this area.

APPENDIX

AECL Mr. C. Seni, Atomic Energy of Canada Ltd., AECL-CANDU, 2251 Speakman Drive, Mississauga, Ontario L5K 1B2. Tel: (905) 823-9060 Fax: (905) 823-8006

CAN/SND Services Coordinator, National Research Council Canada, Canada Institute for Scientific and Technical Information, (CAN/SND), Montreal Road, Bldg. M-55, Ottawa, Ontario K1A 0S2. Tel: (613) 993-3294 Fax: (613) 952-8246, e-mail: cansnd@nrc.ca

CCG David McKellar, e-mail mckellar@ncs.dnd.ca, Department of National Defence

CCOSH Candian Centre for Occupational Health and Safety, 250 Main St. East, Hamilton, Ontario L8N 1H6 Tel: 1-800-668-4284, email: custserv@ccohs.ca http://www.ccohs.ca

FACT Prof. W. Thompson, Royal Military College, Kingston, Ontario K7K 5L0. Tel: (613) 541-5010

GEOCAN Gordon Plunkett, Manager, GIS Operations, Geographic Information Systems and Services Division, Geomatics Canada

GLC Carol Ratza, Great Lakes Commission, Argus II Building, 400 Fourth St., Ann Arbor, MI, 41802-4816, Tel: (313) 665-9135, e-mail: carol.ratza@um.cc.umich.edu

GSC Geological Survey of Canada, 601 Booth Street, Ottawa, Ontario K1A 0E8 Tel: (613) 996-3919

NRC-HIA National Research Council Canada, Herzberg Institute of Astrophysics, 5071 West Saanich Road, Victoria, British Columbia V8X 4M1 Tel: (604) 363-0025

NRC-IMB Dr. Mark Ragan, National Research Council Canada, Institute for Marine Biosciences, 1411 Oxford Street, Halifax, Nova Scotia B3H 3Z1. Tel: (902) 426-1674 Fax: (902) 426-9413

NRC-IMI Mr. Georges Salloum, National Research Council Canada, Industrial Materials Institute, 75 de Mortagne Blvd., Boucherville, Quebec J4B 6Y4
Tel: (514) 641-5222

NSIS Nova Scotian Institute of Science, Science Service Killiam Library, Dalhousie University, Halifax, Nova Scotia B3H 4H8. Tel: (902) 494-3339

SRL Dr. J. Sangster, Sangster Research Laboratory, Suite 402, 3475 de la Montagne, Montreal, Quebec H3G 2A4 Tel: (514) 694-1215

STATS CAN Statistics Canada, Publications Sales, Room 1710, Main Building, Ottawa, Ontario K1A 0T6 Tel: (613) 951-7277 or (1-800-267-6677

UOS Dr. Paul Mezey, University of Saskatchewan, Saskatoon, Saskatchewan S7N 0W0 Tel: (306) 966-4654 Fax: (306) 966-4730