ISBN 0-9689321-0-X

Label Data Standards for Terrestrial Arthropods

A brief prepared by the
Biological Survey of Canada (Terrestrial Arthropods)

Biological Survey of Canada (Terrestrial Arthropods)

Document Series No. 8 (2001)

Biological Survey of Canada

Commission biologique du Canada

Published by the
Biological Survey of Canada (Terrestrial Arthropods)
Ottawa
2001

Prepared on behalf of the Biological Survey by

Terry A. Wheeler
Department of Natural Resource Sciences, McGill University,
Macdonald Campus, Ste-Anne-de-Bellevue, QC H9X 3V9

John T. Huber
Canadian Forest Service, c/o ECORC, Agriculture and Agri-Food Canada,
Ottawa, ON K1A 0C6

and

Douglas C. Currie
Centre for Biodiversity and Conservation Biology, Royal Ontario Museum,
100 Queen's Park, Toronto, ON M5S 2C6

The Biological Survey of Canada (Terrestrial Arthropods) develops and coordinates national initiatives in systematic and faunistic entomology on behalf of the Canadian Museum of Nature and the Entomological Society of Canada.

The document series of the Biological Survey of Canada comprises invited bibliographies and other miscellaneous publications that are especially relevant to the fauna of Canada.

Additional copies of this document are available in limited numbers from the Biological Survey of Canada (Terrestrial Arthropods), Canadian Museum of Nature, P.O. Box 3443, Station "D", Ottawa, Ontario, Canada K1P 6P4 or as an Adobe Acrobat file on the Survey's web site: http://www.biology.ualberta.ca/esc.hp/bschome.htm

Label Data Standards for Terrestrial Arthropods

Abstract

The data associated with specimens and recorded on their labels are a permanent record of research that is as important as the specimens themselves. This brief provides recommendations on how to prepare data labels for collections of terrestrial arthropods. Given here are standards for label data, to ensure that the data associated with the collecting event are clearly presented and organized, as well as standards for label preparation, to ensure that the labels are clear, useful and permanent. Labels should provide accurate, unambiguous locality information that includes latitude and longitude. Specific recommendations are also provided on how to format information about the date, collector, collecting method and habitat that should appear on labels, and about unique identifier codes if used. Guidelines for preparing computer-generated specimen labels are given, as well as recommendations on paper and printers for both dry (pinned) specimens and wet specimens (preserved in fluid). Label data should be in a format that maximizes the efficiency with which the data can be extracted into databases, data retrieval systems and geographic information systems.

Les normes d'étiquetage pour les arthropodes terrestres

Résumé

Les informations associées aux spécimens sont des données scientifiques permanentes qui sont tout aussi importantes que les spécimens en tant que tel. Vous trouverez ici les normes sur la préparation des étiquettes afin de vous assurer que les informations associées avec l'événement de collecte soient présentées clairement. Les étiquettes devraient fournir de façon précise (incluant longitude et latitude) l'information sur le lieu de collecte. Des recommandations sur la structure de l'information (date, collecteur, méthode d'échantillonnage, habitat) et des codes uniques d'identification (s'il y a lieu) sont aussi fournies. Un guide de préparation d'étiquettes sur ordinateur et des conseils sur les types de papiers et d'imprimantes (pour des spécimens préservés dans l'alcool ou épinglés) sont fournis. Les étiquettes devraient être conçues de façon à faciliter le transfert d'informations dans une base de données et dans un système d'information géographique.

Introduction

Specimens are permanent data points in studies of systematics, biodiversity and ecology and few informed scientists would question the need for careful preparation and preservation of specimens. The collection data associated with specimens are as important and permanent as the specimens themselves. Therefore, standards of presentation of label data should be adhered to as carefully as those for the preparation of specimens.

There has been a recent increase in biodiversity studies and faunal inventories, and widespread use of passive sampling methods like Malaise traps and pitfall traps means that great numbers of specimens are being added to research collections. Many of these specimens are not in the collector's taxon of interest, which means that the material deposited in collections may not be studied for some time. Another result of the increasing volume of material is that the collector often does not label the specimens; this task falls to students and technical assistants not directly involved in the research. Such personnel must also be made aware of the importance of clear, unambiguous, persistent label data. Moreover, there has been a marked increase in the availability of computer equipment and software for generating specimen labels since the publication of guides such as Martin (1977), and an increase in the amount of data incorporated in field studies and sampling programs.

