



Cook Islands Biodiversity Database

- a Case Study

8th Conference of the Parties to the CBD
Curitiba, Brazil, 20-31 March 2006

Gerald McCormack
Cook Islands Natural Heritage Trust

Introduction

The Cook Islands, a small island-state, has been developing a national biodiversity inventory to cover all species - plants and animals, terrestrial and marine - in a single multimedia database. This project is presented as a case study at COP 8 in relation to the CBD's proposed Island Biodiversity programme of work, and its Global Taxonomy Initiative (GTI).

The database, which is available on the web and on active CD, presently has information on 4,200 species, including 2,500 with photographs. We describe the main features of the database (Section 1), discuss some of the technical challenges and solutions (Section 2), and identify the present challenges (Section 3).

The database and website have been developed by the Cook Islands Natural Heritage Trust, an agency of the Cook Islands Government. We hope this case study will be useful to other small island-states developing national multimedia biodiversity databases.

The website and database are at:

<http://cookislands.bishopmuseum.org/>

Background

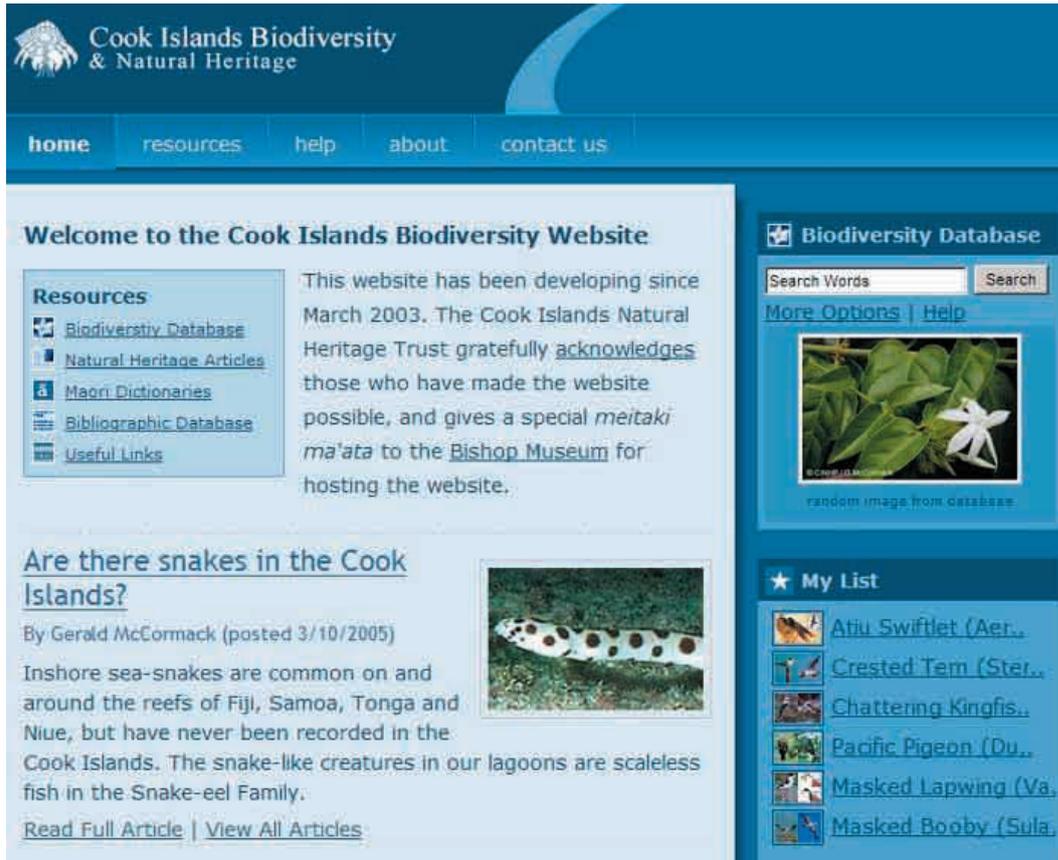
The publication of visual guides for biodiversity is a major hurdle to small island developing-states. In affluent countries the wide range of visual guides empower the general public through increased knowledge and awareness. Much loss of biodiversity is occurring in small island states, and the international community has given increasing support to programmes to manage and conserve particular species and habitats. In addition to direct conservation action, there is a need for more programmes to empower the public in small island states by increasing their awareness and knowledge of their biodiversity through suitable visual guides. In addition to empowering the public, our multimedia database partially fulfils our commitment to develop inventories of biodiversity, as required under Article 7 of the Convention on Biological Diversity.

The Cook Islands developed the Natural Heritage Project in 1990 with three goals: (1) to record local plants and animals with photographs and drawings; (2) to record related traditional and scientific knowledge; and, (3) to make such information available to the public to facilitate awareness and communication. Apart from a few posters and books on particular taxonomic groups or habitats, the long-term focus was to develop a single electronic system to record and present all Cook Islands plants and animals. This approach had three predicted advantages: (1) the public would need to master only one system of information retrieval and presentation; (2) multi-taxon groupings, such as all endemics or invasives of an island, would be in the same system; and, (3) popular groups, such as birds, fish, and ornamental flowers, would help to carry and open the door to less popular groups, such as insects, echinoderms, and ferns.

The Cook Islands is a very small country of about 20,000 people on fifteen widely separated islands in the tropical South Pacific. It has a total land area of 240km² and an oceanic EEZ of near 2,000,000km². The estimated plant and animal biodiversity is about 7,000 species, divided almost equally between marine and terrestrial. The present biodiversity database lists 4,200 species, including about 2,500 with photographs, and a similar number with descriptive texts. Some groups, such as flowering plants, ferns, birds, mammals, fishes, lizards, butterflies, and molluscs, have more than 90% of their predicted totals already in the database.

