

A Relational Database Structure for Taxonomy  
Records from the Department of Fisheries and  
Oceans (DFO) Long Range Transport of Air  
Pollutants (LRTAP) Biomonitoring Programme

E. Anita Hamilton, Martyn N. Futter and N.H.F. Watson

Department of Fisheries and Oceans  
P.O. Box 550  
Halifax, Nova Scotia  
Canada B3J 2S7

1997

Canadian Technical Report of  
Fisheries and Aquatic Sciences 2164



Fisheries and Oceans  
Canada

Pêches et Océans  
Canada

Canada

Canadian Technical Report of Fisheries and Aquatic Sciences

March, 1997.

A Relational Database Structure for Taxonomy Records from the Department of Fisheries and Oceans (DFO) Long Range Transport of Air Pollutants (LRTAP) Biomonitoring Programme

by

E. Anita Hamilton<sup>1</sup>, Martyn N. Futter<sup>2</sup> and N. H. F. Watson<sup>3</sup>

---

<sup>1</sup> E. Anita Hamilton, Department of Fisheries & Oceans, P. O. Box 550, Halifax, N. S. B3J 2S7

<sup>2</sup> Martyn N. Futter, Dorset Environmental Science Centre, P.O. Box 39 Bellwood Acres Road, Dorset, Ontario.

<sup>3</sup> N. H. F. Watson, Department of Fisheries & Oceans, P. O. Box 550, Halifax, N. S. B3J 2S7

© Minister of Public Works and Government Services  
Canada 1997  
Cat. No. Fs 97-6/2164e ISSN 0706-6457



Correct citation for this publication:

Hamilton, E. Anita, M. N. Futter and N. H. F. Watson. 1997.  
A relational database structure for taxonomy records  
from the Department of Fisheries and Oceans (DFO) Long  
Range Transport of Air Pollutants (LRTAP) Biomonitoring  
programme. Can. Tech. Rept. Fish. Aquat. Sci. 2164.

---

## ABSTRACT

A relational database structure was designed to manage collections of fauna for the Department of Fisheries and Oceans (DFO) Long Range Transport of Air Pollutants (LRTAP) Biomonitoring programme. The present report describes the component of the database structure used for managing taxonomic information. This database structure has been implemented in a number of software packages including dBase IV, Oracle 7.0 and Microsoft Access. The database includes records of taxa collected from all Canadian geographic regions associated with the project. The taxonomic component of the database structure is designed to manage four types of data; the taxonomic hierarchy, allowable names for each taxonomic level, synonymies where two or more names have been used to identify the same organism, and an audit trail to track changes to taxonomic identifications. The taxonomic database can be sorted and/or searched at any level of taxonomic complexity. This flexibility allows for extraction of biological data at any systematic level for analysis associated with other data sets within the LRTAP database

## INTRODUCTION

In 1985, UNECE (United Nations Economic Commissions for Europe) SO<sub>2</sub> reduction protocols were ratified by Canada. Part of Canada's commitment included an agreement to monitor the physical, chemical, and biological responses to the anticipated reduction in atmospheric pollutants. In 1987, protocols for biological monitoring and the location of field sites were established (Shaw, et. al., 1992.). Monitoring has continued at least annually from 1987 to 1995 (Shaw, et. al., 1995.).

A consulting company (MONENCO-AGRA Ltd.) was contracted to design a database and manage data from all contributors between 1987 and 1994 (Minns, Sayer, and Tardioli, 1994.) After 1994 this function was turned over to staff at the Ontario Ministry of Environment and Energy (OMEE), Dorset Environmental Science Centre (OMEE). The original MONENCO-AGRA database design was subsequently modified and has been implemented in ORACLE 7.0 and Microsoft Access.

Early in the project it was recognised that many thousands of faunal records, especially those for arthropods and specifically insects would need to be tabulated for analysis of community composition, etc. Relatively little was known about the constituent fauna of aquatic habitats throughout each of the study areas in the programme. Additionally, the identification of many organisms was in dispute or periodically changed as the taxonomy was updated. For these reasons a method of recording taxa collected which would accurately tabulate records of organisms across regions and could be updated easily was required. When the database was being assembled, several large biological databases were examined<sup>4</sup>. None of these systems were on-line at the time the LRTAP biomonitoring database was being developed. It was clear that, for

---

<sup>4</sup> Storage and Retrieval of U. S. Waterways Parametric Data (STORET) database for planktonic organisms developed by the National Oceanographic and Atmospheric Administration (NOAA) and the National Museums of Canada, Canadian Heritage Information Network (CHIN) database of invertebrate animals.

our purposes, each system contained many extraneous species and lacked many which we had found during the biomonitoring project.