This brief provides recommendations on two aspects of labelling specimens of terrestrial arthropods: standards for label data (presentation and organization of the data associated with the collecting event); and standards for label preparation (guidelines to ensure clarity, usefulness and persistence of label data).

Label Data Standards

Most guides on insect collecting include general instructions on labelling specimens (e.g., Martin 1977, Borror et al. 1989) and more detailed guidelines are occasionally published, usually directed at specific user groups, such as biocontrol workers, quarantine officers and other non-systematists (Huber 1998).

The absolute minimum label data required on any specimen are locality and collecting date. However, the addition of data about the collector, collecting method, hosts (where applicable) and habitat greatly increases the information value of the specimen.

Label data must be unambiguous. Abbreviations and codes that the collector understands may be adequate as long as that individual works with the specimens, but may be incomprehensible to someone studying the specimens 50 or 100 years later. Accordingly, the use of abbreviations and codes should be minimized. Specific examples are discussed in the following sections.

Locality Data

Locality data on insect labels are traditionally organized in the following sequence: **COUNTRY** (in all capitals and unabbreviated, except for USA and, frequently, CAN); **lesser political unit** (province, state, etc., written in full or using two letter postal abbreviations for Canadian provinces and territories and American states); **precise locality** (distance and compass direction to nearest inhabited community, administrative unit such as a park or reserve, prominent landmark such as a mountain or lake, or mileage along a road or railway). While these points of reference work well in most cases, there are many examples of potential problems arising from their use without further locality information:

- Many political entities have changed over time, or ceased to exist altogether. The most obvious are changes in the status or boundaries of countries (e.g., Soviet Union, Yugoslavia), but there have also been historical changes in provinces, territories and cities. For example, in the early part of the 1900s, the Northwest Territories included much of British Columbia, Alberta, Saskatchewan and northern Manitoba.

- Localities that are sometimes well represented in collections are no longer (or never were) found on maps: Laggan is an early name for Lake Louise, Alberta; Frobisher Bay, Northwest Territories is now known as Iqaluit, Nunavut; Mistassini and Mistassini Post, Quebec are actually Baie-de-Post, Quebec and not the village that currently appears on Quebec maps as Mistassini; Erebia Creek, Yukon is a site well-known to entomologists but will not be found in a gazetteer.

- Collecting localities along the Alaska Highway are frequently given in terms of milepost markers; however, the Alaska Highway is approximately 40 km shorter than it was in 1942 and road improvements continue to re-route and shorten it every year. Accurate location of a milepost, therefore, would require cross-referencing to the collecting date. To further complicate matters, Alaska uses historical mileposts (calibrated to 1942 distance), the Yukon uses historical mileposts converted to kilometers, and British Columbia uses actual mileage (expressed in kilometers).

- Localities along roads are often tied to the highway number. However, highways are occasionally renumbered (as recently happened in Ontario), rerouted or replaced by other highways. The now defunct Route 66 in the United States is probably the most famous example of this.

- Private buildings, research stations, businesses, etc. are questionable landmarks because they tend to change name or even disappear. Localities expressed in terms of distance from the Spuzzum Café in British Columbia were rendered useless when that building burned down in 2000. Many specimens deposited in museums were collected at field stations that no longer exist.

- Local names for physical features often do not correspond to official names. A small lake near Carcross, Yukon is known as both Rainbow Lake and Emerald Lake. However, the lake is unnamed on topographic maps, and a search of the Canadian Gazetteer for both names gives one lake of each name in the Yukon, neither of which is the one near Carcross.

- There are often multiple occurrences of the same name. A search of the Canadian Gazetteer for “Beaver Lake, Ontario” produced 41 records, spread throughout the province. Other provinces have a similar predilection for Beaver Lakes.

In many cases obscure localities (including historical names) can be found with the aid of electronic gazetteers, which increasingly are becoming available. Table 1 provides URL addresses for on-line gazetteers for most regions of the world.

