The development of an invertebrate database for the lower North Island

Gene N. Browne

DOC RESEARCH & DEVELOPMENT SERIES 269

Published by
Science & Technical Publishing
Department of Conservation
PO Box 10420, The Terrace
Wellington 6143, New Zealand

DOC Research & Development Series is a published record of scientific research carried out, or advice given, by Department of Conservation staff or external contractors funded by DOC. It comprises reports and short communications that are peer-reviewed.
Individual contributions to the series are first released on the departmental website in pdf form. Hardcopy is printed, bound, and distributed at regular intervals. Titles are also listed in our catalogue on the website, refer www.doc.govt.nz under *Publications*, then *Science & technical*.

ISSN 1176-8886
ISBN 978-0-478-14176-4 (Hardcopy)
ISBN 978-0-478-14177-1 (Web PDF)

This is a client report commissioned by Wanganui, East Coast/Hawke's Bay and Wellington Conservancies and funded from the Science Advice Fund. It was prepared for publication by Science & Technical Publishing; editing and layout by Amanda Todd. Publication was approved by the Chief Scientist (Research, Development & Improvement Division), Department of Conservation, Wellington, New Zealand.

In the interest of forest conservation, we support paperless electronic publishing. When printing, recycled paper is used wherever possible.

CONTENTS

The development of an invertebrate database for the lower North Island

Gene N. Browne

Kingett Mitchell Ltd, PO Box 33849, Takapuna, North Shore City 0740, New Zealand
Current address: Flybusters/Antiants, 56B Barry's Point Road, Takapuna, North Shore City 0662, New Zealand. Email: gene@flybusters.co.nz

ABSTRACT

A database of invertebrates that have been found in the lower North Island, New Zealand, was developed. Various organisations were asked to supply data to populate the database. Information that was provided was split into five sections: data on the animal itself, the location where it was recorded, the reference for the original source of the information, data on any voucher specimens maintained, and information on the species' relative rarity. In total, 17 519 individual records were stored within the database; some of these were recorded slightly north of Napier, the arbitrary cut-off boundary. The database was designed to enable easy searches of all the data records both now and in the future. This database is to be a starting point for the Department of Conservation's BioWeb database.

Keywords: invertebrate, database, BioWeb, lower North Island, Geographic Information System, GIS

Browne, G.N. 2007: The development of an invertebrate database for the lower North Island. *DOC Research & Development Series 269*. Department of Conservation, Wellington. 16p.

1. Introduction

The Department of Conservation (DOC) has a responsibility to collect data on the natural and historic heritage of New Zealand. In the past, such data have been stored by individual staff members. However, the year 2000 strategy saw the development of the BioWeb database system, which was developed to integrate various data on a range of subjects into a central, manageable database for DOC staff (DOC 2001).

A Geographic Information Systems (GIS) database of terrestrial invertebrates was developed for the lower North Island, New Zealand, to collate information about indigenous terrestrial invertebrates found in Wellington and Wanganui Conservancies, and in East Coast/Hawke's Bay Conservancy south of the Napier–Taihape Road. This document records the establishment of this GIS database, which is a starting point for the BioWeb database.

2. Invertebrates

'Invertebrate' is a term used to describe any animal without a backbone. Invertebrates are an extremely diverse group, both in form and function, and represent 95% of species belonging to the animal kingdom (Monaghan 1999). To some extent, this categorisation is an artificial one and is not necessarily based on common phylogenetic origins.

The current abundance and diversity of invertebrates is attributable to their relative success. Invertebrates inhabit virtually every type of environment found on earth and are an essential component in the maintenance of ecosystem function. They cycle nutrients, break down pollutants, produce soil, assist with the fertilisation of many plant species, and serve as a source of food to other animals.