Since 2003 the database has been available on the Web, and on active CD for those without convenient access to the Web. In addition to the multimedia database, the website and CD make available: (1) natural heritage articles; (2) a bibliography, mainly related to local natural heritage; and, (3) two Māori dictionaries of words related to natural heritage. The biodiversity website and database is a work in progress, and it will become a primary source of knowledge on local plants and animals as information technology increases in schools and among the general public. See figure 1.

Figure 1: Website Home Page



The stable support of the Cook Islands Government for a long-term researcher has enabled steady progress for fifteen years. A breakdown of the researcher's time would be about five years for database design and data input, five years of photography and fieldwork, and five years for other activities. In addition, data input gained about three years from four short-term assistants (Fritha Stalker, Paul Kos, Gwen Welland, and Sonja Miller). Resident specialists in traditional knowledge have contributed the equivalent of two years, and overseas scientific experts have provided about three years of time, including a year of short visits by a series of scientists to undertake fieldwork sponsored under the NBSAP enabling activity.

Section 1: Features of the Database

Microsoft Access was selected because it is widely used, and easy to use. The main database is the editorial database in the office, and from it we output customised data-tables for the web database. The web database uses Active Server Pages (.asp) to access and present the information from the data-tables. Images are stored as image files in folders and they are called using data-table information when required.

The three basic components of the web database are:

- (A) the **Search Page** where search words are input, with or without other criteria from menus;
- (B) the **Results Page**, which displays a list of the species found, with or without images; and
- (C) the **Species Page**, which presents all information on a particular species, including all multimedia.

A: Search Page

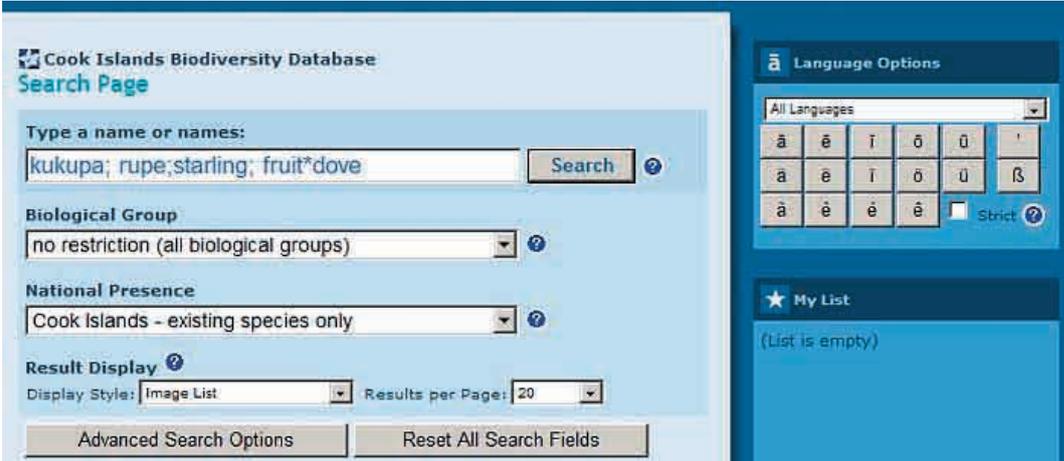
i) Input Textbox

The Search Page has a input textbox into which the user can type any Māori, English, or Latin name for a species, or a few names separated by semi-colons (;). The user can also search for one or more taxonomic groups by entering appropriate English or Latin names, such as Lepidoptera;Coleptera to find all butterflies and beetles.

Compound names pose an interesting challenge – is it spelt fruit bat, fruit-bat or fruitbat? The user could search for the three possibilities with fruit bat;fruit-bat;fruitbat. However, this is such a common situation that we have provided the asterisk (*) as a simple solution. The asterisk acts as a wildcard representing “a space”, “a dash” (-), or “nothing” so that fruit*dove will find fruit dove, fruit-dove and fruitdove. See Figure 2. The main reason for not finding an expected species is typing the endings of words, such as -s, -ium, and -ia. We have not programmed against such endings but they are best avoided.

The user can search for Māori names, with or without the diacriticals, for example Utū or ‘Utū or ‘Ūtū. The use of Māori in the database posed a number of challenges and these are discussed in the Challenges and Solutions section.

Figure 2: Search Page - basic



ii) Basic dropdown menus

The Search Page has two basic dropdown menus – Biological Group and National Presence. The Biological Group menu lists some major taxonomic groups of plants and animals, such as Birds, Fishes-Sharks-Rays, and Flowering Plants; as well as a few pseudo-taxonomic groups used by the public, such as Trees. These very broad taxonomic groups are useful to restrict a names search. For example, searching for butterfly will find various plants, insect butterflies, and butterflyfishes. The user can select <Insects> from the Biological Group menu to restrict the search to the insect butterflies. This menu can also be used to restrict searches, based on the Advanced Search Criteria, to a major biological group. For example, Cook Islands endemic birds can be found by combining an Advanced Search Criteria (<endemic>) with a Biological Group (<birds>).

The National Presence (formerly called Country) enables the user to include or exclude extinct species, Quarantine intercepted species, and Potential Invasives. The default is set to <Cook Islands - existing species only>.

The hidden section of the Search Page, the Advanced Search Criteria, offers a wide-range of search criteria not based on biological taxonomy, such as social usefulness, invasiveness, endemism, and habitat. This complex but most useful aspect of the database is discussed in a later section.

B: Results Page

The Results Page has three basic display modes: (1) <Image List> where each species has its thumbnail image and its main Latin, English and Māori name – the default; (2) <Images and all names> where each species has its thumbnail image and all its recorded Latin, English and Māori names; and (3) <Main names> which is a one-line-per-species display for printing with the main Latin, English and Māori names. See figure 3.

Figure 3: Results Page – default view



At present 50% of all species have an image, and for popular groups the level is usually more than 80%. As a result we have set the default Results Page display to <Image List> and there is an excellent chance that the user will quickly identify the species they are searching for from the thumbnail images. If the user wants to see their search word(s) highlighted for every species that was found, they can select <Images and All names> from the Display Options on the Results Page.