Source	URL	Coverage
Geomatics Canada	http://geonames.nrcan.gc.ca	Canada
USGS Geographic Names Information System	http://geonames.usgs.gov	United States
US Census Bureau	http://www.census.gov/cgi-bin/gazetteer	United States
GEOnet Names Server	http://164.214.2.59/gns/html/index.html	World except USA, Antarctica
Getty Thesaurus of Geographic Names	http://www.getty.edu/research/tools/vocabulary/tgn	World
Arizona State University Place Names on the Internet	http://www.asu.edu/lib/hayden/govdocs/maps/geogname.html	Links worldwide

Table 1. Selected on-line gazeteers for georeferencing information

There are published sources for collecting localities as well. Notes on historical records are scattered throughout the literature, often in little known places. For example, Huckett (1965) provided maps of several localities in northern Canada, Alaska and Greenland from which insects have been collected. Hamilton (1997) and Scudder (1997) provided names, latitude and longitude for almost 600 collecting localities in the Yukon and Handfield (1999) gave similar information, including historical notes, for over 3000 localities in Quebec and adjacent regions from which Lepidoptera have been collected. Roughley (1990) provided sources of information about historical localities in the Nearctic and Palearctic regions, as well as advice on interpretation of ambiguous data labels.

Although political and administrative entities are not fixed and are not natural divisions for arthropod distributions, alternatives based on "natural" divisions such as ecozones and ecoregions are not viable for several reasons:

- The limits of ecozones are necessarily coarse and encompass large areas. Ecoregions are smaller, but are also not entirely accurate as currently mapped. Habitat heterogeneity within ecoregions would not be expressed on a label tied to an ecoregion classification. Furthermore, currently available maps of ecozones and ecoregions are not accurate enough to pinpoint a collecting locality.

- Some provinces have ecological subdivisions that differ from federal classifications.

- The description of the ecozone, ecoregion, etc. would take as much space on a label as a political locality description, while use of the numbering system currently employed to designate ecozones and ecoregions would introduce uncertainty into interpreting the label (is it a locality, a date or a code?), and visual scanning of labels for records would be cumbersome.

- Many biodiversity inventories, surveys and catalogs are based on administrative regions such as provinces, parks, reserves, etc. and cross the boundaries of ecoregions.

- Limits of ecoregions are as subject to change over time as administrative divisions. Urbanization, agricultural development and climate change, as well as additional analysis, can all result in changes to the distribution and limits of habitats within ecoregions, and the ecoregions themselves.

- It is easier to locate a politically defined location on road maps, topographic maps and atlases than an ecologically based site. This is important for writing labels in the field.

The most practical solution to all of the above problems is to include **geographic coordinates** on each label **in addition to the politically defined locality**. Latitude and longitude are the most widely used, persistent and easily matched to most available maps. Other systems such as Universal Transverse Mercator (UTM) coordinates, military coordinates, or country-specific coordinate systems are not recommended because they are not indicated on many maps, and may be subject to change in the future. An appropriate level of resolution must be used for latitude and longitude readings. Degrees alone are essentially useless for pinpointing a locality; degrees and minutes are better but still are accurate only to within 1-2 kilometers in southern Canada. It is recommended that latitude and longitude readings be given to the second (95°40′12″W) or decimal degree to three decimal places (95.563°W or 95°40.2′W). This information is easily obtained in the field with handheld Global Positioning Systems (GPS).

With the increasing accuracy and availability of portable GPS units collectors can record accurate georeferenced data for every collecting event. Most currently available models employ 12 channel receivers, which can access up to 12 GPS satellites simultaneously to fix their position; these units have less difficulty in obtaining coordinates in areas with obscured access to the GPS satellites, such as dense forest or ravines. The United States, which controls access to the GPS satellite network, discontinued degradation of the signals (known as selective availability) in May 2000, which means that basic GPS units now record location to within 20 meters, and usually much less. More expensive units equipped with Differential GPS (DGPS) are accurate to within 1-2 meters.