The majority of invertebrates are arthropods, which are animals with a chitinous exoskeleton or cuticle that covers the entire body. Phylum Arthropoda includes organisms such as spiders, insects and crustaceans. Other invertebrate phyla include Porifera (sponges), Cnidaria (corals, jellyfish and sea anemones), Platyhelminthes (flatworms), Nematoda (round worms), Annelida (segmented worms), Mollusca (snails, chitons, clams, octopods and squid) and Echinodermata (sea stars, sea urchins, sea cucumbers and sand dollars).

New Zealand has a unique and diverse invertebrate fauna, with a high proportion of endemic species relative to many other countries. Recent estimates suggest that there are c. 80 000 invertebrate species in New Zealand's marine, freshwater and land environments (DOC 2001). It is estimated that c. 20 000 of these species are terrestrial or freshwater arthropods (Watt 1976).

DOC (2001) estimated that invertebrate species outnumber vertebrate species by 230:1 in New Zealand.

New Zealand's invertebrate fauna has evolved over c. 80 million years of geographic isolation (Cooper & Millener 1993), and has been shaped by the changing climate, shorelines, glaciation and volcanism. Unlike invertebrates in other countries, most of New Zealand's invertebrates have evolved without pressure from mammalian predators. Consequently, they lack the behavioural adaptations to counteract the predation strategies used by introduced species. Endemic New Zealand invertebrates also face the additional pressures of reduced habitat, habitat modification and displacement by invading exotic invertebrate pests.

3. Sources of information

Data were gathered from locations throughout the lower North Island of New Zealand (Fig. 1). Data were obtained from the New Zealand Arthropod Collection (NZAC; Landcare Research), Auckland Museum, the Department of Conservation, and Hawke's Bay Regional Council. In addition, various student honours projects and theses from Victoria University of Wellington were examined for relevant information.

Although it was intended that only records of terrestrial invertebrates would be included in the database, some intertidal or estuarine species were also included, as some of these species may be tolerant of a range of freshwater and estuarine habitats.

Information in this database was not checked for applicability, bias or errors (i.e. it was assumed that information on a species that was purported to be an invertebrate found in the lower North Island was correct). Responsibility for the accuracy of specimen identification and location records lies with the authors of the original data. In addition, there may be some spelling errors in names that were brought across from the original texts or databases. However, where possible, checks were put in place to identify any obvious errors.