The Results Page <Image List> defaults to 20 to work efficiently with a 56k dialup connection, but this can easily be changed on the Search Page for those with faster connections or working from the CD. The user can click one or more thumbnail images to popup an enlarged image for more detail (see figure 8). In all display modes, clicking the scientific name opens the Species Page for that species.

On the Results Page we see one of the challenge of handling the different dialects of Māori. For each species a main or “national” Māori name was selected, and this is the default main Māori name. However, it is obvious that students on ‘Ātiu displaying the native trees of ‘Ātiu would want to display their own ‘Ātiu Māori names. They do this

on the Results Page by setting the Display Option Traditional Language to <'Ātiu>, and this changes the main Māori names to 'Ātiu Māori names followed in square brackets by the abundance of the species on 'Ātiu, if known. The students can also sort an <Image List> display, or a <Main Names> display based on their Māori names via the Display Options. See figure 6.

C: Species Page

The Species Page displays all the data and multimedia on a particular species. The most fundamental areas presently are the Latin, English, and Māori names; the list of multimedia; a simple description of the main features, and features important in distinguishing similar species; the abundance on each island; voucher information (evidence of presence); lists of biological and social significance, such as invasiveness, medical use, food use, etc.; and, general text on biological and social significance. See figure 4.

Figure 4: Species Page (top section only)

Cook Islands Biodiversity Database
Species Page

Back to: [Search Page](#) | [Results Page](#) ★ [Add to My List](#)

Ptilinopus rarotongensis
Cook Islands Fruit-Dove **Kūkupa**

Multimedia & Additional Resources

Type	Description	Download
Image:	Adult & juvenile - plumage differences	36KB
Image:	Adult - resting, breast & head feathers	40KB
Image:	Adult - Ātiu subspecies	32KB
Image:	Nestling and juvenile	57KB
Sound:	Call - Rarotonga (1) (M.Taylor)	345KB
Sound:	Call - Rarotonga (2) (M.Taylor)	676KB
Article:	Cook Islands Fruit-Dove (Kūkupa) Gerald McCormack, March 2005	Read

General Information

COMMON NAMES: Cook Islands Fruit-Dove; German Cook Insel Fruchttaube, Rarotongafruchttaube

TRADITIONAL NAMES: Kūkupa (RR AT MK)

GLOBAL DISTRIBUTION: NATIVE Cook Islands (Rarotonga, 'Ātiu)

COOK ISLANDS STATUS: Native, Resident Breeder, Endemic of Cooks; S,Group only (RR and AT - common); Land, makatea and central hills; Feeds on small fruits and berries.

SIGNIFICANCE LIST: ; Globally endangered (moderately) Ecotourism

KEY FEATURES: To 25cmTL. Plump with a short bill. Pale grey anteriorly, green upperside, and yellow underside; with a magenta cap and white tail-tip. Rarotonga birds have red patch on lower breast. VOICE: haunting "OOOOO-OOOO-ooo-oo" or "OOOO-OOOO-ooo-oo".

Cook Islands Distribution [View Distribution Map](#)

The Species Page shows the presence and abundance for each island in table form, and it also links to a distribution map presenting the same information over a map of the Cook Islands. See figure 9.

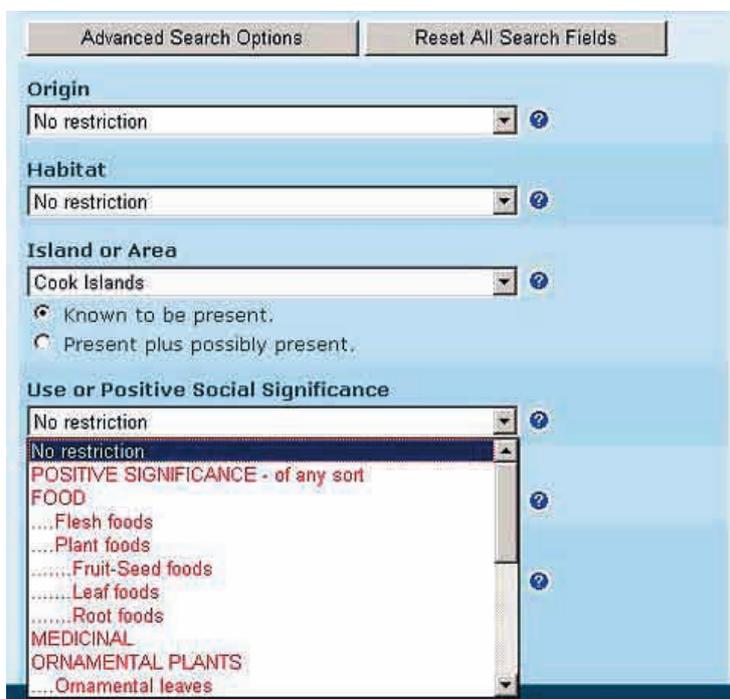
The user can add Species Pages to <My List>, which is maintained throughout the session and allows one-click reopening of the listed Species Pages from anywhere in the database. The user can also display <My List> as a Results Page in any of the three formats, for convenient viewing or printing.

Relationships between species, such as pests or pathogens are primarily presented on the Species Page. For example, the Species Page for *Homo sapiens* lists species related to people, such as those that cause disease, spread disease, are toxic, etc. Clicking on any species in these lists takes the user to the appropriate Species Page, or the user can display the whole list on a Results Page in any of the three Display Options. The inputting of this relationship data is at an early stage.

Advanced Search Options on the Search Page

The Advanced Search Options enable the user to find special-interest groups of species with or without restricting the search to a particular taxonomic group. See figure 5.