With the popularity of GPS for outdoor activities, there is a wide range of brands and models available at most outdoor stores. A guide to features and prices (in the USA) of several models can be accessed at http://www.gpsnow.com. The Lyman Entomological Museum has used the Garmin® GPS12 since 1998. A newer version of this model (the GPS12 XL) is available from Mountain Equipment Co-op (http://www.mec.ca) for CDN $325.00 (as of summer 2001). It takes less than a minute to obtain a position reading, so that by the time specimens are collected (e.g., flies swept from a roadside ditch) the latitude and longitude can be added to the field label. If used only to record latitude, longitude and elevation for collecting events, a set of batteries will last several weeks. A lower priced model, the Garmin eTREX incorporates many of the same features but costs less than CDN $200.00. Another advantage of most handheld GPS units is that several localities can be entered and saved as waypoints so that the unit can recall the location of a line of insect traps at a field site.

In the absence of a GPS unit, electronic gazetteers are an *a posteriori* source of latitude and longitude for localities. The electronic gazetteers listed in Table 1 will provide coordinates for localities worldwide.

In summary, every insect label should have accurate latitude and longitude data, and this is the only locality information really necessary to incorporate the associated specimen into specimen databases or Geographic Information Systems for geographic analysis and distribution mapping.

Other Collection Data

The **collection date** is important in establishing phenology, activity periods (e.g., flight times), ecological interactions, etc. The preferred sequence is day.month.year, separated by periods, with the year written in full (13.iv.2001). Sampling of longer duration, such as in Malaise traps or pitfall traps, should include start and end dates, separated by a hyphen (10-16.vi.1998 or 28.vi-3.vii.1998). The day and year should be in arabic numerals, the month in roman numerals. Using lower case roman numerals for months will reduce potential confusion between the months i and ii and the dates 1 and 11. In addition to date, time of day may be important for some species (e.g., swarming species, crepuscular or nocturnal species). Time should be expressed using a 24-hour clock.

The name of the **collector** should appear on the label, partly for credit, but also because it can often help to link the specimen to additional data, especially if field notes are published or archived. Multiple collectors should be listed or alternatively a field party or study label can be used (e.g., ROM Field Party, Lyman Gaspé Survey).

Sweeping and hand collecting are no longer the most widely used collecting methods. A range of traps and techniques is used, often with markedly different results in terms of diversity and abundance of species collected. Because of these differences, the **collecting method** is valuable information on a label. Some authors (e.g., Finnamore et al. 1998, Huber 1998) have compiled lists of abbreviations for a range of collecting techniques. These abbreviations are good for workers who have the master list, but can create problems for others in interpreting label data. There are also linguistic considerations; many English-speaking collectors would be confused by abbreviations like "PM" or "PC" if they were not aware of the French terms piège Malaise (Malaise trap) or piège a cuvette (pan trap). If space allows, the collecting method should be written out as fully as possible to avoid ambiguity. It is better to add an extra line to a label than to risk confusion.

Habitat information should be brief, but as informative as possible. Host plant or animal, rearing information and habitat type are all important, especially in ecological studies. Detailed information can be linked to a field code or unique identifier (see below).

Unique Identifier Codes

Unique identifier codes for individual specimens may take the form of bar codes or other machine readable symbols, or an alphanumeric code.

There are published discussions on the advantages and desirability of bar codes (Janzen 1992, Thompson 1994) and in 1993 the Entomological Collections Network passed a resolution on a standard for specimen bar codes (see http://www.sel.barc.usda.gov/Diptera/barcodes.htm). Bar codes have the advantage of being quickly accessible and instantly entered into a database, provided one has the appropriate bar code reader and software. This is often not the case except in the institution that originally attaches the bar code to the specimen; despite the increasing use of bar codes in biodiversity studies, it is unlikely that their use will become standard in smaller collections in the near future, largely because the initial cost of the bar code reader and software and the ongoing cost of producing printed bar code labels is prohibitive for many small institutions on a limited budget. J. Pickering (http://dial.pick.uga.edu) gives approximate prices and requirements for a bar code label system based on a different standard than that adopted by most other institutions to date. One advantage of the system described by Pickering is that the bar code and conventional collecting data are printed on the same side of the same label and are, therefore, both visible. Some institutions place the printed collection label and bar code back to back so that each specimen must be removed from the unit tray to read one or the other types of labels.