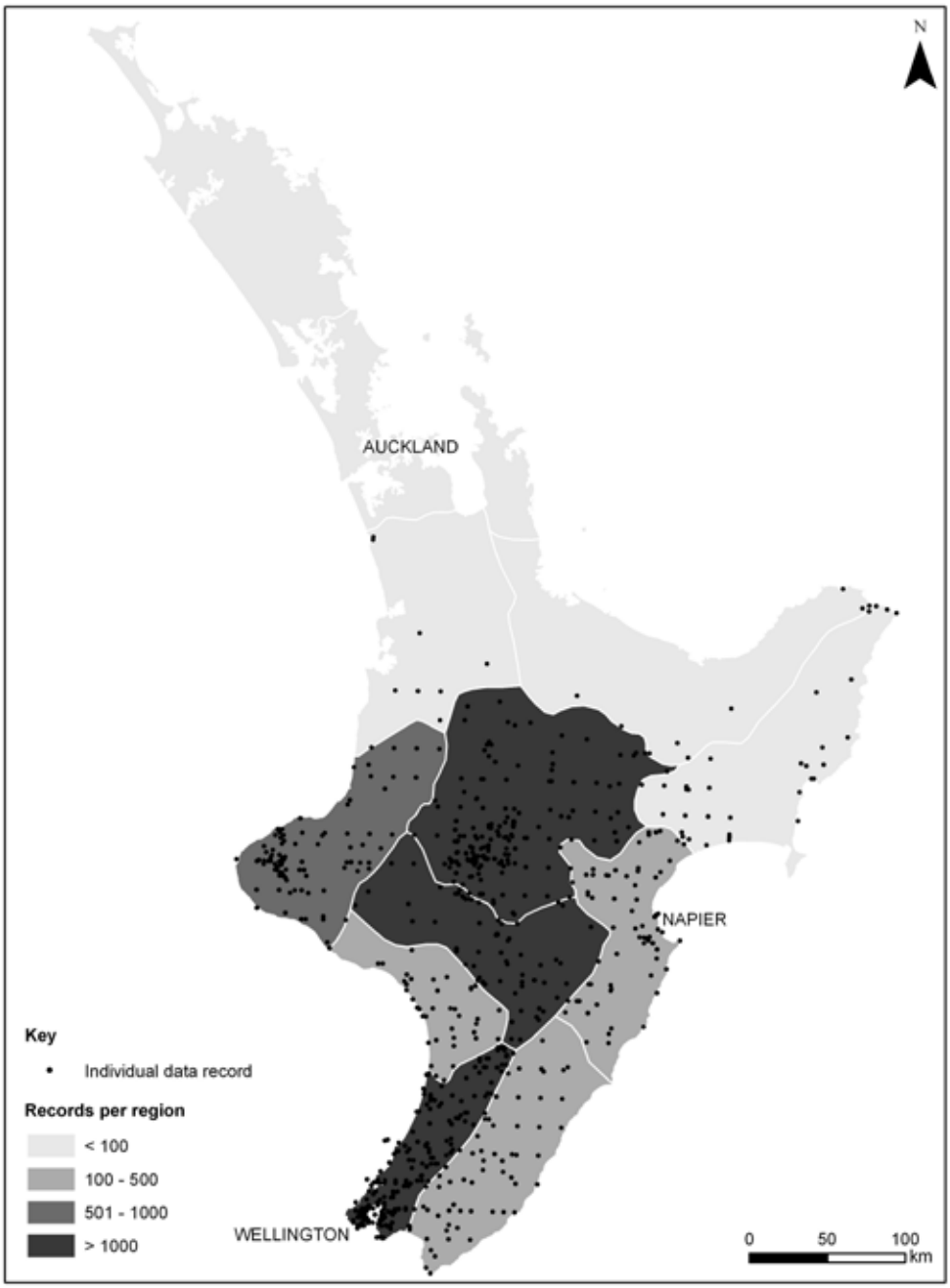

Figure 1. Map of the North Island, New Zealand, showing the locations of invertebrate records included in the database. Depicted regions are based on area codes from Crosby et al. (1998).

4. Database structure

The database follows a hierarchical structure and is divided into five broad sections:

1. The animal: taxonomy and common names of the animal
2. Location: where the specimen was found
3. Reference: the original source of the information
4. Voucher specimens: information about any voucher specimens
5. Rarity: information based on its relative rarity, as provided by DOC (based on Molloy et al. 2002)

Within each of the five sections are various categories. These are specific to each section. The five sections and their associated categories are listed and described in Table 1. The remainder of this report discusses features of the database and provides summaries of the various sections.

4.1 THE ANIMAL

The information within the taxonomy section has been provided so that database users can collate or sort information at different taxonomic levels. Specifically, these are phylum, subphylum, class, order, superfamily, family, subfamily, genus and species. Alternative, common and Maori names have also been provided to add to the completeness of the information.

Some of the data that were forwarded for inclusion were limited in that the original records only listed the genus and species. It was beyond the scope of the database development project to examine and update the full classification tables for such data. In other datasets, individuals had only been classified to a higher taxonomic level, so that genus and species details were missing from the final database (for example, the original data as supplied may have recorded 'Diptera' or 'spider' as the lowest taxonomic level). Where the taxonomic classification was unknown or data were missing, a '.' has been recorded in the database.

In total, the database included 17 519 individual invertebrate records. Arthropoda dominated the database with 16 621 records (95%), followed by Mollusca with 573 records (3%) and Annelida with 201 records (1%) (Table 2). Records for the remaining phyla were relatively scarce.

Within the phylum Arthropoda, Class Insecta contributed 15 348 (88%) of the records. Arachnida was the next most numerous class, contributing 674 records (4% of the total) (Table 2).