Figure 5: Search Page – Advanced Search Options



The screenshot shows the 'Advanced Search Options' section of a web application. At the top right is a 'Reset All Search Fields' button. Below it are four main categories, each with a dropdown menu and a help icon (question mark in a circle):

- Origin:** Dropdown menu set to 'No restriction'.
- Habitat:** Dropdown menu set to 'No restriction'.
- Island or Area:** Dropdown menu set to 'Cook Islands'. Below it are two radio buttons: 'Known to be present.' (selected) and 'Present plus possibly present.'.
- Use or Positive Social Significance:** Dropdown menu is open, showing a scrollable list of options: 'No restriction', 'POSITIVE SIGNIFICANCE - of any sort', 'FOOD', '.....Flesh foods', '.....Plant foods', '.....Fruit-Seed foods', '.....Leaf foods', '.....Root foods', 'MEDICINAL', 'ORNAMENTAL PLANTS', and '.....Ornamental leaves'.

The numerous search criteria are grouped into a series of dropdown menus, such as Origin, Habitat, Island or Area, Use or Positive Social Significance, Harmfulness or Negative Social Significance Positive Usefulness, and Endangered. The list of categories in each menu is relatively long and complex, because the criteria need to apply to a wide range of special-interest users, and to a wide range of taxonomic groups. For example, the idea of being Native or Indigenous is a straightforward concept, or is it? In our database we have taken Native to refer to any species that arrived by natural means, unaided by people, even though environmental changes brought about by people enabled the species to survive and reproduce. When a flowering plant is native it obviously breeds locally and it is not necessary to state this, however, for swimming and flying animals being Native is a more complex issue. For example, birds can be

breeding Natives or non-breeding Natives, and they can also be annual migrant breeding Natives or annual migrant non-breeding Natives, or they can be vagrant natives (which visit at irregular intervals). To deal with a wide range of taxonomic groups the Advanced Search Options menus are often very complex and take some time to master.

The Advanced Search Options take the database beyond searching for a particular species, or taxonomic group of species. The Advanced Search Options menus empower the user to find a wide-range of groupings based on biological aspects (such as invasiveness, habitat, or origin), and social aspects (such as use as fruits, toxicity, or use for adornment). Below we show two Display Options for part of the Results Page for Biological Group = <Birds>, Origin = <Native breeding locally>, and Island/Area = <‘Ātiu>. See figure 6.

Figure 6: Results Page – Image View and Name View for ‘Ātiu birds with ‘Ātiu Māori names

Cook Islands Biodiversity Database
Results

Displaying results 1 to 14 of 14 found [Back to: Search Page](#)

[Key to Symbols](#)
[Your Search Criteria](#)
 Biological Group: Birds
 Country: Cook Islands - existing species only
 Origin: Native - breeding locally
 Area: ‘Ātiu Display Option = Default



Anous stolidus
Brown Noddy
Ngōio [++]

Todiramphus tuta
Chattering Kingfisher
Ngōtare [+++]

Gygis alba
White Tern
Kākāia [++++]

Aerodramus sawtelli
Atiu Swiftlet
Kōpeka [++]

Egretta sacra
Pacific Reef-Heron
Kōtuku [+++]

Ptilinopus rarotongensis
Cook Islands Fruit-Dove
Kūkupa [++++]

below: Display Option = Names List with Scifamilies

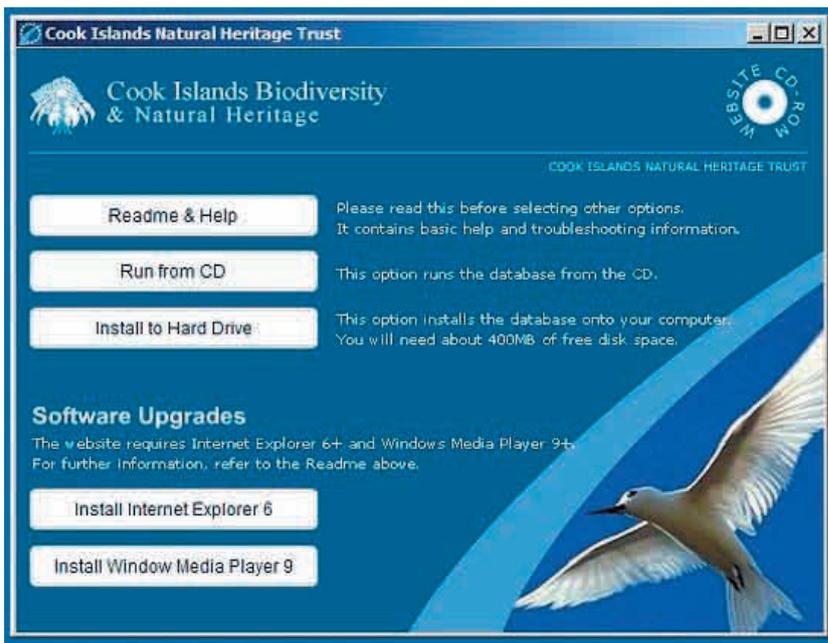
Family	Sci. Name	English Name	Atiu
STERNIDAE	Anous stolidus	Brown Noddy	Ngōio [++]
ALCEDINIDAE	Todiramphus tuta	Chattering Kingfisher	Ngōtare [+++]
STERNIDAE	Gygis alba	White Tern	Kākāia [++++]
APODIDAE	Aerodramus sawtelli	Atiu Swiftlet	Kōpeka [++]
ARDEIDAE	Egretta sacra	Pacific Reef-Heron	Kōtuku [+++]
COLUMBIDAE	Ptilinopus rarotongensis	Cook Islands Fruit-Dove	Kūkupa [++++]
ANATIDAE	Anas superciliosa	Grey Duck	Mokorā Taetaevao [+++]
RALLIDAE	Porzana tabuensis	Spotless Crake	Mo'o [++]
PHAETHONTIDAE	Phaethon lepturus	White-tailed Tropicbird	Fitake [++]
COLUMBIDAE	Ducula pacifica	Pacific Pigeon	Rupe [+++]

Increasing accessibility

Accessing our multimedia biodiversity database is dependent on students and the general public having access to computer technology. All schools now have a computer room for students, and computers are becoming more common with the general public. Internet access is also becoming more widespread and affordable, although the slow connections and high costs will continue to be a relative obstacle to browsing a complex multimedia database on the Internet. To meet this challenge we provide the website and database on a CD. The CD has a programme, Dynamic-CD, which presents the website actively from the CD, as though it was from the Internet. The CDs are provided free to local schools and Government agencies, and are available to the public at a nominal cost. The website database is updated every six months, and a new CD are issued each mid-year.