Unique alphanumeric identifier codes on labels may be useful for a time, but the master list of codes is subject to loss after several years. Collectors and preparators should ensure that essential locality and date information appears on each printed label. Specimens should **never** be deposited in a collection bearing only a code label; a specimen labelled only LRB01002983 would provide little useful information to a future systematist. Keys to field codes should be posted online or archived in multiple accessible locations, especially if the field notes contain additional ecological information not reproduced on the specimen label.

Label Preparation Standards

Label Data Format

The following sequence is recommended:

COUNTRY: lesser political unit: exact locality (latitude
  and longitude)
date(s), collector(s),
collecting method, habitat
field code or unique identifier

Locality data should always be listed first and in the order specified. The sequence of date, collector, method and habitat data is more flexible and can be altered to better fit the available space. If a unique identifier code is used, it should appear last, on a separate line if possible, to facilitate retrospective data capture. Sample labels are shown below:

CAN:MB: Winnipeg, St. Charles
Rifle Range, Arrowhead Block
(49°54.6’N,97°20.5’W), sweep in
tallgrass prairie, 13.vii.1999
V. Crecco & T. Wheeler

CANADA:QC: Lac St-Francois
Nat. Wildl. Area, NE of Aménag
Therrien (45°00.17’N, 74°30.63’W)
26.v-03.vi.1999, F. Beaulieu
Carex meadow, pan trap T2d

ARGENTINA:Salta: Rosario de
Lerma (24.983°S, 65.583°W)
FIT in disturbed forest
29.i-04.ii.1996, J.P. Johnson
LEM960200203

USA:NH: Coos Co., 1km E Stark
(44°36’N, 71°24’W), sweep along
Ammonoosuc River, 07.viii.2000
T. Wheeler, J. Savage, J. Forrest

USA:WA: Douglas Co., Chelan
(47°50’N,120°W), 150m
2.vii.1993, J.T. Huber
feeding on *Picea alba* foliage

CANADA:QC: Stoneycroft Pond
(45°25.8’N, 73°56.4’W) ex. *Phalaris*
arundinacea, 31.v.2000, emerged
5.vii.2000, F. Beaulieu

Labels will generally have 4-5 rows of data, especially if habitat information or unique identifier codes are included. Where possible, all collecting data should be on the same label; this avoids having to turn specimens upside down to read a second label back to back with the first, or twisting the upper label around to read the lower one.

Sans-serif fonts with uniform line thickness, such as Arial or Univers, are recommended because they are easy to read at small font sizes and parts of letters are less likely to be lost or filled in when printing. Font size for labels should be 3 or 4 point; 3 point is recommended for labels with a lot of data. Single line spacing should be sufficient, but this will depend on the combination of word processing software and printer used, and it may be necessary to reduce the line spacing to 0.9; lines on labels should be close together, but not crowded. Once a suitable combination of font/size/line spacing is found for a particular computer-printer combination, the specifications can be posted near the computer for subsequent users. This is especially useful in busy laboratories or on large projects where a number of people may be generating labels.

Printed labels should be no larger than 17 mm long by 6 mm wide. Using 3 point Arial font, this size will allow 5 lines of data with 29-32 characters per line (including spaces), which will be sufficient for most labels incorporating the recommended information. Labels of the same dimensions using 4 point font will be limited to 4 lines of 22-25 characters.

Cutting and Mounting Labels

Printed labels should be cut as closely as possible to the printed text so that extra white space is minimized. However, care must also be taken to avoid cutting pieces off lines of text. Although guillotine-type paper cutters are fast and efficient, they can produce rough or bent edges on labels and increase the chance of cutting off pieces of text or leaving wide margins around the edge of the label.

The pin should be placed through the centre of the label for pinned specimens; for specimens glued to points or cards the label should be offset so as to provide maximum protection to the specimen above it, and to minimize the space the specimen occupies in the drawer. Pins should be inserted through a space between letters on the label, not directly through a letter.

Labels should be properly oriented on the pin. The long axis of the label should be parallel to the long axis of the body of a directly pinned specimen, or the point or double mount. The text on the label should be readable with the specimen, point or double mount facing to the left.