Within the Insecta, Family Coleoptera (beetles) were most abundant, with 5139 records (33% of the insects). Coleoptera, Lepidoptera (moths and butterflies), Hemiptera (true bugs), Diptera (flies), Hymenoptera (bees, wasps and ants) and Trichoptera (caddisflies) combined constituted 90% of all insects included (Table 3).

One of the fields in the database holds information on species abundance (i.e. the number of individuals of a particular species that were recorded). However, only 17% of records ($n = 3072$) included abundance data.

The database also records the observer of the invertebrate record. This information was provided for 84% of all records. In addition, 35% of all records provided information on the observed stage of the life cycle for the recorded specimens.

TABLE 1. THE STRUCTURE OF THE INVERTEBRATE DATABASE.

SECTION	CATEGORY	DESCRIPTION
The animal	Phylum	Taxonomic classification
	Subphylum	Taxonomic classification
	Class	Taxonomic classification
	Order	Taxonomic classification
	Superfamily	Taxonomic classification
	Family	Taxonomic classification
	Subfamily	Taxonomic classification
	Genus	Taxonomic classification
	Species	Taxonomic classification
	Genus_species	Genus_species name
	Common name	Common name, if species has one
	Naming authority	Reference used as basis for scientific name for species
	Alternative name	None, one or several
	Maori name	Maori name, if species has one
	Abundance	Data on abundance, if provided
	Life-cycle stage	Information on life-cycle stage, if provided
	Observer name	Name of observer of specimen
Location	Location	Location where specimen observed
	Place name	Place name where specimen observed
	Eastings	Geographic information
	Northings	Geographic information
	Level	GIS descriptor/layer
	Region	Region where specimen observed
	Position	GIS descriptor
	Observation data text	Date, range of dates, or indication of when specimen observed
	Observation date	Date (yyyymmdd) when specimen observed
	Map sheet no.	Geographic information
	Latitude	Geographic information
	Longitude	Geographic information
	Imperial map sheet no.	Geographic information
	Imperial East	Geographic information
	Imperial North	Geographic information
	Host	Information on where animal found
	Position	Inside/outside terrestrial New Zealand
	Notes	Sundry information on animal recorded
Reference	Author	Author of observation data
	Type	Type of reference
	Date	Reference date of observation
	Publisher	Publisher or producer of document
	Title	Title of source of data
	Reference	Further reference information, if provided
	Source	Source of database, if provided
Voucher specimens	Repository	Name of repository, if held in collection
	Voucher number	Voucher number, if held in collection
	Number of vouchers	Number of vouchers, if held in collection
	Reference	Reference for specimen, if held in collection
	Date determined	Date specimen determined, if provided
	Voucher notes	Additional notes from repository, if provided
Rarity	Status	Threat status

TABLE 2. TOTAL NUMBERS OF INVERTEBRATES INCLUDED IN THE DATABASE, CATEGORISED BY PHYLUM AND CLASS.

PHYLUM	CLASS	TOTAL
Annelida	Hirundinea	1
	Oligochaeta	77
	Polychaeta	34
	(Unknown)	89
Arthropoda	Arachnida	674
	Branchiopoda	6
	Copepoda	37
	Diplostraca	1
	Malacostraca	141
	Ostracoda	3
	Chilopoda	15
	Diplopoda	11
	Insecta	15348
	(Unknown)	385
Chordata	(Unknown)	4
Ciliata	(Unknown)	2
Cnidaria	Hydrozoa	1
	(Unknown)	3
Echinodermata	Asteroidea	3
	Echinoidea	1
	Ophiuroidea	1
	(Unknown)	2
Gastrotricha	(Unknown)	1
Mollusca	Amphineura	5
	Bivalvia	28
	Gastropoda	412
	(Unknown)	128
Onychophora	(Unknown)	25
Platyhelminthes	Turbellaria	21
Porifera	(Unknown)	2
Rotifera	Monogononta	10
Sipunculida	(Unknown)	1
Kinorhyncha	(Unknown)	1
Nematoda	(Unknown)	46

TABLE 3. TOTAL NUMBERS OF INVERTEBRATES INCLUDED IN THE DATABASE BELONGING TO CLASS INSECTA, CATEGORISED BY FAMILY.

