SCIENCE AND RESEARCH
INTERNAL REPORT NO. 88

REPORT ON THE DEPARTMENT OF CONSERVATION'S NATIONAL DATABASE WORKSHOP

by

Helen M. Adcock and W. Mary McEwen

Head Office,
Department of Conservation,
P.O. Box 10-420,
Wellington,
New Zealand

October 1990

ISSN 0-114-2798

ISBN 0-478-01221-7

Keywords: Workshop, Database, Computer systems, Standards, Structures

CONTENTS

LIST OF PARTICIPANTS

NAME	FROM
John Adams	Hawkes Bay
Helen Adcock	Science & Research
David Bowie	Library
Phil Brady	Wellington
Bill Chisholm	Otago
Terry Connor	Estate Protection Policy
David Crawford	Resource Use & Recreation
Duncan Cunningham	Science & Research
Janis Freegard	Protected Species Policy
Mairie Fromont	Hawkes Bay
John Galilee	East Coast
Ken Hales	Historic Places Trust
Jane Hare	Southland
Bob Harper	Historic Places Trust
Malcolm Harrison	Science & Research
Liz Humphreys	Waikato
Graeme Jane	Nelson
Paul Jansen	Bay of Plenty
Les Jones	Information Systems Development
Mary McEwen	Science & Research
Nick Mein	Stewart Island
Nero Panapa	Bay of Plenty
Geoff Patterson	Information Systems Development
Ross Pickard	Science & Research
Brian Rance	Southland
Richard Sadleir	Science & Research
Brian Sheppard	Science & Research
Dick Veitch	Threatened Species Unit

Visiting Speakers:

Alan Edmonds	Deputy Director-General
Garth Harmsworth	DSIR (Division of Land Resources)
Diana Kelly	DSIR Library Centre
Ian Payton	Forest Research Institute
Neil Puller	Land Information New Zealand (LINZ)
Jody Richardson	MAF Fisheries

REPORT ON THE DEPARTMENT OF CONSERVATION'S NATIONAL DATABASE WORKSHOP

by

Helen M. Adcock and W. Mary McEwen

Science and Research Division, Department of Conservation,
P.O. Box 10-420, Wellington.

SUMMARY

This is a report of the proceedings of the Department of conservation's National Database Workshop, held at Riverslea Lodge, Otaki Gorge Road from 20th March to 23rd March, 1990. It includes resolutions made at the workshop, notes made during talks given and any subsequent discussion, and copies of papers presented by speakers at the workshop.

1. INTRODUCTION

This report was compiled from notes taken during the workshop and includes copies of written papers prepared by some of the speakers. An additional paper by G.O. Eyles is also included. This had been prepared for the DSIR workshop on the use of remote sensing which was run in February 1990 for Department of Conservation staff.

A series of important resolutions resulted from this workshop:

2. RESOLUTIONS FROM THE DATABASE WORKSHOP

Participants at the National Database Workshop expressed serious concern that the development and operation of the DOCnet information system is being driven by non-departmental personnel with a limited field of expertise. The procedures needed to develop a strategic information plan have not been fully and properly defined, nor has the advice of departmental staff with appropriate expertise been sought. As a result we believe that the Executive Management Team (EMT) is receiving incomplete advice.

2.1 Information Policy

RESOLUTION

The Department develop an information policy which ensures that DOC:

accepts that information is a major DOC resource and that the use of information is a major activity essential to achieving DOC's mission.

implements information systems in a planned, logical sequence by following a Strategic Information Plan.

regards training as an integral part of system installation and maintenance.

2.2 Strategic Information Plan

RESOLUTION

EMT commit the Department to the development of a strategic plan, and that this plan be developed before any major system change or upgrade is considered.

2.3 The RFP for DOCnet Stage II

It is not the role of vendors to establish the department's business or strategy. The department must do this itself.

The Request For Proposals (RFP) has major flaws, contains misleading and inaccurate statements, and is based on inadequate policy.

Before entering into large expenditure on hardware and software, the Department must develop a policy on information and a strategic information plan to manage the implementation of that policy.

RECOMMENDATION

It is our strong recommendation that DOCnet Stage II (excluding servicing of new Conservancies) be immediately stopped and that the RFP for DOCnet Stage II be rewritten once the above issues have been satisfactorily addressed.

Money earmarked for DOCnet Stage II be redirected to policy development, strategic information planning, staff consultation and staff training.