When inserted, the CD automatically opens a menu with a Readme, and options to open the website on the CD, or to install the website to the computer's hard drive for much faster access. See figure 7. Installing the website and database to the hard drive also makes a desktop icon for convenient access, and it provides program menu items to open the website on the hard drive or to uninstall the website from the hard drive. At present the entire website with its database and multimedia occupies about 60% of a CD, so it will be quite a while before we need move up to a DVD.

Figure 7: CD Options Menu



Section 2: Technical Challenges and Solutions

In designing this biodiversity database the focus was on easy and efficient access to the information for students, the general public, and special interest groups. In this case the designer had been the Cook Islands secondary school science advisor for six years, and director of the Conservation Service for two years – experiences that not only highlighted the need for photographs and traditional names, but also indicated the types of searches a teacher, student, or conservation worker might find useful. The designer worked with several programmers to develop the editorial database (principally Jack Stoffers), and with two programmers to develop the web database and website (Rich Pyle and Justine McCormack). Unfortunately, access to professional programmers has been very limited, and the total time these three were able to assist was about five months.

Below we discuss our solutions to some of the basic technical challenges that were encountered in developing the multimedia biodiversity database.

Fonts for Māori macron-vowels and the glottal

Microsoft Windows 3.1, released in 1992, had the new TrueType font system, which made characters similar and scalable for the screen and in print. TrueType fonts were based on the ANSI character set, which extended the earlier ASCII character set of 128 characters to 256 characters. The system provided the ASCII Arabic numbers and Roman alphabet for English, and added the common accented characters of other Western European languages, and various useful symbols. The macron-vowels (ĀāĒēĪīŌōŪū) were not included, nor a dedicated left-apostrophe (‘) for a glottal. The nearest characters to macron-vowels were umlaut or diaeresis-vowels (ÄäËëĬĭÖöÜü), and these were often used in manuscripts as markers and then typesetters turned them into macron-vowels for publication. The apostrophe was often used for a glottal although it had the disadvantage that some programmes converted them to “smart quotations” thereby turning glottals within words into right apostrophes, for example ‘ava’ava.

We decided to strive for real macrons and a dedicated glottal for use in the database and in our in-office typeset publications. In 1994 we used a font-editing programme to modify the dieresis-vowels into macron-vowels, and modified a little used superscript character, the Ordmasculine (°), into a dedicated and stable glottal. In this way we developed a Times Maori font and an Arial Maori font, and these solved our Māori orthography problems in all Windows-based programmes. However, in early 2003 when the database was ready for the web we discovered that users could not display our Māori names unless they download our Maori fonts – and, apart from the inconvenience, what would Microsoft have said about modifying one of their font!

Fortunately, although unbeknown to us, Adobe and Microsoft had provided the solution in 2000 with the release of the first OpenType fonts, based on the Unicode character set, which can have up to 65,500 characters – enough to cover characters for all languages in the world. The Unicode character set in the OpenType fonts, sometimes called UTF-8 after the system of encoding the characters, contained the macron-vowels (ĀāĒēĪīŌōŪū), and a dedicated single left apostrophe (chr#2018) for use as a glottal.

By 2003 many major programmes could still not handle the extended character set of OpenType, but fortunately for us Microsoft Office on Windows 95 or later could handle them and we were therefore able to access dedicated macron-vowels and a dedicated single left-apostrophe. With a one-time exhausting effort we changed all our Word and Access files from TrueType to OpenType with macrons and glottals where they should have been. Having achieved success with Unicode in Access we were a tad disappointed when the macron-vowels and glottals would not display on our website. After considerable messing around to discover that Windows uses the number 65001 to specify Unicode, our problems were solved with two pieces of code: (1) `session.codePage = 65001`; and (2) `<META HTTP-EQUIV="Content-Type" CONTENT="text/html; CHARSET=UTF-8">`. For more information on the macronisation of New Zealand Māori visit: http://www.tpk.govt.nz/using/macron_paper/index.html

Searching for Māori names

Making macrons and glottals available did not solve the challenges of searching for Māori names, because the spelling of Māori is inconsistent and in a state of flux. In the early 1800s the Missionaries selected the Rarotonga dialect for use in a national bible. They wrote Māori using the English alphabet, ignoring the fact that the vowels have a short and a long form, and mainly ignoring the glottal stop that indicates a very muted consonant. Native speakers became the masters of seeing a particular written word and pronouncing it differently according to meaning based on its context. For example, they would see *Utu* but say '*Utu* for Fish-poison Tree or '*Ūtū* for Mountain Banana based on context. Or, seeing the word *ua* and knowing by context whether to say *ua* (rain), '*ua* (egg), *u'a* (female) or '*ū'ā* (thigh).

When the reader knows how to speak a particular dialect of Māori the system of pronouncing words based on context works well. The problem arises when a reader is unfamiliar with the particular dialect. For example, very few Rarotongans can speak the Pukapukan language, and many young Cook Islanders growing up in New Zealand or Australia are not native speakers of Māori. In both cases, correct orthography is required to enable them to pronounce words correctly.