This report amplifies and updates a previous report on the LRTAP database design (Minns, Sayer, and Tardioli, 1994.) describing the new database structure. The suite of taxonomy tables now resident with the database are defined and linkages among tables are identified.

## **TAXONOMIC DATABASE DESIGN**

There are several unique problems involved in storing a large biological data set. First, of the total number of possible taxa only a few are found at every sampling site. Most are restricted to only a few sites within the entire set of locations examined. The resultant large, sparse matrix makes analysis and interpretation of such data extremely difficult. Secondly, many organisms are relatively difficult to identify to the genus or species level, except by specialists. However, all can be reliably identified to some higher level in a hierarchical taxonomic classification scheme. In addition, names, and to an extent classifications, used by different workers may differ and names and relationships of a particular organism may change with time. It was for these and other reasons that an open-ended, multi-field, cross-linkable relational database structure was designed to manage information on collections of fauna.

The taxonomic database design (Appendix 1, and Figure 1.) manages all data associated with taxonomic identifications. This includes the full taxonomic hierarchy for an organism, any synonymies for that identification, reasons for any changes to taxonomic identifications, published references which substantiate taxonomic identifications, and the location of voucher specimens and reference collections from the biomonitoring programme.

The taxonomic hierarchy is managed by using a table containing all levels of taxonomy for each taxon entered in the database, a series of lookup tables with the names of entries at the different taxonomic levels, as well as a unique numeric identifier for a particular taxonomic level from phylum to "qualifier" (Figure 2.). The taxon code is built from a concatenation of the numeric codes from the lookup tables' codes. Each taxon is also given a unique short-form numeric code called the access code which points to the taxon code and "name to use" in the database.

A second set of tables exists to document taxonomic authorities. These are the published references which are used as the basis for the taxonomic identification. There are three tables in this set. The first contains the names of authors who have published a reference. The second contains the bibliographic citation. The final table links reference citations to taxonomic identification codes (access codes).

The third set of tables contains information on the location of reference or voucher specimens. These table structures are site dependent. Different parameters are recorded at OMEE Dorset than are recorded at Scotia-Fundy. The common information that must be recorded in all cases is the taxonomic identifier, the location of the reference specimen, and a means of uniquely identifying that specimen.

The tables described above were combined to produce a single master listing of currently valid taxonomic identifications. The master table was produced for users unfamiliar with relational databases and for those preferring to work from a single, comprehensive listing.

Twenty-one fields were defined in the master taxonomy table (Table 1). Each field was named, identified as a field containing either numbers (numeric) or letters (character) and sized to the largest expected record within that field.

The "access code" identifies each common name or "name to use" numerically and the "taxon code" identifies each "name to use" within its taxonomic hierarchy (Figure 2.).

The "region" is a three letter code which indicates the geographic region(s) where a taxon has been collected (Shaw, et. al. 1992).

The "type #" field indicates that archived definitive specimens collected during the LRTAP biomonitoring study are available and a linking table (SPECIMEN\_LOCATION table) indicates where they are located.

A "memo" may be associated with any record to indicate that taxonomic complications exist. Where there are complications such as name conflicts, equivalent names or spelling differences for the same organism, the user is referred to tables associated with and linked to the master. Naming complications are detailed in the ACCESS\_CODE\_TO\_USE and the TAXON\_CHANGE tables.

The "bibliographic reference" field refers the user to the defining article(s) used to "key" or describe the taxon to a specific "name to use". A list of references accompanies the taxonomic list (Hamilton, et. al., 1997, in press) and is linked to that list through the REFERENCE\_AUTHORS table, the REFERENCE\_INFORMATION table, and the REFERENCE\_TAXON table.

The "taxon code" is a numeric code which identifies each "name to use" within its particular taxonomic hierarchy. Each level of the hierarchy, from phylum to species to qualifier, has a unique identifying number as specified in the LOOKUP\_CODE table. Each organism collected and identified to whatever level within that hierarchy has its own "taxon code", and unique "name to use" and "access code".

The "name to use" identifies each unique taxon collected according to strict protocols (Appendix 3.) established by the biomonitoring working group. Examples of "name(s) to use" include "*Salmo salar*", "*Aeshna canadensis*", "Chironomidae (Family)", and "*Enallagma* sp.". These names refer to unique taxa identified by taxonomists at many levels of the taxonomic hierarchy. In most cases, animals are identified to the species level (e.g. *Salmo salar*, *Aeshna canadensis*). Where this is not possible, either because the specimen is damaged, is an immature form, or is beyond the taxonomists' ability to identify it fully, animals are identified and recorded to the lowest taxonomic level possible.