Label Production For Pinned Specimens

Paper for pinned labels should be white with a smooth surface. For long-term stability paper should be acid free (or archival quality); this information is generally indicated on the packages of commercially available papers. The weight should be heavy enough to hold the pin securely and resist curling. A weight of 36-40 pounds is recommended as a minimum; the cover stock listed in Appendix 1 is heavier, up to 60-pound. Some recommended brands and sources of label paper are given in Appendix 1.

Laser printers and better quality inkjet printers produce print of high enough quality for permanent labels. The resolution should be set as high as possible (at least 600 dots per inch is recommended, 1200 dpi is preferable) to ensure sharp edges and good separation of letters. Many computer programs have been developed over the years for generating insect labels; most of these are variations on Microsoft Word® or Corel Wordperfect® macros. Each label program has its strengths and weaknesses as well as a limited distribution and an apparently brief life expectancy. Many collectors simply generate labels using a standard word processing program and block and copy sufficient labels for their needs. This process is easy and rapid in larger studies where most of the labels have similar data except for selected fields. A file of master labels can be generated and altered as necessary.

Label Production for Fluid Preservation

Two different types of wet storage are considered in this section. The first is the permanent labelling and storage associated with immature specimens, soft-bodied adults or arachnids in ethanol, formalin or other fluids. The second relates to the increasingly widespread use of bulk sample residues as a source of research material. Residues are often viewed as "temporary" storage, but the amount of time specimens may spend in residues, and the repeated disturbance of the sample for sorting in different laboratories, make labelling issues anything but temporary.

Regular label paper should not be used for fluid-preserved specimens because it will break down after a period of immersion. Resistall® paper is manufactured to withstand immersion in fluids (including ethanol and formalin) without losing its dimensional stability. Resistall paper is not acid free; it has a pH of approximately 5.3. However, the acid leaches out of the paper quickly in fluid and, if desired, the fluid can be changed after the paper has been immersed for a short period of time. Paper for fluid preservation does not have to be as heavy as paper for pinned specimens; the Resistall paper listed in Appendix 1 has a weight between 28 and 36 pounds. Resistall paper was almost impossible to find in the early 1990's but, largely in response to requests from the museum community, some manufacturers now produce suitable paper. Sources of label paper with Resistall are given in Appendix 1.

There are still conflicting views on the long-term stability of labels produced on laser printers for fluid preservation. Many users in the early days of laser printers reported letters flaking off wet labels until the labels were unreadable. The problem lies in the strength with which the ink or graphite is bonded to the paper during the printing process. Smooth papers hold ink more securely than papers with a rougher texture, as do heavier weight papers. The difference in paper quality may explain some of the reported discrepancies in longevity of labels. The Royal Ontario Museum has been using laser printed labels in vials of ethanol for over 10 years with no apparent loss of label quality. In contrast, a set of test labels printed on regular bond paper in 1996 in the Lyman Entomological Museum is already partly illegible and a shipment of ethanol-preserved trap residues received at the Lyman Museum from a study in the summer of 2000 had lost entire letters on some labels after only 9 months. Heating a printed page from a laser printer for one minute in a microwave oven may help to bond the text more securely to the page, but there is no indication of how much longer the labels will last using this method. The safest alternative, as used at the Royal Ontario Museum, is to ensure that the back of each printed label bears a unique identifier code written on the label in India ink.

Inkjet printers may provide more resistant labels than laser printers. At the Canadian Museum of Nature, labels for fluid preservation are generated using an inkjet printer with a Lexmark® model 12A1970 black ink cartridge, which prints using fadeproof and waterproof pigment ink. In addition to Lexmark printers, this cartridge is also compatible with some other brands of inkjet printers. Text on labels printed using this cartridge is not soluble in ethanol, acetone, ethyl acetate or ammonia provided that labels are allowed to dry for at least 24 hours before immersion in alcohol (F. Genier, pers. comm.).

The ink used for hand-written labels must be alcohol-proof and dark enough to be readable after long time periods. Ballpoint ink is obviously undesirable as it dissolves in alcohol; soft pencil is better but can still fade enough that it becomes difficult to read after just a few weeks. Labels for preservation in fluids should be written in India ink, or other waterproof black ink. Technical pens, like those made by Staedtler® or Koh-I-Noor®, work well but are expensive and subject to clogging in the field. Disposible permanent ink pens such as Micron Pigma® pens or

Staedtler® Pigment Markers are highly recommended and come in a range of point sizes (.01 or .005 give a very fine line for small script). They are available at art and drafting supply stores for approximately $3.00 each. The ink dries quickly and does not fade or run after several years in ethanol.

