THE DATABASE

File Structure and Data Validity

The unique identifier for each record on the database is the **Laboratory Number**. The information on file for each sample is as concise as possible, but in order to allow the maximum degree of flexibility when searching the database, each record has an average of at least 20 indexes. The database presently consists of 36 fields ranging in size from 1 to 230 characters, plus a text field containing up to 8000 characters (Fig. 1). Each record requires only *ca.* 1 kbyte of disk storage space.

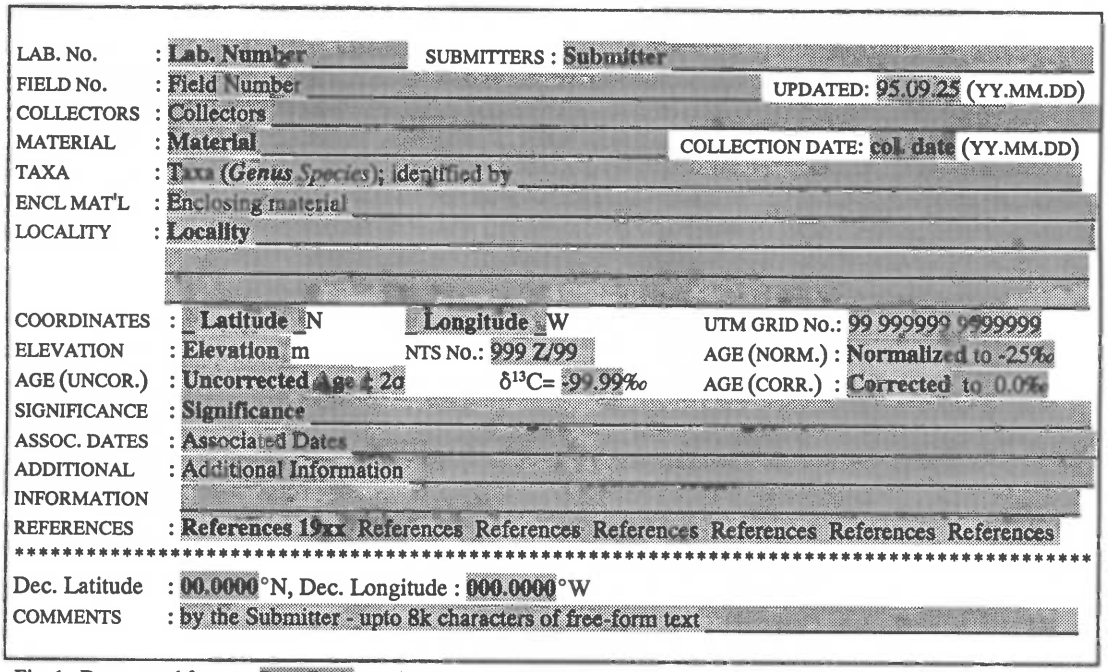

Fig. 1. Data record format. Data fields are shaded; **indexed words** are bolded.

Data verification is the responsibility of the submitter. The Laboratory initiates the addition of new records to the file, but once on file, existing data is modified only with the express knowledge and consent of the submitter. Although the database has been developed as a scientific tool for the research community, its completeness and overall utility depend largely upon altruism, *i.e.,* the clientele's ongoing commitment to add and validate data.

Data Searches and Retrievals

Selected fields such as the **submitter**'s name, **locality**, coordinates (**latitude** and **longitude**), sample **material, taxon** and **significance** have been designated as KEY fields, and entries within these fields have been selectively indexed to facilitate searches, although non-index searches can be processed on a character string in any field. Because of the speed of the searches, the database is typically interrogated in an interactive mode. A request can be simple—for example, "all dates on paleosols" (requires one KEY; select **material** = paleosol); a more complex selection could be "only those dates on peat in Nova Scotia related to Holocene deglaciation" (requires four KEYS; select **material** = peat, **locality** = Nova Scotia; **significance** = deglaciation; and **age** = <10 ka).

[RADIOCARBON, VOL. 38, NO. 2, 1996, P. 271–275]

GEOLOGICAL SURVEY OF CANADA SOIL DATABASE

ROGER McNEELY

Radiocarbon Dating Laboratory, Terrain Sciences Division, Geological Survey of Canada
601 Booth St., Ottawa, Ontario, Canada K1A 0E8

ABSTRACT. The Geological Survey of Canada (GSC) has developed, over the past decade, a user-oriented database, Date Locator File, of Canadian samples dated by the ^{14}C technique. This database presently contains >3500 soil and soil-related dates. The primary category in this suite of dates is peat, as a large portion of the Canadian landscape is covered with this type of organic soil. The data is available *gratis* to all researchers in a large variety of formats from simple lists to complex tables for inclusion in publications. The site localities can also be plotted on base maps suitable for publication. The database is actively augmented on an ongoing basis, but to continue to be relevant, it depends largely on the altruism of the scientific community.