The "species type" identifies taxa as fish, benthic invertebrates, plankton, etc.

Fields 9-21 place each "name to use" in its respective taxonomic hierarchy as specified in the TAXON table. The "qualifier" field identifies taxa within groups, or taxa of newly described or incompletely described species (*Enallagma cf. geminatum*) and species complexes (e.g. some planktonic organisms).



## DISCUSSION

The design of the LRTAP biomonitoring database has been developed to store, record, analyze and present the large quantity of biological information collected over years from aquatic habitats in various regions of Eastern Canada. The use of simple access codes or scientific names rather than the more cumbersome taxon code simplifies access to the faunal records at any taxonomic level.

Relational databases allow modifications of table designs with relative ease. When organisms new to the database are collected, new entries can be added to the table. Updates of taxonomic nomenclature can be accommodated with relative ease while at the same time preserving old names for historical reference.

The importance of establishing strict naming strategies for organisms must not be overlooked. Any attempts to develop similar designs and lists of animals must include strict protocols for naming organisms. Consistent naming assures that table searches are comprehensive, dependable and accurate. Taxonomists with expertise specific to the organisms collected in field sampling efforts are the main resource for naming these animals. The advice of these taxonomists is integral to the development of the naming protocols used in the taxonomy tables. Controversies regarding the systematics of some animals do exist as do spellings of animal names. These controversies can be minimised in the database by linking the less preferred to more preferred names thereby making searches inclusive rather than exclusive of the preferred and less preferred name(s).

The taxonomy database facilitates sorting records by species, by genera and by family or by any level of taxonomic hierarchy in the taxonomy tables as required. This sorting and search function allows complete and comprehensive handling of organism collections during study assessments and analyses.

## CONCLUSION

This database design allows the inclusion into the LRTAP biomonitoring database of organisms which are identified to the species level equally with those identified at all levels of the taxonomic hierarchy. This facilitates analysis of the whole dataset at appropriate taxonomic levels without the use of many intermediate steps.

We consider that this design is a major advance in data handling over conventional "spreadsheet" approaches to data storage and retrieval systems. The outline we have provided should allow the construction of similar databases for similar studies involving large numbers of taxa.

The database design used for organizing taxonomic information and specimen abundance data from the DFO LRTAP Biomonitoring Programme allows researchers a way of storing and retrieving data from a relational database in a way which fits the traditional hierarchical view of biological community structure.

This taxonomic database design is part of the national LRTAP database which presently resides at the Dorset Environmental Science Centre, Ontario Ministry of Environment and Energy, Dorset, Ontario and can be accessed via the "Sealane.ca" Internet web site. Currently,

there are just over 4000 entries in the master taxonomy list which include freshwater aquatic macroinvertebrates, fishes, amphibians, reptiles and various plankton.

## ACKNOWLEDGEMENT

The authors acknowledge with thanks the helpful comments in the review of this paper by I. J. Davies and E. L. R. Kenchington.

**TABLE 1:**

### MASTER TAXONOMY TABLE STRUCTURE:

| Field Number | Field Name              | Field Type | Field Width |
|--------------|-------------------------|------------|-------------|
| 1            | Access Code             | Numeric    | 6           |
| 2            | Region                  | Character  | 8           |
| 3            | Type                    | Character  | 4           |
| 4            | Memo                    | Memo       | 10          |
| 5            | Taxon Code              | Character  | 39          |
| 6            | Bibliographic Reference | "          | 10          |
| 7            | Name to Use             | "          | 60          |
| 8            | Species Type            | "          | 21          |
| 9            | Phylum                  | "          | 21          |
| 10           | Class                   | "          | 21          |
| 11           | Sub Class               | "          | 21          |
| 12           | Order                   | "          | 21          |
| 13           | Sub-Order               | "          | 21          |
| 14           | Family                  | "          | 21          |
| 15           | Sub-Family              | "          | 21          |
| 16           | Tribe                   | "          | 21          |
| 17           | Genus                   | "          | 21          |
| 18           | Sub-Genus               | "          | 21          |
| 19           | Species                 | "          | 21          |
| 20           | Sub-Species             | "          | 21          |
| 21           | Qualifier               | "          | 21          |



## Appendix 1: Taxonomic Database Table Structures

This appendix contains a description of all taxonomic identification tables currently implemented in the LRTAP biomonitoring database maintained at OMEE Dorset. Each table is described in a similar manner. There are three columns of information for each table. The first column contains the column name for each column in a table. The second column identifies key fields; 'P', indicates that the column has a primary key role, 'AP', indicates that the column is an alternate primary key, and 'F' indicates that the column is a foreign key. The third entry in the table description provides a brief description of the purpose and contents of a column.