Sample residues often spend a long time in storage, with specimens, vegetation, small rocks, etc. rubbing against the label. Recommendations for regular fluid labels apply here as well, with the caution that the increased debris in residues can result in label ink being abraded more easily. An extra printed label or hand-written label inside the sample is insurance against damage. In addition, an extra label attached to the outside of the container can make things easier in several ways: the residue can be identified without having to open the sample to locate the label; a large exterior label can be annotated to indicate which taxa have already been removed from the residue; and a unique identifier code can be associated with the residue to track its progress through various institutions.

Data Retrieval Standards

Carefully prepared labels facilitate retrieval of specimen data for entry into databases for further analysis. Label data should be presented in such a way that they are as widely (and easily) transferable to a range of purposes as possible.

Georeferenced data are easily compatible with computer database programs, since many commercially available database programs us latitude and longitude for spatial organization of data. Latitude and longitude are also compatible with programs for mapping and Geographic Information Systems (GIS) without the need for conversion of locality data.

Biodiversity management software such as Biota® (http://viceroy.eeb.uconn.edu/biota) and BioLink® (http://www.biolink.csiro.au) incorporate label data to build specimen-based databases. Both use latitude and longitude to generate spatial databases, although they are equipped to convert locality data. BioLink has an associated electronic gazetteer and mapping facility. Each of these programs can import data from a wide range of other formats including spreadsheets, database programs and other biodiversity packages. This removes the necessity of retyping or converting data each time one changes programs. Both Biota and BioLink are relatively inexpensive and are designed for biologists as opposed to computer programmers. Many researchers managing specimen information prefer to generate custom databases using more flexible database programs such as FileMaker Pro® or Microsoft Access®. This has the advantage of allowing customized fields and menus, but requires more programming to set up the database. With the recent emphasis on bioinformatics in managing biodiversity information we should expect significant advances in the next decade in how specimen label data are incorporated and organized. Recent initiatives in Canada include the Integrated Taxonomic Information System site (http://sis.agr.gc.ca/itis/), the Global Biodiversity Information Facility (http://www.gbif/org/index.html) and the Biota of Canada Information Network.

Data retrieval systems like bar codes obviously make incorporation of specimen data into a database faster, easier, and less subject to errors. However, bar codes are still relatively rarely used and are much more feasible for newly accessioned material. The cost of retroactively labelling museum material with bar codes is prohibitive. Institutions without machine-readable methods such as bar codes, or with extensive older collections that are to be databased must find other alternatives. A recent comparison of speech-recognition software for retrospective data capture (Sabourin et al. 1999) indicated that computers equipped with inexpensive hardware and software for speech recognition can significantly reduce the time required for data entry from specimen labels compared to keyboard entry. However, even the speech-recognition method is still considerably slower than automated methods. Because of the time and cost required to enter data by currently available methods, any efforts to standardize label data will ultimately save time and money.

Conclusion

The label data associated with specimens are a permanent record of research that is as important as the specimen itself. The data may have to serve multiple functions for multiple users over a long time period; different studies and researchers require different information. Systematic revisions often use label data to establish the identity of examined specimens, to plot general distribution maps or to determine activity periods of a species. Ecological studies require more detailed information on habitats or host species. Ecological monitoring, conservation and management planning require precise georeferenced locality data to determine differences in species composition from site to site and to incorporate specimens into Geographic Information Systems for further analysis.

Collecting events also vary from multi-year, multi-taxon studies yielding hundreds of thousands of specimens, to roadside stops to take a single specimen. Both categories of specimens can be a source of information to future users, provided **accurate**, georeferenced label data are obtained.

A complete label incorporates sufficient data to allow future researchers to repeat the locality, date, ecological conditions and methods of the collecting event. Furthermore, those data must be clear and unambiguous. Ensuring complete label data may seem time-consuming and inconvenient on a sunny summer afternoon, but taking a shortcut now will be a disservice to future users of specimen data. Poorly labelled specimens, no matter how rare or carefully mounted, are ultimately useful only in expendable teaching collections or displays; they are lost to the research community.