2.4 Training

There has been insufficient computer training at a basic level for DOC staff. Large sums of money have been, and will be, invested in computer and communications equipment. The level of knowledge and experience of most staff in using such equipment is less than satisfactory. In order to the benefits from this investment, training is mandatory.

RECOMMENDATION

> Training should commence immediately, according to the list forwarded from this workshop to the Training Officer.

2.5 Communication

During the past 18 months there has been a serious lack of contact between Head Office staff and Conservancy end-users of the current DOCnet functions and there has been little or no contact with end-users when considering future developments of DOCnet.

RECOMMENDATION

> A system be established to ensure wide consultation with staff and keep them fully informed about computer use and development.

2.6 Database Development

There has been a considerable duplication of effort in establishing databases. Without the necessary central co-ordination and defined standards the databases that have been developed are generally incompatible.

RECOMMENDATIONS

> Information Systems Development (ISD) should be responsible for the development and co-ordination of all nationally applicable computerised databases.
>
> Standards for all databases be established.

2.7 Staff Levels

The lack of an information systems manager and other information systems staff has had a severe negative impact on the development and implementation of an information systems policy and strategy.

RECOMMENDATIONS

> The appointment of the information systems manager be accorded very high priority.
>
> The vacancy for a user requirements analyst within Information Systems Development be advertised and filled as soon as possible.

3. NOTES TAKEN AT THE WORKSHOP (in chronological order)

Tuesday 20 March

The Workshop began with a welcome from Ross Pickard, who chaired the workshop committee. The official programme for the database workshop is attached (Appendix I).

3.1 Ross Pickard - Existing Databases

Ross listed some of the databases reported as being used by DOC. He indicated that a variety of databases were being developed on PCs in various parts of DOC with little attempt to standardise, the main limitation to such database development being the availability of space on PCs.

Examples include:

SUBJECT	LOCATION
<u>Finance</u>	
DOCfin	All
Library Payments/Receipts	Library
Financial monitoring	Wanganui
Budgeting	Masterton
Hut pass returns	Nelson, Masterton, East Coast
Concessions	Dunedin
Leases and licences	Nelson, Waikato, Hokitika
Contracts commitments	Science & Research
<u>Permits</u>	
Kakariki permit holders	Christchurch
Permit issuing	Head Office
Permits	Masterton
CITES	Head Office
<u>Species Lists</u>	
Plants	Nelson, Waikato, Rotorua
Blue duck distribution	S & R, Nelson, Waikato
Amphibian and reptile dist.	S & R
Bat records	Waiakto
Whale strandings	Head Office
Native fish	Waikato, Rotorua
Kokako distribution	Waikato

Hunting

Deer hunting	Waikato
Helicopter hunting returns	Waikato
Goat kill records	Waikato

Ross asked the question - Should all these databases follow standards, and does it really matter?

Discussion

Which of the PC based databases should be on the Virtual System (VS)?
Wouldn't networked PCs work just as well?

Graeme Jane - VS does not have enough space.

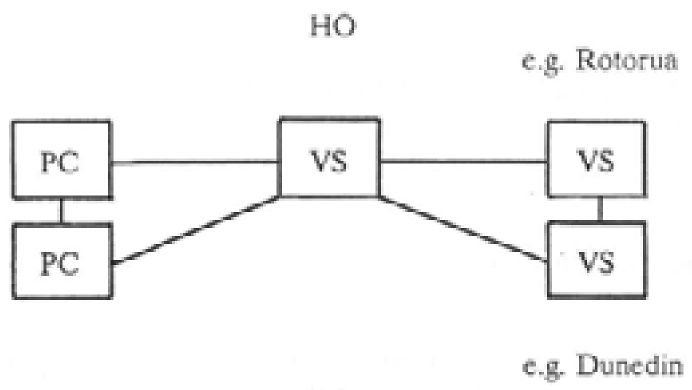
Paul Jansen - DOC data should be fully available to anyone; information is highly useful to all sorts of people; he thinks it is a problem that there is no guidance from Head Office.

Dave Crawford - Management of databases is separate issue from ownership. Must have someone or one group who own and are responsible for updating and managing the database.

Paul Jansen - Best if Conservancies update and manage own parts of the national database

3.2 Geoff Patterson - Types of Computer System

Geoff illustrated the type of system we have now



He explained that if offices can't afford to buy dBASE IV, Information Systems Development can compile databases for them using a compiler.