Primary keys (P) uniquely identify a single row in a table. Alternate primary keys (AP) also uniquely identify a row in a table, but they generally are not used. Foreign key (F) values uniquely identify a row in another table. Foreign keys are used to create links, or joins, between two or more tables.

Figure 1, showing the table associations and linkages follows these table descriptions.

The first set of tables is used to generate the master taxonomic list. This list is generated from the TAXON table, via a series of LOOKUP\_CODE tables, and the ACCESS\_CODE\_TO\_USE table. The dates and nature of changes to the approved taxonomy are documented in the TAXON\_CHANGE table.

The **LOOKUP\_CODE** tables contain codes and the names of entries at the different taxonomic levels. There is a lookup code table for every taxonomic level from phylum through qualifier. The lookup code tables link codes with the taxonomic name at a node in the taxonomic hierarchy. A concatenation of these CODES is used to develop the TAXON\_CODE for each name to use. For example *Leptophlebia cupida* is an Arthropod of Class Insecta, Order Ephemeroptera, Family Leptophlebiidae, Genus *Leptophlebia*, Species *cupida*. Species are exclusively associated with their respective genus; genera are associated within their respective families; families within their respective orders, etc.

| LOOKUP_CODE Table |     |   |
|-------------------|-----|---|
| Column Name       | Key | Description   |
| CODE              | P   | A numerical code of specified field width for each taxonomic level from phylum to qualifier. Each DESC_CODE has a unique CODE of the specified field width for that particular taxonomic level. |
| DESC_CODE         |     | This column contains the taxonomic name of the record.  |
| UPDATE_DATE       |     | This is the date on which the entry was last changed.   |
| VERSION           |     | The version of the table. This column is used primarily to track historical changes.  |
| COMMENT           |     | This column is a comment field providing information about who added a record, and the rationale for its addition.  |

The **ACCESS\_CODES\_TO\_USE** table maps obsolete or invalid taxonomic identifications to their accepted synonyms. The **ACCESS\_CODES\_TO\_USE** creates a mapping between obsolete or incorrect taxonomic identifications in sample result tables to the approved taxonomic identifier. There is almost always a 1:1 correspondence between access codes and the access code to use. The only time this is not the case is for taxonomic identifiers which have been split into two or more groups.

| <b>ACCESS_CODE TO USE Table</b> |            |   |
|---------------------------------|------------|---|
| <b>Column Name</b>              | <b>Key</b> | <b>Description</b>  |
| ACCESS_CODE                     | F          | This field is the code which uniquely identifies a node in the taxonomic hierarchy.       |
| ACCESS_CODE_TO_USE              | F          | This field is the code which is to be used to identify a node in the taxonomic hierarchy. |

The **TAXON\_CHANGE** table identifies when changes were made to the taxonomic hierarchy and what the changes were. This table is an audit trail which allows researchers to see the history of changes to any taxonomic entry. This information can prove especially useful when taxonomy is subject to change.

| <b>TAXON_CHANGE Table</b> |            |  |
|---------------------------|------------|--|
| <b>Column Name</b>        | <b>Key</b> | <b>Description</b>   |
| CHANGE_ID                 | P          | This column is a key field which uniquely identifies a taxonomic change.   |
| OLD_ACCESS_CODE           | F          | This is an access code which identifies a row in the TAXON table containing an obsolete or invalid taxonomic entry.  |
| NEW_ACCESS_CODE           | F          | This field refers to a record in the TAXON table which has the corrected taxonomic hierarchy for an entry identified by an OLD_ACCESS_CODE.  |
| CHANGE_TYPE               |            | The type of change may be either a merge where two or more taxa are merged to a single taxonomic identifier, a split where one taxonomic identifier is split into two or more identifiers, or a rename where there is a one to one correspondence between the old and new names. |
| REASON                    |            | The reason for which a change was made. This field should reference either a memo from the taxonomist or some other authority for the change.  |
| DATE                      |            | The date on which a change was made.   |

The **TAXON** table contains the name of a taxon, and pointers to the components of that taxon's taxonomic hierarchy. This taxonomic hierarchy is not exhaustive for all groups of organisms. For example, super-family and infra-class are not included.