INTRODUCTION

Like most scientific disciplines, Quaternary research has undergone an information explosion in the last few decades. Since the inception of the radiocarbon dating technique more than 45 years ago, the number of samples dated by this technique has steadily increased, especially since the advent of accelerator mass spectrometry (AMS) dating. The large quantity of dates now available has made it increasingly difficult for Quaternary researchers to locate and manage the existing information.

More than a decade ago the GSC Radiocarbon Dating Laboratory developed a database, Date Locator File, which was designed specifically to address the requirements of the laboratory's scientific clientele. We have been using this database to manage the samples dated at the GSC and published in our annual date lists. The utility of this computerized information (database) is only now becoming evident as regional and global syntheses are developed.

THE GSC DATE LOCATOR FILE

The Date Locator File is a computerized database of selected information on published ^{14}C-dated samples. Because the database has been developed to be of use to the Laboratory's clientele, it has been based on the Geological Survey of Canada Radiocarbon Age Data Form and contains primarily field parameters, rather than laboratory variables. The file does not attempt to mimic ^{14}C date lists or other types of reports, but instead provides the most pertinent information in as concise a format as possible, yet allowing for extensive descriptive text. When a date of interest is identified by a search of the database, the user can consult both the references and the submitter to obtain additional information on the sample.

HISTORICAL DEVELOPMENT

In 1982, an interactive proto-database was developed on a mainframe computer using the Scientific Information Retrieval (SIR) database management system, but the increasing availability of microcomputers and associated database programs provided an alternative mode for developing the database. By September 1983, a dBASE prototype of the database was installed on a VICTOR 9000 microcomputer; early in 1985, the database was transferred to the database program CARDBOX-Plus because the database involves primarily text management and does not require the capability for calculations. For the novice user, this program has a very short learning curve and incorporates all the capabilities required to manage textual information. Data entry and verification as well as enhancement of the database capabilities have continued to the present.

The few search limitations that exist are primarily acceptable ones. The more specific a request, the greater the likelihood that records will not be included because of missing data. As with all databases, the search terms must match exactly, although wildcards can be used when there is a known variability in the spelling of a word, *e.g.*, "paleosol" and "palaeosol". If a term of interest occurs in a text field, this text field can be searched, although this is seldom done because of the additional time required to process the request, especially on slower processors (*e.g.*, 80286). In most cases, the required selection can be roughly obtained using KEYS such as coordinates and then a text-search made on the reduced data set.

Data reports

Once a retrieval has been made, the information can be printed or transferred in a wide variety of formats, including the complete data record, and tables or lists of data for publication. For example, the most pertinent data can be printed onto index cards for the client's reference file. Standard formats have been developed of the commonly required information, but customized reports can be designed to fulfil almost any requirement. After a search has been made, the data can be sorted and then transferred directly to a word processor for incorporation into a manuscript. This facility eliminates the necessity for retyping the information, and proofreading and editing the tables or lists.

Data transfers

The master Date Locator File is maintained at the GSC, but extracted subsets of the database, in a variety of formats (*e.g.* ASCII, CSV, tab-delimited), can be transmitted to other locations either electronically or on diskette for the use on compatible microcomputers, or converted to be used on non-DOS machines. After this type of transfer, the maintenance of the data subset is the responsibility of the user. To avoid a multiplicity of databases that may not correspond to the master file, the wholesale transfer of the database is not usually undertaken. Periodic upgrades can be requested from the GSC and reciprocal data-sharing arrangements are encouraged between the GSC and active users.

Plotting

The master file contains the site coordinates, both as the submitter provided them and decimal degrees. The retrieval of subsets of data with decimal degrees facilitates the plotting of site locations on graphics or in Geographic Information Systems (GIS) applications. As an example of this type of plotting, Figure 2 shows the distribution of all dated peat samples in Canada on the database. In addition, the complete UTM coordinates are available on file, and can be used for plotting overlays to verify the original site locations, and thus the latitude and longitude provided by the submitter.

Soil database

Because the Date Locator File has been developed as a tool for Quaternary researchers working in Canada, it includes dates on soil-related materials as old as 40 ka BP. The materials on this database include organics associated with soils and paleosols (including paleosols developed on, and organics related to tephras of various ages), and organic soils (peat) that cover a very large part of the Canadian landscape are a key component of this database. The database is presently comprised of 4350 records on samples from Canada and adjoining states and countries. Because peat is an important entity within the Canadian landmass an Atlas of Peatland Development During the Holocene has been developed (Dyke 1995a). The dating of basal peats as a minimum age estimate of deglaciation in various regions of Canada has indicated that the development of peatlands has occurred during