Most databases in the Department of Conservation are on PCs using dBASE, but a few are on the VS.

The database program on the VS is SPEED:

- it does its own backups (i.e. when the system is backed up),

- it has better security than databases on PCs.

The Land Register is on the VS. It is not a very large program, but disk space on the VS has been a limitation.

The Department of Conservation is thinking of hiring more disk space in the development of DOCnet.

One idea is to be able to network PCs within a Conservancy - and linking to the Conservancy VS - including links with field station PCs.

They are trying to find a system which will enable links of the VSs (e.g. so Head Office could compile summary information from databases held on Conservancy VSs).

<u>Discussion</u>

The need for more training was highlighted, e.g. about VS housekeeping.

The New Zealand Forest Service had very strict protocols – e.g., to control programs on PCs. Should the Department of Conservation also have strict protocols?

It was suggested that operations staff in Conservancies should be able to keep an eye on housekeeping on PCs.

Paul Jansen - There should be standard programs prepared by Head Office for such things as permits.

Dick Veitch - There are two sides:

(1) The people writing the programs.

(2) The people using them.

There are relatively few people in DOC who are designing databases -they should be prepared to visit and talk to Conservancy staff to find out their need.

Dave Crawford - Resource Use & Recreation are developing a standardised Recreation and Resource Use program -involving conservancies in the development.

Graeme Jane - What is very important is communication so that people do know what is available.

3.3 Les Jones and Malcolm Harrison - Establishing Computer Database Structures.

Les Jones spoke first:

Question: What is a database?

Answer: Must contain data or information
Set of information with defined relationships used for one or more purposes
Information that you wish to retrieve (retrievable)
Amendable
Maintainable
Structured
Doesn't have to be computerised, e.g. even a dictionary or an encyclopaedia is a database

Question: How do we determine when a database is required?

Answer: When we want an output
report
record
access
to retrieve information in whatever form you want, to be able to sort it, to facilitate transfer of information as a means of communication.

i.e. when a requirement exists

Question: How do we ensure that the database will meet future needs?

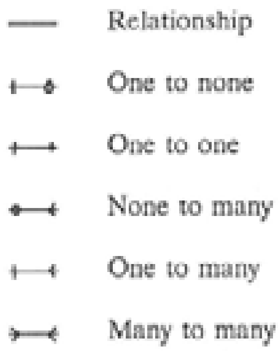
How do we ensure that the database is part of an integrated DOC system?

Answer: Information analysis - data modelling

The modelling has behind it a data dictionary:

- file
- objects
- entity
- activities/transactions
- organisation

The symbols of data modelling:



e.g. a library

People  Book (i.e. many people to many books)

A data model can also be called an information map.

Malcolm Harrison continued the subject:

The Department must aim to integrate our information - in the Conservation Act there is the requirement to advocate.

Ideally, we need a database that can be "everything for everybody".

We have a range of users from generalists to specialists.

The link for most DOC staff is the location or the "SITES" where what they are interested in take place or exist/occur.

Malcolm referred to the SITE database which Geoff Patterson has developed from a data model which Malcolm prepared.

(Note: This database structure is available on disk from Information Systems Development.)

3.4 Terry Connor - Structuring data: ways of data collection

Terry sees data like a tree rooted in the earth. He described some of his own philosophy and emphasised that at this stage in the Department's development we need to concentrate on the bottom section of the tree - user needs etc.

It is essential to know what information is needed and in what form. Once you know specific need, suitable systems can be developed to meet that need.

As an example he described the development and setting up of the Land Register.

The present state of the Land Register:

Canterbury has finished entering its data in the register.

Auckland was a pilot location but not yet complete.

About three other conservancies have begun to enter data in the register.

Wednesday 21 March

Before beginning Wednesday's programme, the workshop decided to list a number of issues that we felt it would be valuable to reach agreed resolutions about by the end of the workshop.

These included:

(1) Training
(2) Communication
(3) Staff levels
(4) DOC computer policy
Sale of Data
Charging for Data extraction
(5) Strategic information plan
(6) The RFP
(7) Database standards

People were assigned to develop draft resolutions for each of these issues.