| <b>TAXON Table</b> |            |  |
|--------------------|------------|--|
| <b>Column Name</b> | <b>Key</b> | <b>Description</b>   |
| ACCESS_CODE        | P          | This numeric column uniquely identifies a record in the taxon table.   |
| TAXON_CODE         | PA         | This column is a concatenation of all the taxonomic level codes for a name to use from the LOOKUP_CODE tables. It can be used to summarize taxonomic categories at a higher level than that inferred by name to use. |
| NAME_TO_USE        |            | This is the approved name for the taxon described by the current record.   |
| STATUS             |            | This field identifies whether or not a taxonomic identification is approved, provisional or obsolete. All approved identifiers have been given that status by the LRTAP biomonitoring working group taxonomist.      |
| PHYLUM             | F          | A code field identifying the phylum to which the taxon belongs.  |
| CLASS              | F          | A code field indicating the class to which the taxon belongs.  |
| SUB_CLASS          | F          | A code field identifying the sub-class, if any, to which the taxon belongs.  |
| FIRST_ORDER        | F          | A code field indicating the order to which a taxon belongs.  |
| SUB_ORDER          | F          | A code field identifying the sub-order, if any, to which a taxon belongs.  |
| TRIBE              | F          | A code field identifying the tribe, if any, to which the taxon belongs.  |
| FAMILY             | F          | A code field indicating the family to which a taxon belongs.   |
| GENUS              | F          | A code field identifying the genus to which a taxon belongs.   |
| SUB_GENUS          | F          | A code field identifying the sub-genus, if any, to which a taxon belongs.  |
| SPECIES            | F          | A code field identifying the species to which a taxon belongs.   |
| SUB_SPECIES        | F          | A code field identifying the sub-species, if any, to which a taxon belongs.  |
| QUALIFIER_TYPE     | F          | A code field identifying the qualifier type for a taxon.   |
| QUALIFIER_NAME     | F          | A code field identifying a qualifier name for a taxon.   |

## Appendix 2: Taxonomic Database Bibliography

Three tables have been created for managing reference information for taxonomic identifications. These tables identify the authors of publications, other publication data, and a table linking references to the taxa they have been used to identify. A routine has been written in Microsoft Access to produce a reference bibliography from the information contained in these three tables.

The **REFERENCE\_AUTHORS** contains author names for publications. A separate table is required to deal with the variable number of authors for a reference. For example, the current document would require three rows in the **REFERENCE\_AUTHORS** table to deal with the three authors.

| REFERENCE_AUTHORS Table |     |   |
|-------------------------|-----|---|
| Column Name             | Key | Description   |
| REFERENCE_ID            | PF  | This is a key field which uniquely identifies an individual reference.  |
| RANK                    | P   | The RANK is the position of an author in the list of authors for a document.                                    |
| LAST_NAME               |     | This is the last name of an author. The LAST_NAME column includes ancillary information such as 'Jr.' or 'III'. |
| FIRST_NAME              |     | This is the first name, or names, by which an author is usually identified.                                     |
| INITIALS                |     | This column contains the initials of the author's first name.   |

The **REFERENCE\_INFORMATION** table contains all data except author information necessary to identify and catalogue a publication. Depending whether the publication is a book, or a serial different columns in the REFERENCE\_INFORMATION table will be filled in.