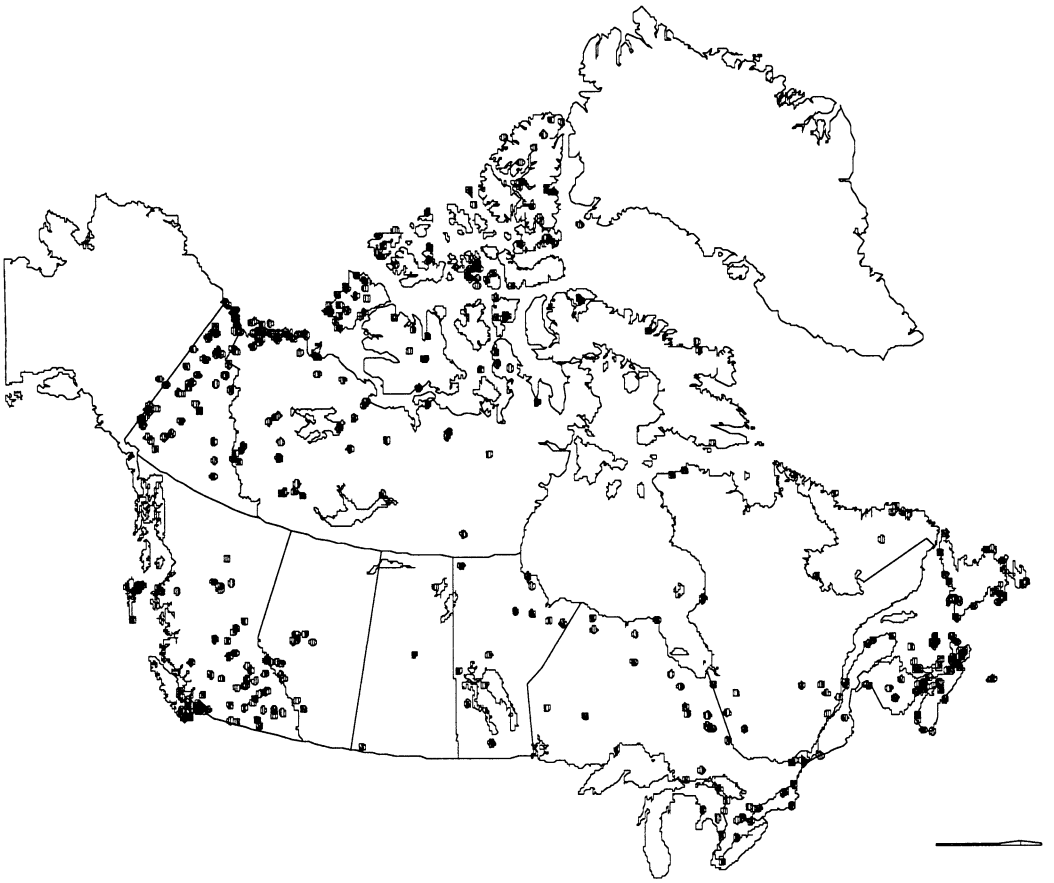

Fig. 2. Dated peat samples in Canada

the Early to Late Holocene and follows a regional pattern of initiation (*cf.* Dyke 1995b for detailed treatment of this data).

In addition to the foregoing uses of the database, the GSC has initiated a project dealing with the distribution and stratigraphy of peatlands (I. Kettles, personal communication 1995) to identify paleoenvironmental changes, and geochemical pathways of heavy metals and anthropogenic contaminants, thereby providing a basis for mapping and characterizing Canadian peat deposits. Many of the Provincial Geological Surveys and Departments of Mines have also initiated programs to ascertain the distribution and thickness of their peatland resources. Dr. Kettle's project will also assess the release of both CO_2 and methane from Canadian peatlands, as it is anticipated that global warming will have a marked worldwide effect on peatlands by increasing the release of "greenhouse" gases from the oxidation of existing peats. On the other hand, global warming, in favorable environments, may cause an increase in the productivity of some peatlands (see Watson *et al.* 1990 for an extensive discussion).

Distribution

Because this database has been developed for the benefit of the scientific community at large, and the information is in the public domain, copies of this database can be obtained *gratis* as a DOS Open File on diskette from the GSC by writing to:

Dr. J.-S. Vincent, Director
Terrain Sciences Division
Geological Survey of Canada
601 Booth St.
OTTAWA, Ontario K1A 0E8
Canada

Acknowledgments

The development of this database is primarily a result of the conscientious work of a large number of summer students over the past decade; to all, my sincerest thanks. We at the GSC would also like to acknowledge our appreciation for the kind generosity of Dr. P. J. H. Richard, University of Montreal, who has allowed his database of samples from Québec to be incorporated into the GSC Date Locator File. Our thanks to Dr. A. S. Dyke, who kindly provided many non-GSC dates to augment the soils database as a result of his projects on peatland development during the Holocene.

References

Dyke, A. S. 1995a (ms.) Paleogeographic Atlas of northern North America. GSC Bulletin. In preparation.

____1995b (ms.) Chronology of wetlands and drylands in northern North America. In preparation.

Watson, R. T., Rodhe, H., Oeschger, H. and Siegenthaler, U. 1990 Greenhouse gases and aerosols. *In* Houghten, J. T., Jenkins, G. J. and Ephraums, J. J., eds., *Climate Change: The IPCC Scientific Assessment.* New York, Cambridge Press: 1–40.