3.5 Les Jones/Malcolm Harrison - Ways of identifying those databases that need to be computerised.

Les Jones

<u>The workshop brainstormed:</u>

1. On the Good Points of what we have got already

"Office" communications (electronic mail, etc.)
VS Network -available in various locations
High level of motivation?
Talents
Word processing
Hardware mix (PCs and VSs) - some flexibility
Office servicing
Ability to transfer data
Facility for use of distribution lists, e.g. for newsletters
dBASE - comparatively easy to use
Expandability
Printer quality
Service record

2. On the Current Problems

Response times
Difficulty in data transfer (PC and VS)
Under-utilisation (communications)
Training -lack of appropriate/usable documentation
Wang Office printing
Support - for the VS network
No national direction -lack of standards
PC support
People doing their own thing -results in duplication of effort.
Waste of resource (amateurs doing professional's work e.g. typing -and wasting time which should be used for field work).
Lack of rules
Lack of communication
Lack of trained professional operators
A lot of stand alone PCs - it could be an advantage to network PCs
Lack of disk space
Lack of reliability
Lack of terminals
Demand exceeds supply
Lack of strategy
Bad planning
Little progress
Lack of consistency - finance
Lack of external audit

Malcolm Harrison

Malcolm began by commenting that yesterday someone had asked "Why has nothing happened in the last couple of years?"

He then set out to explain the reasons for apparent lack of progress. In spite of appearances, some good progress has in fact been made.

The Director General has a contract with Government to carry out/perform the key objectives, and we now have a Corporate Plan and a series of Business Plans which identify a series of key outputs. Key outputs equal the way of putting into practice the key objectives.

The main business of the Department of Conservation is information. In the great majority of cases the information is related to subjects and places. In this respect DOC is fundamentally different from most businesses, which usually deal in products or services.

The reason nothing much seems to have happened in the last few years is because it has taken till now to get the present state of having about 75 key outputs, each of which is the responsibility of one manager.

Now it is time to go through the department interviewing all these managers to find out what information they need to meet each key output.

At the same time these managers may need to be educated about the advances in technology. So now we can make a list of the information requirements to meet each key objective.

DOC's Conservation Contract:

1 Act	)	
1 Mission	)	— Stable
3 Goals	)	
10 Objectives	)	
75 Key outputs		— Change from year to year

We need a system flexible enough to deal with this. Next we need Priorities -To set priorities:

Job quality
$$
People
Skills
Resources

Knowledge
Hardware
Software
Urgency and complexity

With limited amounts of money the department may have to reduce job quantity in order to maintain quality.

Computers may be used - depending on the priority.

High Priority Candidates for Computing:

Operations which are long term
Systems which won't change
Nationally important projects

Operations which are short-term would have lower priority.

The Public Service should be quietly efficient but we must live up to public expectation.

The planning is an enormous task, but Malcolm believes it must be done.

It may also seem to be very expensive.

Software is available for analysing the information requirements of a corporation (DOC is equivalent to a corporation).

e.g. the software costs $39,000 to run on a PC (compared with $16K for PC-GIS for software).

Plus need to interview managers of 75 key outputs.

But - Because information is ca. 90% of DOC's business, the $40000 -$50000 required to analyse our information needs properly is a small sum compared with the value of a proper strategic information plan.

Subjects within the information system.

Facilities - Recreational
Visitor centre
Access opportunities
Resources
References
Corporate services
Values
Concessions

Territorial
Species and artifacts

A strategic information plan would involve a team of several people working for about a year.

3.6 Establishing Standards

3.6.1 Neil Puller (LINZ)

Standards Linking
Geographic reference (NZ map grid)
Land appellation (Cadastral/Legal description)

Access keys
Street address

Common Data Elements

- Territorial local authority codes
- Departmental codes
- Area
- Standard land use classification (an interim)
- Data transfer standard
- based on American standard -see LINZ News 10
    US SDTS (Spatial Data Transfer Standard)

One of the important things for transfer of data is the **data model**. LINZ staff are interested in the part of our data model relating to information we want to transfer. Appellation (land ownership) is the key. Appellation is not simple though it is gradually being simplified.

LINZ are working in Auckland, and other areas, on an LIS integrating
Maori Land Court records
DCDB
Lands and deeds title index
Valuation NZ database

A Central Index is being developed which includes the following information.
Street address
Appellant
Certificate of title
Value
Owner name
Maori land claim
Gazetteer reference

Neil Puller's strong recommendation to DOC was go into the GIS game with caution. Maintenance of this information is an ongoing overhead, a fact that should always be recognised when establishing a database. He recommended that DOC join the Australasian organisation AURISA; next year their conference will be in NZ. Two final comments were that GIS is a multidisciplinary game and that data quality report very important. The Department of Conservation should maintain close links with LINZ.