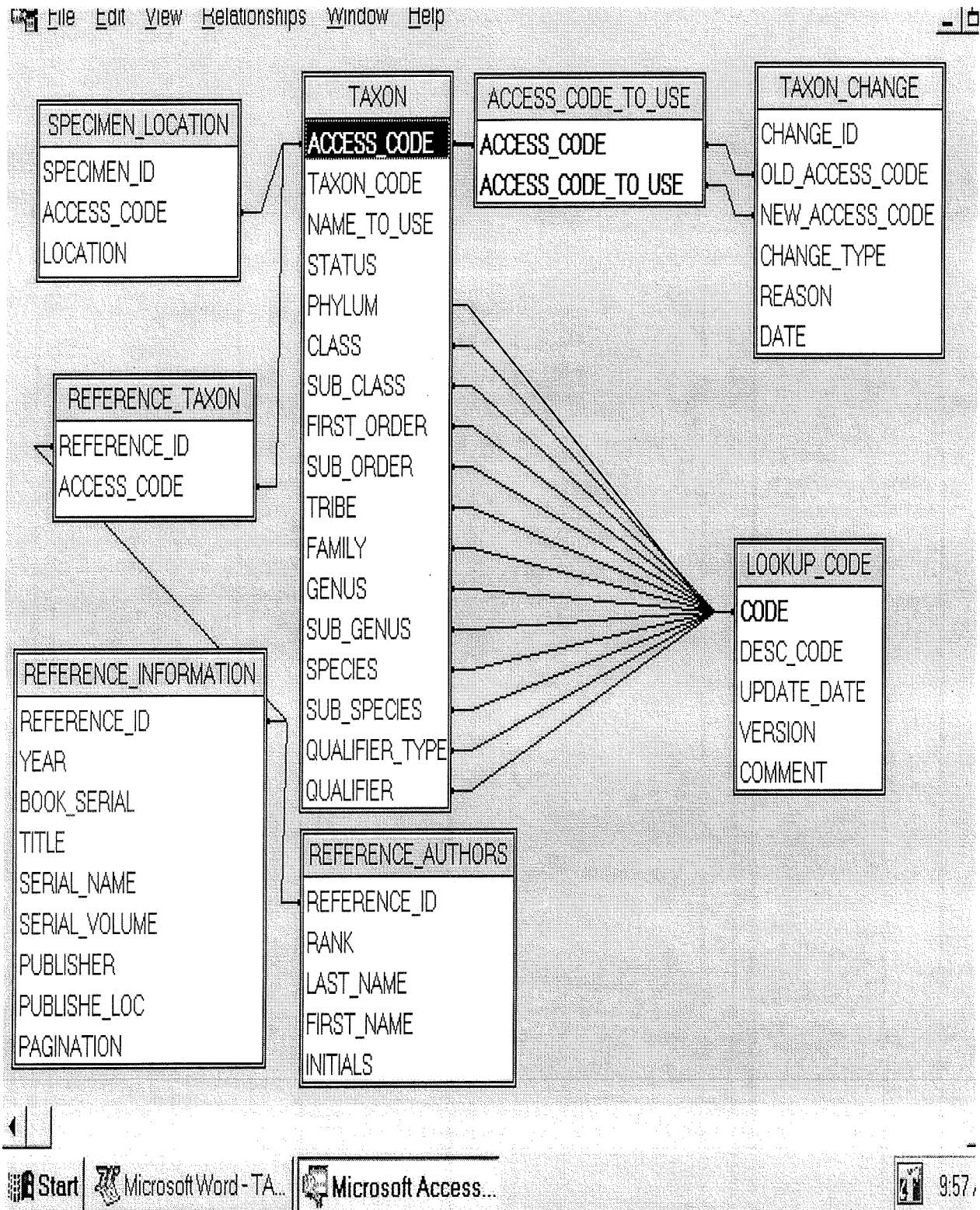
| <b>REFERENCE_INFORMATION Table</b> |   |  |
|------------------------------------|---|--|
| REFERENCE_ID                       | P | This is a key field which uniquely identifies an individual reference.   |
| BOOK_SERIAL                        |   | This column contains a flag field which identifies whether a reference is a book or a serial, such as a journal.                     |
| TITLE                              |   | This is the title of the reference.  |
| YEAR                               |   | This is the year of publication of the reference.  |
| SERIAL_NAME                        |   | This is the name of a serial publication, such as a journal. The approved short form of a serial name should be used in this column. |
| SERIAL_VOLUME                      |   | This is the volume number of a serial.   |
| PUBLISHER                          |   | This is the name of the publisher of a book.   |
| PUBLISHER_LOC                      |   | This is the city or country in which a book was published.   |
| PAGINATION                         |   | This column contains the page numbers of a serial, or the total pages of a book.   |

A separate **REFERENCE\_TAXON** table has been created to deal with the possibility that there could be more than one reference which is used for a single taxonomic identification. If there is never any more than a single reference used to identify a taxon, the REFERENCE\_TAXON and the taxon table may be merged.

| <b>REFERENCE_TAXON Table</b> |            |  |
|------------------------------|------------|--|
| <b>Column Name</b>           | <b>Key</b> | <b>Description</b>   |
| ACCESS_CODE                  | PF         | A code used to uniquely identify a taxon.                              |
| REFERENCE_ID                 | PF         | This is a key field which uniquely identifies an individual reference. |

As well as tracking actual taxonomic identifications, it is important to track the location of voucher specimens. This is done using a **SPECIMEN\_LOCATION** table. The structure of these tables varies from site to site. The following table structure is a minimum supported at all sites.

| SPECIMEN_LOCATION Table |     |   |
|-------------------------|-----|---|
| Column_Name             | Key | Description   |
| SPECIMEN_ID             | P   | This column uniquely identifies an individual organism used as a reference or voucher specimen for a particular taxon. Depending upon implementation, this identification may use more than one column. |
| ACCESS_CODE             | F   | A code used to uniquely identify a taxon.   |
| LOCATION                |     | The LOCATION column must provide sufficient information so as to uniquely locate an individual organism in a museum or reference collection.  |

**Figure 1: Taxonomic Database Table Associations**



**Figure 2.** This illustrates the process of creating the taxonomic code from appropriate entries in the “look\_up code tables” i.e. *Ophiogomphus* is a member of the Phylum Arthropoda, Class Insecta etc. and has no specific epithet or qualifiers, questionable identification or membership in a “species group”.