FAMILY	NUMBER	%
Coleoptera	5139	33
Lepidoptera	3706	24
Hemiptera	1658	11
Diptera	1219	8
Hymenoptera	1200	8
Trichoptera	795	5
Orthoptera	384	3
Ephemeroptera	341	2
Plecoptera	224	1
Homoptera	186	1
Thysanoptera	116	<1
Megaloptera	58	<1
Odonata	45	<1
Psocoptera	33	<1
Phasmatodea	28	<1
Mantodea	18	<1
Collembola	13	<1
Dermaptera	11	<1
Heteroptera	11	<1
Blattodea	5	<1
Neuroptera	4	<1
Protura	2	<1
Phthiraptera	1	<1
(Unknown)	151	1

4.2 LOCATION

Geographic Information Systems (GIS) provide a spatial context for viewing and managing environmental data. Since many of the invertebrate records that were obtained included some kind of spatial reference information, they can be incorporated into GIS to enable species distributions to be viewed and analysed in relation to other environmental data.

GIS was used to integrate the invertebrate data into a single common framework. The location information associated with the invertebrate data came in a variety of forms, including New Zealand Map Grid (NZMG) coordinates, latitude and longitude, NZMS 260 map sheet references, and decimal degrees. Further location data were interpreted from images and inferred from descriptions. For consistency and compatibility with most national spatial datasets, all records for which coordinate-based locational data existed were converted to the NZMG coordinate system.

4.2.1 Conversion

Data that were supplied as NZMG coordinates were retained in that format. Latitude and longitude references that were expressed as degrees, minutes and seconds were converted to decimal degrees using MS Excel prior to being imported into GIS. They were loaded into GIS as WGS84 coordinates and were then projected to NZMG, to generate coordinates for each record. Where data had both NZMG and latitude and longitude references, the NZMG reference was used preferentially. Where records had NZMG coordinates and an associated place name, other occurrences of that place name were assigned the same coordinates if these did not already exist.

4.2.2 Level

As each record was supplied with different spatial information, these were standardised using the hierarchical approach of NZMG: 'Place', 'Location', then 'Region'. Some data were supplied as latitude–longitude, which, although better than a place name, was less accurate than NZMG. Although these were converted to NZMG coordinates, they were rounded approximates (due to their relative inaccuracy). To ensure that NZMG and latitude–longitude data sources can be distinguished, the latter have been recorded as 'LL' in the 'Level' field of the database.

Some data that were supplied placed the NZMG coordinate outside terrestrial New Zealand (i.e. in the ocean). Although this was an obvious mistake, other discrepancies in the location records may not be as easy to distinguish. Therefore, the end user of the database needs to consider this before use. As an example, some spatial data were provided as latitude '41 20 S, 174 59 E'. This crude value has arisen because a latitude–longitude value has been interpreted from a drawn map. It is unclear whether this information was based on degrees minutes or decimal degrees (we assumed the former due

to no minutes values being greater than 60, which would be expected if these were decimal degrees). In either case, the information provided used a crude rounding that may push some beach locations outside terrestrial New Zealand when converted to NZMG for the database. Another discrepancy occurred when the location information provided used a beach name. In this case, a single GIS point was allocated to a middle point of the beach using a map.

The GIS database was constructed with an inbuilt hierarchical structure. Consequently, if there were any details missing in the original reports this could not be corrected for in the final database. Therefore, the original data were left uncorrected in the database. However, where the data point was clearly incorrect (e.g. located in the ocean), it has been indicated in the 'Position' column of the 'Location' section (either 'inside' or 'outside' terrestrial New Zealand, or 'no location'). There were 8068 suspect data records (46%) that fell into this category (outside) in the database, with another 1334 (4.6%) that had no location information.