3.6.2 Duncan Cunningham -Standards for DOC databases

See paper attached (Appendix II)

Duncan discussed the need for data to be of consistently high quality; for the same (consistent) language; for using the same universal expressions.

Map Series

Duncan recommends maintaining both NZMS 1 and NZMS 260 series because of practicalities; many DOC offices have sets of NZMS 1 maps but not NZMS 260.

We should continue to accept the six digit reference as a basic standard for collecting field information. But for putting into databases we should use full eight or ten figure references.

Discussion

* **Neil Puller** - What Duncan is presenting is OK for dBASE databases BUT is NOT good enough for GIS.

NB -DOC must follow this up with LINZ as soon as possible.

Neil Puller also suggested we should have a field to say what the accuracy of the information is.

Richard Sadleir - We should revert to well-established standards = military standards. To define a square you use bottom left grid reference.

Duncan does NOT want to see blank spaces left in a computer database where there are blanks in grid references. This question was discussed but no resolution was reached. Note the need for further discussion.

Within the database structure Duncan believes we don't need a field for grid system. There was general agreement on this point.

Helen Adcock - Says the program can say which map series is being used.

There was some discussion on dates. Separate fields could be used for day, month and year (e.g. 15 JAN 80), or one field for all (e.g. 15/01/80), but no resolution was reached.

3.7 Jody Richardson – The MAF Fisheries Freshwater Fish Database

See papers attached (Appendix III)

Information on the database forms needs to be checked by a knowledgeable fisheries expert before data is entered.

It is important to make sure all fields are standardised.

At present DOC users have to channel requests through Jody, though they are looking at alternatives.

Map outputs cannot be scaled, only printed on an A4 page.

The database is not linked to any other database.

MAF has a 2-tiered system of charging

e.g. Contributors are charged at a 3:1 basis; i.e. if you contribute 3 records you are entitled to 9 free records.

Discussion

There are 30-50 DOC requests per year.

3.8 Ian Payton - Forest and Wildlands Ecosystems Division, FRI

See brochure attached (Appendix IV).

Indigenous vegetation databases

Effective land management decisions require knowledge of:

What is there (inventory)?

How is it changing (monitoring)?

What is causing the change (process)?

The New Zealand Forest Service undertook two major surveys of indigenous forests:

The National Forest Survey 1947-1956 (with emphasis on timber resource); and the Ecosurvey 1956-67 (North Island).

More recently computer databases of information from these surveys have been established. The National Forest Survey (NFS) Database is managed in Rotorua by John Leathwick. The need for information related to animal control led to the National Indigenous Vegetation Survey Database which is managed in Christchurch. This is a database of vegetation plots established in indigenous forest, scrubland and grassland.

FRI (Christchurch) developed standard data collection methods which were produced as manuals. These methods were used in inventory and monitoring; Conservancies carried out the surveys.

In April 1987 DOC took over ownership of vegetation survey data, and a decision was made by FRI to set up a computerised database.

Why put vegetation information into a database structure?

A database structure can

* provide security
* ensure continuity
* allow integration

However the **investment is enormous**. We don't have resources to do it again.

NB: The database is structured on basis of Ecological Regions and Districts.

FRI is beginning to integrate the database into Terrasoft, using the South Management planning area survey data as a pilot.

Ian was pleased to see Alan Edmonds memo asking for information about the location of field data from previous PNA Programme surveys. Some PNAP data had been entered into their computer system for analysis using their programmes and it was found that some computer files of PNAP data had been scrambled etc.

Ian Payton suggests **that the rest of PNAP data be integrated with their systems.**

3.9 Alan Edmonds - Who co-ordinates the setting up of databases?

About 80% of DOC's work involves resource information (as opposed to management information).

To co-ordinate - we need to understand everything the department does.

We have to understand what DOC's business is about.

We need to set standards so that the consolidation of different data sets is possible.

Different parts of DOC have different needs.

Similarly we need standards so we can use data from other organisations and vice versa.

Financing - who pays for initial development? There will be subsequent users other than the initial developers.

Who pays for the maintenance?

For reasons of economies of scale -the department needs to try to keep to standards, e.g. purchasing sets of software.

It is necessary to clarify the roles of Information Services Unit within the Science & Research Division and Information Systems Development.

The Information Services Unit (S & R) has been a driver - and communicator of what is happening, and has also collated and disseminated information about information.