Scientific “Name To Use”:

***Ophiogomphus (Ophionurus) rupinsulensis***

**Access Code (1533)**

**Taxon Code: 0412000490217302949043809809120000000**

|    |    |    |     |    |     |     |    |      |     |      |    |    |     |
|----|----|----|-----|----|-----|-----|----|------|-----|------|----|----|-----|
| 04 | 12 | 00 | 049 | 02 | 173 | 029 | 49 | 0438 | 098 | 0912 | 00 | 00 | 000 |
| a  | b  | c  | d   | e  | f   | g   | h  | i    | j   | k    | l  | m  | n   |

**Lookup Code tables:** \* segments of Lookup\_Code tables are shown for taxonomic levels specified

**a**

| *PHYLUM<br>ID NAME   |
|----------------------|
| 03 Annelida          |
| <b>04 Arthropoda</b> |
| 05 Bryozoa           |

**b**

| *CLASS<br>ID NAME |
|-------------------|
| 11 Hydrozoa       |
| <b>12 Insecta</b> |
| 13 Oligochaeta    |

**c**

| *SUB_CLASS<br>ID NAME |
|-----------------------|
| <b>00</b>             |
| 01 Branchiopoda       |

**d**

| *ORDER<br>ID NAME  |
|--------------------|
| 048 Neuroptera     |
| <b>049 Odonata</b> |
| 050 Orthoptera     |

**e**

| *SUB_ORDER<br>ID NAME |
|-----------------------|
| 01 Amphipoda          |
| <b>02 Anisoptera</b>  |
| 03 Arthropleona       |

**f**

| *FAMILY<br>ID NAME   |
|----------------------|
| 172 Glossosomatidae  |
| <b>173 Gomphidae</b> |
| 174 Gordiidae        |

**g**

| *SUB_FAMILY<br>ID NAME |
|------------------------|
| 028 Gomphaeshinae      |
| <b>029 Gomphinae</b>   |
| 030 Hageniinae         |

**h**

| *TRIBE<br>ID NAME        |
|--------------------------|
| 48 Nymphulini            |
| <b>49 Onychogomphini</b> |
| 50 Orthoptera            |

**i**

| *GENUS<br>ID NAME        |
|--------------------------|
| 0437 Onocosmoecus        |
| <b>0438 Ophiogomphus</b> |
| 0439 Optioservus         |

**j**

| *SUB_GENUS<br>ID NAME |
|-----------------------|
| 097 Oecata            |
| <b>098 Ophionurus</b> |
| 099 Panchaetoma       |

**k**

| *SPECIES<br>ID NAME       |
|---------------------------|
| 0911 rugglesi             |
| <b>0912 rupinsulensis</b> |
| 0913 rusticum             |

**l**

| *SUB_SPECIES<br>ID NAME |
|-------------------------|
| <b>00</b>               |
| 01 canadense            |

**m**

| *QUALIFIER_TYPE<br>ID NAME |
|----------------------------|
| <b>00</b>                  |
| 01 cf.                     |

**n**

| *QUALIFIER<br>ID NAME |
|-----------------------|
| <b>000</b>            |
| 001 a                 |

### Appendix 3:

#### General Naming Protocols Used to Develop "Name to Use" for the LRTAP-BEM Taxonomy Listing

The naming protocols listed below are not unusual and cover situations in the taxonomic literature. They are described here to indicate how a "name to use" was assigned within the taxonomy table. In every case, animals were named to the level of accuracy of their identification. Each "name to use" entry included its complete hierarchical record from Phylum through to its level of identification.

A routine has been written in ORACLE structured query language (SQL) to enforce the naming protocols within the database.