Although it was intended that the database include invertebrates from only the lower North Island, with a boundary roughly south of Napier (Fig. 1), in some circumstances it was economic to obtain data parcels that included records from areas slightly north of this boundary as well (i.e. to avoid extra charges being incurred by the suppliers to sort and remove data from the requested dataset). Data that were located north of the Napier area have been left in the database, as some of these records may be of additional value once the DOC BioWeb database is more fully populated with various data.

TABLE 4. NUMBER OF RECORDS IN THE INVERTEBRATE DATABASE, BY REGION.

REGION	TOTAL
Wellington	5649
Taupo	3970
Rangitikei	2057
Taranaki	2046
Hawke's Bay	1346
Wairarapa	1153
Wanganui	801
Gisborne	396
Bay of Plenty	38
Waikato	28
Northland	3
Unknown[a]	26
No region[b]	6

[a] Unknown: Locality could not be found or was in confusion.

[b] No region: No information provided.

The database has been designed so that the 'Region' search field can be interrogated with selections or exclusions for different regions. Some GIS packages are able to define records below a critical northing location in a similar manner.

The majority of data records were located around major cities (e.g. Wellington and Taupo) (Fig. 1; Table 4). Records from Wellington accounted for 32% of all data. The regions defined in Fig. 1 were based on entomological area codes from Crosby et al. (1998). Figure 1 shows that many data points originated from above the Napier cut off. These have been included for completeness based on the entomological area codes.

4.2.3 Spatial accuracy

Original locational information fell along a hierarchical scale of accuracy. From finest to coarsest, this consisted of NZMG coordinates, latitude and longitude, place, and region. Even where a record has been assigned NZMG coordinates, the user should take note of the original scale at which the data was recorded, as some accuracy may have been lost in the conversion process. Entomological regions were digitised into GIS using an image from Crosby et al. (1998), so that those records with 'Region' as their highest spatial resolution could at least be assigned to some location. If a point had not been assigned a region, spatial queries were made in the GIS database to allocate a region.

4.2.4 Date accuracy

As the observation data came from a variety of sources, some of which were older documents, there was a variable level of accuracy in the dates recorded. More modern data records recorded the day and month, whereas older records sometimes only provided a season and year, just a year, or a range of dates when the invertebrate was observed. Where a full date was provided, this was recorded within the database as such (format yyyymmdd). The remaining observation dates were recorded as text (e.g. spring 1996 – winter 1997).

4.2.5 Observations with no locations

For some data that were forwarded for inclusion, no spatial information was included, or the location information was too broad (e.g. North Island). Where no location information was provided, a '.' appears under 'Location'. Where multiple species were recorded in the same study, they were typically allocated the same spatial coordinates.

4.3 REFERENCE

All reference information that was provided for the invertebrate data was collated and stored so that any future users of the database can track the data back to the original texts if necessary. The reference data were split in such a manner as to allow various publication reference formats to be easily used.

Over half the records (54.5%) were from relatively recent information (i.e. since 2000) (Table 5). Less than 2% of the data were forwarded with no information about when the information was collected or reported.

Information about the source of the data was less complete, with 45% of all the information being obtained from unknown references (Table 6). Typically, this means that the records were provided by Auckland Museum, the NZAC Bugs database at Landcare Research, or the Museum of New Zealand Te Papa Tongarewa (a small number). These can be identified by searching under the 'Source' field in the database). Auckland Museum alone provided 31% of all the information included in the database. Various books, including

TABLE 5. NUMBER OF RECORDS IN THE INVERTEBRATE DATABASE, BY DATE.

REFERENCE DATE	NO. RECORDS	%
Pre-1990	2824	16.1
1990–2000	4837	27.6
2000–2005	9541	54.5
No date	317	1.8
Total	17519	100.0

TABLE 6. NUMBER OF RECORDS IN THE INVERTEBRATE DATABASE, BY REFERENCE TYPE.