Information Systems Development develops systems.

<u>Discussion</u>:

Funding and co-ordination for national databases are needed with a group to run it. Conservancies should be funded so that they can create their own databases, but must follow standards. Therefore we need a policy. A case needs to be made for a policy unit for information systems.

Paul Jansen - Raised subject of conservancy staff writing database programs when there should be basic programs being written in Wellington.

Discussion about compatibility of systems to enable data exchange.

Alan explained that **Peter Andrews** is doing three tasks:

(1) He is a consultant with MBA.
(2) In Head Office he has oversight over computer development.
(3) He is carrying out some of the functions (strategic) of an Information Systems Manager.

As far as computer needs are concerned:

Some urgent needs must be solved now and can be – e.g. slow response time, not enough terminals, not enough processing capability on VSs.

In terms of extending into further database (applications) development, we are not in a position to do that in the next two months.

However, there are some people who need databases now.

Up to now we have been focusing on legal arguments and management issues.

The evaluation process for DOCnet Stage II proposals should develop a plan including modification and expansion of the draft information technology policy and strategy.

We might use other agencies to store and manage our information for us.

We could develop a pilot to try and test our data model (or part of it) -including GIS.

Alan Edmonds - Approves involvement of outside consultants in the GIS area to help develop criteria.

Q. Why can we not network the existing PCs?

A. Wang has proposed that we network the PCs as Local Area Networks with gateway to VSs - we could leave DOCFin on a reduced number of PCs (e.g. three).

3.10 Garth Harmsworth - DSIR Division of Land Resources

The New Zealand Land Resource Inventory - a computerised physical resource data base.

Garth described the New Zealand Land Resources Inventory. His paper is attached (Appendix V) as well as one presented by Garth Eyles (Appendix VI) to the workshop on the use of remote sensing in land use management (The New Zealand Land Resources Inventory - and management of the DOC Estate, G O Eyles).

3.11 Diana Kelly - DSIR Library Centre

DSIR Library Databases

See paper attached (Appendix VII).

The databases described are on BASIS software and are run on the DSIR VAX network.

SIRIS is also on the National Library NZBN Database; all DOC librarians can do a search but it is quite expensive.

Searches may be about $70 per hour.

A form of BASIS to be run on PC is being developed.

3.12 David Bowie - PAPYRUS - Bibliographic software

PAPYRUS is a PC based product.

A site licence costs about US $206, and you can make as many copies of the software as you can manage to keep up to date; e.g. S & R could buy a site licence and serve S & R + CASs.

PAPYRUS interfaces with all the standard major word processing packages (but not Wang WPPlus).

For each scientist it can store a "card index" of references from which the references of a particular paper can be produced.

e.g. an S & R bibliography could be created to be shared.

If you type a new reference in your text, PAPYRUS can be set up to take it and add it automatically to your bibliography, giving it a unique number.

It has the reference formats for the major international journals (e.g. Nature) built in.

Article)
Book)
Chapter) for each of these there is a predefined data
Map) structure
Thesis)
Patent)
Quote)
Other)

e.g. for articles - a format

Within the bibliography you can set up a group.

It keeps a record of each group it selects; if you want to keep a group you give it a name.

You can delete the group if you don't want to hold it.

CASs want to keep lists of the publications Jane Napper sends out (S & R and Transfer Funded Agencies).

Bibliographies which used to be maintained by Land and Survey haven't been updated for ages.

Management plans (especially National Park ones) have lots of references.

If all these were entered into PAPYRUS, Central Library staff might would be able to build a large DOC database.

Librarians will provide help, but other parts of DOC will have to pay them.

Thursday 22 March

3.13 Dick Veitch - The Cost of Data Entry

We should be talking about the value of the data rather than the cost of data entry.

There are several different sorts of data in DOC:

Existing data
New local data
New national data.

Existing Data

SSWI is the only national biological data set on card in DOC.

WERI is the only national biological data set which has been computerised.

The first attempt at computerising SSWI (by Linda Hayes) only included part of the data.

If the computer screen exactly replicated the field card, a data entry person could enter the data quite readily.

I.e. the program should allow data entry screens which look just like the field cards.

(NB This would apply equally for PNAP data.)

For all data the question must be asked:

What is the end use?
What is needed?
What is NOT needed?
What STANDARDS should be used?

If everyone follows the same standards then conservancy databases can be combined (e.g. if conservancy boundaries change).