The Cook Islands Māori Dictionary of 1995 introduced modern orthography to show the pronunciation of each word within the spelling of the word – long vowels were marked with macrons (ĀāĒēĪīŌōŪū) and the glottal stop was shown as a single left apostrophe ('). Many people still resist the change – as being unnecessary. The biodiversity database records thousands of Māori names from different islands, and a linguist has checked most words to ensure the orthography represents the way the word is spoken on the particular island. However, we cannot insist on modern spelling so although a system is available on the Search Page to input all macron-vowels and glottals the search will, by default, ignore such characters, unless the user selects <Strict> from the Language Options. Without the selection of <Strict> the words *Utu* or '*Utu* or '*Ūtū* will find both the Fish-poison Tree and the Mountain Banana (plus various other species with names starting with *utu*).

Sorting Species Taxonomically

Computers sort alphanumerically – numbers ascending or descending, and the alphabet ascending or descending. They therefore sort the orders Reptilia, Aves, and Mammalia as Aves, Mammalia and Reptilia. This problem can be solved in a simple manner by developing an alphanumeric system to sort the taxonomic groups into a biologically acceptable sequence. If somebody had arranged the million or so taxonomic groups and given each a number we would have simply used that. And in some cases it was almost that easy because some taxonomic groups have had numbering systems developed at the Family level and when we found these, we used them.

In our case although we had a limited biodiversity we did not know what groups would need to be included, so we needed an alphanumeric system that could maintain sequences, yet accommodate new groups within the sequences. For example, Nelson (1994, Fishes of the World, 3rd Ed.) developed a numeric system for all the known families of fishes, and within this system it is easy to move a family by appending a letter, for example 073, 074, 075 can have a family added to become 073, 074a, 074b, 075.

We developed an alphanumeric system for higher taxonomic groups separately from the numbers for families. For example, the 28-spot Ladybird (*Epilachna vigintioctopunctata*) is sorted by the alphanumeric [a^5B^p29D_A\$c12B_A%o27a_a&01a] plus [084], while the user sees the higher taxonomy as [k.ANIMALIA, p.ARTHROPODA, sbp.ATELOCERATA, spc.HEXAPODA, c.INSECTA, sbc.PTERYGOTA, o.Coleoptera] plus [COCCINELLIDAE]. See figure 8. We could have used numbers only, but we found that dividing the sequence into sections for phylum, class, and order made it easier to adjust numbers as taxa were moved or new taxa added. The main advantage of the system is that we can add to it piecemeal as new taxonomic groups are discovered without upsetting its integrity. During ten years of use, the system has proved to be robust, efficient and adaptable.

Figure 8: Alphanumeric System for sorting taxonomically

HighTax_No	Fam_No	SciFamily	SciName
a^5B^p29D_A\$c12B_A%o27a_a&01a	063	NITIDULIDAE	<i>Stelidota nigovaria</i>
a^5B^p29D_A\$c12B_A%o27a_a&01a	067	SILVANIDAE	<i>Cryptamorpha desjardinsi</i>
a^5B^p29D_A\$c12B_A%o27a_a&01a	067	SILVANIDAE	<i>Oryzaephilus surinamensis</i>
a^5B^p29D_A\$c12B_A%o27a_a&01a	075	CRYPTOPHAGIDAE	<i>Cryptophagus</i> QQAS-A
a^5B^p29D_A\$c12B_A%o27a_a&01a	083	ENDOMYCHIDAE	<i>Holoparamecus?</i> QQAS-A
a^5B^p29D_A\$c12B_A%o27a_a&01a	083	ENDOMYCHIDAE	<i>Trochoideus desjardinsi</i>
a^5B^p29D_A\$c12B_A%o27a_a&01a	084	COCCINELLIDAE	<i>Chilocorus bipustulatus</i>
a^5B^p29D_A\$c12B_A%o27a_a&01a	084	COCCINELLIDAE	<i>Chilocorus circumdatus</i>
a^5B^p29D_A\$c12B_A%o27a_a&01a	084	COCCINELLIDAE	<i>Chilocorus schioedtei</i>
a^5B^p29D_A\$c12B_A%o27a_a&01a	084	COCCINELLIDAE	<i>Coccinella repanda</i>
a^5B^p29D_A\$c12B_A%o27a_a&01a	084	COCCINELLIDAE	<i>Coelophora inaequalis</i>
a^5B^p29D_A\$c12B_A%o27a_a&01a	084	COCCINELLIDAE	<i>Cryptolaemus montrouzieri</i>
a^5B^p29D_A\$c12B_A%o27a_a&01a	084	COCCINELLIDAE	<i>Epilachna vigintioctopunctata</i>
a^5B^p29D_A\$c12B_A%o27a_a&01a	084	COCCINELLIDAE	<i>Harmonia octomaculata</i>
a^5B^p29D_A\$c12B_A%o27a_a&01a	084	COCCINELLIDAE	<i>Olla v-nigrum</i>
a^5B^p29D_A\$c12B_A%o27a_a&01a	084	COCCINELLIDAE	<i>Rhyzobius lophanthae</i> QQID

Handling uncertain identifications

Imagine you have just found a fascinating insect in your garden and have taken great photographs, but you cannot add them to your database because you don't know its scientific name, or you think you know but are not 100% certain. Over the years, we have gained the support of a wide range of overseas taxonomists who can often identify species from detailed photographs, or who will provide identifications for specimens sent to them. However, even with the best taxonomic support in the world, there are going to be many species in the database where the Latin name being used for a species is questionable. Furthermore, there can be a level of questionability at any taxonomic level, although it is more common at the level of family, genus or species.

Marking questionable names with question marks, such as ?LIMACIDAE, ?*Lamprocystis* or *Vaginulus* ?*alte*, is a very obvious system. However, searching for question marks is technically difficult in Access and also with file managers. We therefore decided to indicate questionable names in the database and in filenames with an unused pair of letters, namely "QQ", which are easy to search for.