- 1) The "name to use" (n2u) is the specific name of an animal whether it be its binomial scientific name (e.g. *Leptophlebia cupida*) or another recognised name depending on the degree of accuracy of the animal's identification (e.g. Ephemeroptera (Order)).
  - 1.1) For animals identified with certainty to the species level;
 

e.g. n2u= *Leptophlebia cupida*;  
For *L. cupida* the "Genus" field is LEPTOPHLEBIA; and the "Species" field is CUPIDA.
  - 1.2) For an animal with a subgenus name identical to its genus name (sensu strictu);
 

e.g. n2u= *Cricotopus (ss) festivellus*  
The "Genus" field is CRICOTOPUS; the "Subgenus", CRICOTOPUS; and the "Species" field, FESTIVELLUS.
  - 1.3) For an animal with a subgenus name different from its genus name;
 

e.g. n2u= *Simulium (Eusimulium) congareenarum*  
The "Genus" field is SIMULIUM; the "Subgenus" field, EUSIMULIUM; and the "Species" field, CONGAREENARUM.
  - 1.4) For animals identified to the genus level only;
 

e.g. n2u= *Leptophlebia* sp.;  
This case has LEPTOPHLEBIA in the "Genus" field and no entry in any more specific level of identification .
  - 1.5) For animals having less specific identifications (partial loss of specimen, dried out body parts, etc.);
 

e.g. n2u= Ephemeroptera (Order);  
The file entry for this case goes only to the "Order" field EPHEMEROPTERA.

2) The "Qualifier" field name is used to associate the following qualifiers; cf. (comparable), clx. (complex), gr.(group), nr. (near), var. (variety) with invertebrate identifications.

2.1) For a confident identification of an animal which is commonly identified as part of a group of similar species;

e.g. n2u= *Eukiefferiella brevicealcar*

For *E. brevicealcar* the "Genus" is EUKIEFFERIELLA; the "Species" is BREVICALCAR; and the "Qualifier" field, GR.BREVICALCAR.

2.2) Where an animal is identified to the genus and group level but where the exact species cannot be defined;

e.g. n2u= *Eukiefferiella* sp. *brevicealcar* group

For *E. sp. brevicealcar* group the "Genus" is EUKIERRERIELLA; and the "Qualifier" is GR.BREVICALCAR.

2.3) For animals identified to the species level but where the species name is uncertain or where speciation has not been clearly defined in the literature; **NOTE:** This naming protocol is accepted only with advance written permission of the LRTAP Biomonitoring Programme working group chair.

e.g. n2u= *Ptilostomis* sp.cf.*ocellifera*

The "Genus" field for this case is PTILOSTOMIS; and the "Qualifier" field is CF.OCELLIFERA.

2.4) For animals which have been identified as "near" a species; that is, the animal is near the species indicated but is missing some small characteristic of that species. **NOTE:** This naming protocol is accepted only with advance written permission of the LRTAP Biomonitoring Programme working group chair.

e.g. n2u= *Ceraclea* sp.nr.*arielles*;

In this case, the "Genus" is CERACLEA; and the "Qualifier" field is NR.ARIELLES.

3) For an animal which has been given a temporary name;

**NOTE:** This naming protocol will be accepted only with advance written permission of the LRTAP Biomonitoring Programme working group chair and must be validated by the taxonomic authority using the temporary name.

e.g. n2u= Type III Epler;

For the purposes of this taxonomic listing, this animal is identified at the genus level. No "scientific name" has been assigned at this time. As in all cases the animal is identified within its hierarchical taxonomic structure. Therefore the "Genus" field contains TYPE III EPLER until such time as a generic name is assigned which is recognised by taxonomists.

**REFERENCES**

- Hamilton, E. A., B. Bilyj, I. J. Davies, J. Findeis, M. Futter, A. Kemp, M. Lemire, R. Reid, P. M. Ryan, M. A. Shaw, F. Therrien, N. H. F. Watson, W. White. 1997. The taxonomic listing of benthic macroinvertebrates, fishes, etc. collected across Eastern Canada during the Department of Fisheries and Oceans (DFO) Long Range Transport of Air Pollutants (LRTAP) Biomonitoring Programme, 1987-1995. Can. Tech. Rep. Fish. Aquat. Sci., in preparation.
- Minns, C. K., R. Sayer, and L. Tardioli. 1993. Design report for the DFO National LRTAP Biomonitoring Database. Can. Tech. Rep. Fish. Aquat. Sci. 1987. i-v + 96p.
- Shaw, M. A., I. J. Davies, E. A. Hamilton, A. Kemp, R. Reid, P. M. Ryan, N. Watson, W. White, and K. M. Murphy. 1995. The DFO national LRTAP biomonitoring programme: Baseline characterization 1987-1989. Can. Tech. Rep. Fish. Aquat. Sci. 2032: 69p.
- Shaw, M. A., S. Geiling, S. Barbour, I. J. Davies, E. A. Hamilton, A. Kemp, R. Reid, P. M. Ryan, N. Watson, and W. White. 1992. The Department of Fisheries and Oceans National LRTAP Biomonitoring Programme: Site locations, physical and chemical characteristics. Can Tech. Rep. Fish. Aquat. Sci. 1875.