References

Borror, D.J., C.A. Triplehorn and N.F. Johnson. 1989. An Introduction to the Study of Insects. Sixth edition. Saunders College Publishing, Fort Worth. 875 pp.

Finnamore, A.T., N.N. Winchester and V. M. Behan-Pelletier. 1998. Protocols for measuring biodiversity: arthropod monitoring in terrestrial ecosystems. Environmental Monitoring and Assessment Network. http://www.cciw.ca/emantemp/research/protocols/arthropod/samproc.html

Hamilton, K.G.A. 1997. Leafhoppers (Homoptera: Cicadellidae) of the Yukon: dispersal and endemism. Pp. 337-375 in H.V. Danks and J.A. Downes (Eds.), Insects of the Yukon. Biological Survey of Canada (Terrestrial Arthropods), Ottawa. 1034 pp.

Handfield, L. 1999. Le Guide des papillons du Québec. Version Scientifique. Broquet: Boucherville. 982 pp.

Huber, J.T. 1998. The importance of voucher specimens, with practical guidelines for preserving specimens of the major invertebrate phyla for identification. *Journal of Natural History* 32: 367-385.

Huckett, H.C. 1965. The Muscidae of northern Canada, Alaska and Greenland (Diptera). *Memoirs of the Entomological Society of Canada* 42: 1-369.

Janzen, D.H. 1992. Information on the bar code system that INBio uses in Costa Rica. *Insect Collection News* 7: 24.

Martin, J.E.H. 1977. Collecting, preparing and preserving insects, mites and spiders. The insects and arachnids of Canada, Part 1. *Agriculture Canada Publication* 1643. 182 pp.

Roughley, R.E. 1990. A systematic revision of species of *Dytiscus* Linnaeus (Coleoptera: Dytiscidae). Part 1. Classification based on adult stage. *Quaestiones entomologicae* 26: 383-557.

Sabourin, K.J., A.T. Finnamore and J. Nagel. 1999. Advances in data capture for museum collections. *Curator* 42: 245-252.

Scudder, G.G.E. 1997. True bugs (Heteroptera) of the Yukon. Pp. 241-336 in H.V. Danks and J.A. Downes (Eds.), Insects of the Yukon. Biological Survey of Canada (Terrestrial Arthropods), Ottawa. 1034 pp.

Thompson, F.C. 1994. Bar codes for specimen data management. *Insect Collection News* 9: 2-4.

Appendix 1 - Selected brands and sources of paper for specimen labels

Brand	Supplier
Paper for pinned labels	
Dulcet Cover Neutral White, 260M, 8.5"x11" acid free (must be cut to 8.5x11) $404.73 per 5000 sheets	Unisource 560 Hensall Circle, Mississauga, ON L5A 1Y1 *Phone:* 905-276-8400 *Fax:* 905-276-8418
Weyerhaeuser First Choice Premium Cover/Card 57-pound, 8.5"x11", acid free $8.35 per 150 sheets	Staples / Business Depot / Bureau en Gros Stores in most major cities *Web:* www.staples.ca
Georgia-Pacific Colour Copier Paper 60-pound, 8.5"x11", acid free $16.44 per 250 sheets	Staples / Business Depot / Bureau en Gros Stores in most major cities *Web:* www.staples.ca
Paper for alcohol labels	
Resistall Paper 28-pound, 100/pack 8.5"x11" Cat. no. 219-288511 $32.15 per pack (100 sheets) (December 1997)	University Products of Canada Division of BFB Sales 6535 Millcreek Drive, Unit #8 Mississauga, ON L5N 2M2 *Phone:* 1-800-667-2632 *Fax:* 905-858-8586
Label paper with Resistall, 36-pound, 8.5"x11" Cat. no. 1223RA (25 sheets),1223RB (100 sheets) US$6.50 (25 sheets), US$24.30 (100 sheets)	BioQuip Products Inc. 17803 LaSalle Ave. Gardena CA 90248, USA *Phone:* 310-324-0620 *Fax:* 310-324-7931 *Web:* www.bioquip.com