In the database we indicate questionable identifications directly in the scientific name, for example Limacidae QQ, *Lamprocystis* QQ or *Vaginulus alte* QQ. Usually the QQ has a reminder after it referring to a person, a specimen number, a collection location, or a feature. For example, *Vaginulus alte* QRR1 means that this species is based on specimen number one from Rarotonga, and this information is noted in a separate field in the database.

Questionable identifications of specimens in images are shown in the filename as follows: _limacidQQ, _LampQQ, or Vagi-alteQQ. In this case the doubt can have two different meanings: (1) the image is of a species with a questionable name in the database; or (2) the species is correctly recorded in the database but the specimen in the image is not guaranteed to be of that species.

The adoption of this system enabled us to include species with incomplete or questionable identifications in such a way that it would be clear to the user that there was some doubt concerning the name being used. It also enabled us to make full use of the available photographs while ensuring that the user was aware that the specimen in the particular image might not be of the species as named in the database.

Editing image files

Editing thousands of images is one of the major undertakings to develop a multimedia database. Of course, you don't have to edit an image at all to put it on the web, and then have a programme display it at different sizes depending on the user's connection speed. However, on looking over our photographs we concluded that both good and poor images could be greatly improved by editing – usually, cropping, adjusting levels, and a little sharpening (with unsharpen mask!). Initially we cropped according to the shape of the object in the image, but later we decided to standardise the dimensions of our Primary Image – the image used as a Thumbnail and as the associated Main Image. Standardization of the Primary Image had four advantages: (1) it focused the editing decisions by creating a fixed picture frame; (2) it made it easier to present an array of

comparable images on the Results Page; (3) it made printing an array of thumbnails, or a series of main images, much easier; and (4) it facilitated the convenient printing of small field guides with colour images.

The Primary Image is edited to make a Main Image of 600x400pixels, which is viewed in a popup viewer, and a Thumbnail of 150x100pixels, which is used on the Results Page. See figure 9. Most of the image editorial effort is focused on producing a Main Image that visually summarizes the main features of a species. To achieve this in one image it is often necessary to use inserts to make a collage showing different morphs or particular details. For example, the frigatebird is sexually dimorphic, so it is ideal to show both sexes in the Primary Image. In support of the Primary Image there can be numerous Supporting Images at any convenient size to show aspects of interest. In our system the Main Image is accessible wherever the Thumbnail Image is displayed, while the Supporting Images are accessible on the Species Page, along with any video files and sound files – the more the better.

Figure 9: Primary Images – Thumbnails and one pop-up Main Image



Our landscape 600x400 pixel Main Images are typically about 65KB, which means they download conveniently on 56K modems, which are common in rural areas. We have found that images at 600x400 pixels display well on screens from 800x600 through 1024x768 to 1280x1024; and the numbers are easy to handle to arrange inserts for collages, such as 400x400 plus two 200x200, or 300x400 plus another 300x400.

Copyright of images can be a problem on the web. One way to prevent the user making unauthorised use of images is to make them so small or of such poor quality that nobody would want to use them, and another popular system is to write the photographer's name in bold letters so it dominates the image. We could not use the poor quality option because we want every resident of the Cook Islands to have access to good quality images. Initially we decided in favour of the obtrusive photographer's name. However, we eventually refocused on the primary object of the database: to provide quality visual identification material to all Cook Islanders. With the refocus on our goals we changed to put the photographer's name muted but readable at the bottom of the image. See figure 8 (popup image). We also have a simple Use Policy statement on the website which stresses that we encourage private, non-commercial use of our images.

Managing image files

As our few thousand images became many thousand, finding or developing a system to effectively organise and find images became a matter of extreme urgency. There were a few image cataloguing programmes that were effective, but for a remote location we decided to work within the Windows file management system. This led us to focus on putting essential information into the filename rather than into metadata.

In our filenames we record the Latin name of the species, the photograph location, the photographer, and more recently we have added the date. With this basic data we could use ordinary search programmes to find all the images for a species for comparison and review. However we soon realised that it would be ideal if each filename were prefixed by an alphanumeric code to sort the images into the same taxonomic order as in the database. We therefore developed an alphanumeric summary of the database sorting system for use in filenames, ensuring that it allowed for the insertion of images for new taxonomic groups and the moving of images from one taxonomic group to another without loss of integrity. If we were starting out again now, we might have decided to put all this filename information into metadata and use an image-cataloguing programme to find and display the images. However our system works well without any special programmes, and it makes it very easy to view species that are related to each other with any image viewer, such as the freeware XnView or IrfanView.

Like the taxonomic system in the database, a taxonomic prefix must be stable even when new taxonomic groups are added, and our system meets this criterion. For example, an image of the 28-spotted Ladybird might be: 5AUt084_Epil-vigi_RR3_GM1_2005-12_ZC.jpg. The components are: (1) 5AUt084 = alphanumeric taxonomic data to control sort order; (2) _Epil-vigi = 4+4 representation of genus and species; (3) _RR3 = Rarotonga, third series of photos; (4) _GM1 = photographer, first image in the series; (5) _2005-12 = year and month; (6) _ZC = symbols meaning original camera image. When we edit an image we indicate the type of editing by changing the _ZC to (a) _WK for the working Photoshop file, (b) _MX for the Primary Main Image, and (c) _TX for the Primary Thumbnail Image. This system enables us to conveniently recognise how a particular image has been edited and to find all images edited in a particular way.

The main disadvantage to our system of building the taxonomy and Latin name into the filename is that it makes updating this information more cumbersome. With a dedicated image management programme updating names associated with images would be done a single time and would apply to all files with appropriate metadata. In our case, we search across all folders that might contain the image with a file-finding programme and adjust the names with a batch-renaming system. This is not as time consuming as it might seem; for example, XnView can search across more than 100,000 files in dozens of folders in five seconds - although the first such search takes nearly 20 seconds! The batch renaming system in XnView is excellent.