At present we have no proper direction from the top - (computer policy - Strategic Information Plan.)

Another question requiring an answer is:

Who will fund collection of more data?

There are gaps in our information, these gaps must be filled so we can make comparisons between all bits of land.

The data in the database must be carefully chosen so it "doesn't get us into trouble"; it should not be politically sensitive. And it must be easy to interpret.

There must be someone to drive it – i.e. to take the responsibility.

There is also the question of selling of data:

Discussion

Re: selling data - we are constrained by the Official Information Act.

Ian Payton Referred to the Memorandum of Agreement re use of NIVS.

Malcolm Harrison Discussed the cost of collecting the SSWI data which was much greater than the small additional cost of computerising it.

Brian Sheppard Discussed our right to withhold information.

If the courts ask for information we have no right to withhold information and must provide it in whatever form the user needs.

Once we have data collected we cannot withhold it if the courts require it.

Richard Sadleir Discussed the matter of our ability to withhold data for conservation reasons – e.g. rare orchid locations, fossil localities, locations of rare plants which can be used in horticulture.

This relates to the need to take care when setting up databases; think carefully about what goes into the database.

Richard Sadleir - discussed the ownership of intellectual information which is obtained under contract.

David Bowie - the restructured DOC libraries are now centrally run, so that now David Bowie is responsible totally for this key output.

3.14 Ross Pickard - Outputting Data

There are limitations on idea of putting in data directly from data card which relate to the limits of the size of the screen. But you can design the data card on the computer so the card looks nice and the data is subsequently easy to enter.

If people are to be asked to provide data for setting up a database, the card has to look nice - a bit glossy - and be very easy to use.

Ross recommends the use of menu-driven systems (though you can allow other users to use a command-driven system).

DOC has several different sorts of plotters - some with the ability to create points on maps etc.

The question of how to transfer data (how to send it round the country):

> Data can be sent on diskette; but it must be well wrapped.
>
> Or on the VS - through the system administrator, who should know how to do it. If not, ask Ross.

Ross has 20% of his time available to service conservancies.

Discussion

Paul Jansen - Suggested that the scaup, falcon, kea, kaka databases etc. could be in conservancies, i.e. that conservancies could manage conservancy parts of the databases and supply S & R with up-to-date versions of their parts periodically.

Mary McEwen - Suggested that it is a good idea to bring together:

(1) the end users of the data;
(2) experts in the data;
(3) experts in database design;

This was the concept behind the Biological Resources Centre.

Paul Jansen - Commented on central database development:

There should be a database written before a request goes out to collect data, e.g. scaup.

(He set up a dBASE database to enter local scaup data and sent the data to S & R, where it was retyped into the national database).

Dave Crawford - Where we are at now, we have arrived by a natural evolution process. When DOC began the Department was not ready for database standards -we are now. There has been a swing away from centralised information management towards decentralised management, but while there are strengths in the latter, there are also problems.

Richard Sadleir - To sell to senior management that it is a better idea to develop a unified information management system than a series of unrelated systems; it is necessary to show them that it will save money and be more efficient.

Ideally, the workshop should set out a list of problems and recommend a set of solutions to be followed up in six months by a costing of the solutions, but the essential thing is that these solutions save money.

Terry Connor - Without a unified system we cannot communicate with each other.

David Bowie - There are two conflicting ideas:

(1) from conservancies -we need systems now;

(2) from HO we should have standards.

Duncan Cunningham -We should have a working group to guide database developments -to make sure that things follow the main standards but are allowed to vary within guidelines if they need more flexibility.

3.15 Brian Sheppard - Access and Interpretation

See paper attached (Appendix VIII)

Brian described the history of the Archaeological Sites Database, a large database which is a directory to a much greater information system in paper form. It includes 45,000 records, but it is up to date and in daily use.

It is a geographically based system which has potential to be integrated in a GIS.

Brian agrees with the need for standards but standards don't need to be taken too far.

Geographical information identifies the location of the site.

NZMS 1 maps were based on two separate grids - one for each island.

The metric grid is based on a different projection from the NZMS 1 series but a single New Zealand map grid covers the whole country.

Grid References

Eastings and northings are the number of metres away from a false origin to the southwest of New Zealand.

If you see a map number and a grid reference of six digits then you know you are talking about a 100-metre square.

If you see a four-figure grid reference it refers to the whole 1000 m square - but it identifies the lower left hand corner of the square -this is an international standard (i.e. the grid reference does NOT identify the centre of the square).