REFERENCE TYPE	NUMBER	%
Book	4938	28.2
Thesis	1951	11.1
Report	1725	9.8
DOC report	950	5.4
Letter	50	<1.0
Journal article	32	<1.0
Newspaper article	2	<1.0
Unknown	7871	44.9
Total	17519	

the 'Fauna of New Zealand' series, made up 28% of records, and various theses and honours dissertations (mostly from the University of Victoria, Wellington) contributed a further 11%. Newspaper articles, journal articles (including draft manuscripts) and various data held in letters and personal correspondence contributed a further 3%, while DOC reports (including *DOC Science Internal Series 146*, internal reports, and letters, reports and draft reports to DOC) contributed over 5%.

4.4 VOUCHER SPECIMENS

Where information was provided about voucher specimens, the type and format of information was retained in the original format used by the organisation where it was held. The rationale behind this was that any researcher can then take the original data back to the repository in a format that each repository can use.

4.5 RARITY

For the majority (96%) of the species listed in this database, there was no information about the rarity or threat status, as evaluated by DOC (e.g. Molloy et al. 2002) (Table 7). All information that was supplied about the current status of a species was included.

TABLE 7. NUMBER OF RECORDS IN THE INVERTEBRATE DATABASE, BY RELATIVE RARITY.
Threat classifications follow Molloy et al. (2002).

RARITY	TOTAL
Common	34
Data Deficient	98
Extinct	1
Gradual Decline	11
Highest priority/threatened	35
Introduced, paleoarctic origin	1
Nationally Critical	32
Nationally Endangered	54
Native	15
Not threatened	169
Range Restricted	79
Range Restricted: threatened	6
Rare	18
(Regarded as rare insect)	1
Serious Decline	10
Sparse	62
Threatened	2
Under biosecurity eradication	3
Widespread	2
Unknown	16886

5. Discussion

Although it was not intended that this database would be a complete account of all invertebrates in the lower North Island, it has been developed as a starting point for the DOC BioWeb database. The sections included make the database useful for future searches or compilations of records. It is envisaged that over time the database will be extended to include information from throughout New Zealand, and that information and taxonomic updates will be shared between holders of similar databases.

6. Acknowledgements

Many people and organisations have shared their data, knowledge and time to help to complete this database and associated report. Thanks are given to the following for discussions, information or comments on the draft manuscript: Trevor Crosby (New Zealand Arthropod Collection, Landcare Research), George Gibbs, John Early and Rosemary Gilbert (Auckland Museum), Astrid Dijkgraaf, Les Jones, Ian Cooksley and Ian Stringer (DOC), and Paul Dunford (Hawke's Bay Regional Council).

7. References

Cooper, R.A.; Millener, P.R. 1993: The New Zealand biota: historical background and new research. *Tree 8(12)*: 429–433.

Crosby, T.K.; Dugdale, J.S.; Watt, J.C. 1998: Area codes for recording specimen localities in the New Zealand subregion. *New Zealand Journal of Zoology 25*: 175–183.

DOC (Department of Conservation) 2001: BioWeb. Invertebrate user requirements. Internal report for the Department of Conservation (unpublished).

Molloy, J.; Bell, B.; Clout, M.; de Lange, P.; Gibbs, G.; Given, D.; Norton, D.; Smith, N.; Stephens, T. 2002: Classifying species according to threat of extinction. A system for New Zealand. *Threatened Species Occasional Publication 22*. Department of Conservation, Wellington. 26 p.

Monaghan, J.M. 1999: Effect of burning on bark invertebrates. P. 46 in Ponder, W.; Lunney, D. (Eds): The other 99%—the conservation and biodiversity of invertebrates. Transactions of the Royal Zoological Society of New South Wales, Mosman.

Watt, J.C. 1976: A biological survey of New Zealand? *New Zealand Entomologist 6(2)*: 138–143.