Getting Google to Index a Database

The databases used by search engines such as Google are made by robot programmes, which are called spiders because they crawl the links that make up the web. Spiders prefer static html pages, and they crawl along the simple links from one to another. Spiders have neither hands nor brains to select options on an .asp page to crawl into a behind-the-scenes database. They are also programmed to avoid complex connecting paths, especially those that contain questions marks (?).

We experimented with various systems to assist the spiders and they now index our species data, so that searching for a species name with Google can find our web database. The system involved making a `googlelist.asp`, which opens to present twenty species each with a thumbnail image and all their Latin, English, and Māori names; and each species section had a short link to generate its full Species Page. When the spider has crawled down this page, having taken the short detours to each Species Page, then at the bottom it follows a short link to generate the next page with twenty species.

Our experience with this system showed that the Google spider, over a period of several visits, generated and indexed all our Species Pages. We found that if we searched for a species name with Google it would either take us to the appropriate page of the `googlelist` or, much better, take us directly to the Species Page.

Our first `googlelist.asp` had the species sorted in the order that they were added to the database, which meant that when a user was taken to the `googlelist` page it was taxonomic nonsense. We amended the system so that `googlelist` pages were sorted taxonomically and thought we had solved all our problems. Unfortunately the spiders definitely are lacking in the brain compartment, because they were incredibly slow to notice that the already indexed `googlelist` page 324 was not the same as the present `googlelist` page 324, thus users were being directed to `googlelist` pages that no longer listed the species they searched for – what a mess! We are still learning how to teach a spider to index our database intelligently!

Section 3: Present Challenges

The future task is mainly finding and photographing the 2,000 species already listed in the database but without an image; then we need to find, identify, and photograph the estimated 3,000 species that are in the Cook Islands but not yet listed in the database. Below are a few of the other developments we are planning.

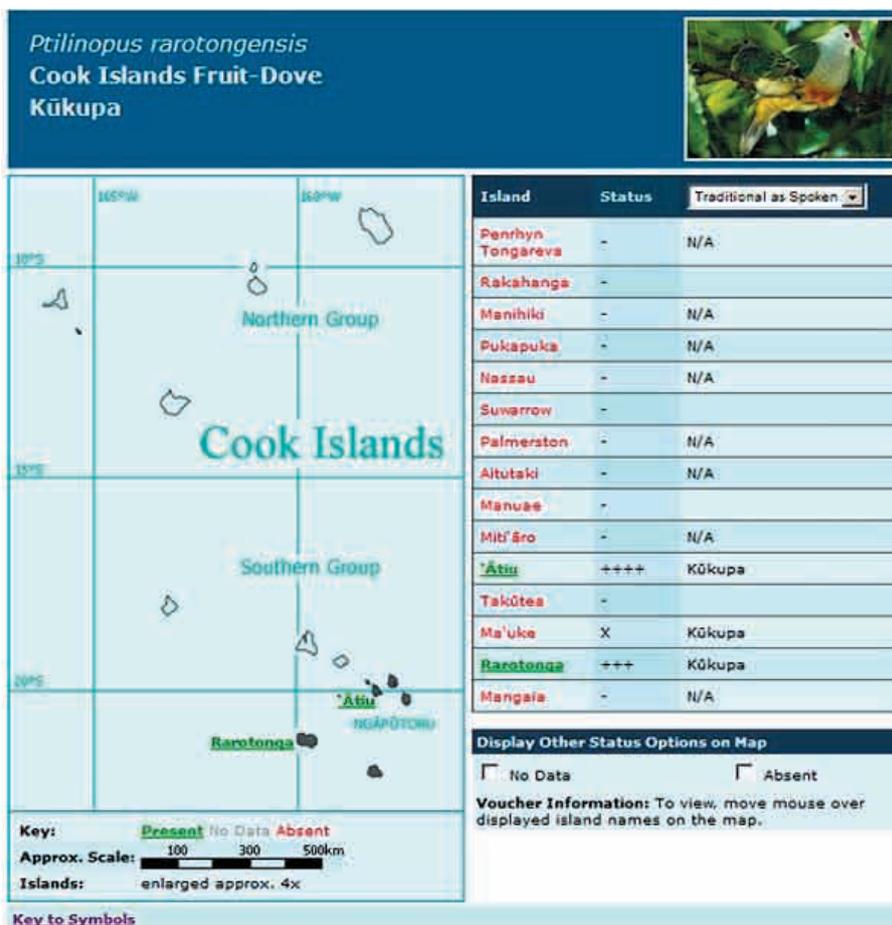
(1) Finding species by features

At present a user must know a Latin, English, or Māori name to find a species or taxonomic group to search the database. We need to develop a system to enable a user to find a species by referring to its features. We have experimented with a few options and have found that the Lucid taxonomic key system would work well within our web database. The main challenge to implementing this system is finding suitable manpower

(2) Geographic Information System (GIS)

At present we have a simple system that develops a Cook Islands distribution map for each species based on its presence on each island, and the system also displays the “voucher information”, which indicates the evidence we have for claiming that the species is on that island. See figure 10. We would like to incorporate a more elaborate GIS system, possibly based on the freeware DIVA-GIS. The main challenge in implementing this system is finding suitable manpower.

Figure 10: Cook Islands Distribution Map



(3) Website

The website started with the biodiversity database, and we have been adding new sections to make the website a more useful local resource. We now have sections presenting: (1) natural heritage articles; (2) Māori dictionaries of natural heritage words; and (3) a bibliography of natural heritage publications. The next major section needed is one to present each of the islands of the Cook Islands with a special emphasis on their natural heritage and biodiversity. The main challenge is to find the manpower to implement this section.

Conclusion

We hope you can benefit from our experience. And, if you do develop a multimedia biodiversity database, we hope we can cooperate with you to mutually lighten the burden of obtaining and editing the thousands of photographs required to make it work.

<http://cookislands.bishopmuseum.org/>