In any mapping system we have to be explicit on the size of the grid square we are talking about.

Archaeological sites are usually small - usually within a 100 m square - but the system does not allow you to locate the site any more accurately than to the nearest 100 square.

> e.g. a road drawn on a map is actually drawn much wider than it really is, and a house is much smaller than it appears on the map, etc.

What this means is that if site locations are converted by computer from the old NZMS 1 grid to the metric grid, it can be that the archaeological site plots out in a different place (even as much as 400 m away from where it should be).

Plots of archaeological sites can, however, be useful in showing patterns of archaeological sites.

The maps used to plot sites are based on cadastral maps as the data is mostly used by planners.

If users see gaps in the sites, however, they don't know whether this is because:

(1) there actually aren't any sites there, or
(2) this area hasn't been surveyed, or
(3) the information is culturally sensitive and does not appear in the database.

Q. **Paul Jansen** - Surely DOC staff should be allowed access to the secure information?

A. (1) We have to ask should the public have access to knowledge:

Are the sites better protected by being known or by being kept secret?

(2) For archaeological sites - especially Maori burial sites - there is an agreement between archaeologists, the Historic Places Trust (HPT) and Maori people that the information should be kept secret.

If DOC staff are made aware of a planning proposal then they should ask the local Maori kaumatua if it is OK to release information. If they agree, then the information is released in whatever form Maori believe is OK.

Richard Sadlier - For sacred sites especially, DOC is trying to keep the trust of Maori people, it is most important to maintain that trust.

When you look at a map you get some impressions of where Maori sites occur, <u>but</u> you need more information before you can be sure that gaps represent real gaps in sites or simply gaps in knowledge.

A GIS is a manager's dream, but Brian fears that the limitations of our knowledge mean that DOC should not go in the direction of a GIS which overlays WERI, SSWI, Geopreservation Inventory etc., because we will always need experts to interpret what is important about each site.

For example there are problems of the scale at which boundaries have been drawn. The systems allow the user to expand too easily, the scale at which information is displayed - BS wishes that GIS developers would design the software so that a big flashing light appears!! to warn you not to display this information at a scale greater than that at which it was digitised.

<u>Discussion</u>

Mary McEwen made three points:

(1) An example of what Brian is concerned about occurred when Ecological Region and District boundaries were digitised on a 1: 500,000 scale non-metric map. When these digitised boundaries were transferred to the NZLRI GIS system and plotted at 1:50,000, the lines did not all occur in the correct places.

(2) We should realise that it is extremely expensive and time consuming to keep databases up to date. The department is not managing to keep all existing databases up to date, let alone new ones such as CRI.

(3) I believe it will be a long time before the department is in a position to consider having a GIS in every office. In the meantime GIS should be used only to create special maps, using specialists to interpret the information for specific purposes.

Nick Mein - Made the point that the quality of data varies, whether you are using GIS or just ordinary databases.

3.16 Dick Veitch - Bulletin Boards

In the past (in the Wildlife Service) it was possible for staff to know what was going on with the management of threatened species, but now this is much more difficult. All offices need to have up-to-date information to answer questions from the public, e.g. about approximately 100 species, including plants, insects, birds, mammals.

For each species we might need to know:

up-to-date information on the threats to it,
what the department is doing about it, especially if that involves the public?

A bulletin board managed on DOCnet could be very helpful in providing this sort of information and could also be used for communicating information about databases.

Dave Crawford - Within Wang Office there is a system which is quite easy to use but at present it is too slow.

Q. Would it actually be used?

Liz Humphreys - It could be useful for use in recreation, members of the public often want to know about places in the next conservancy. Information about tracks, huts, etc. could be kept on the bulletin board.

Richard Sadleir - There would be a need to control input.

Nick Mein - Remember that most field stations don't have computers.

Paul Jansen - Conservancies now have journalists who could write these public information articles.

Ken Hales - How far has DOC got with a register of expertise (Consultants Register - is in Wren Green's area on a PC).

These things are a good idea but to keep up-to-date and reliable is a big job. Have staff got time?

What is needed is some way to:

Inform staff
Inform public
Control input
Print output

Terry Connor - Management - for any system to be viable it needs proper management; and training - everyone would need to be trained that this information exists, and how to access it.

Paul Jansen - Only certain people in each office need to be able to read the bulletin board – e.g. archaeologists know about the archaeology part, ecologists the ecology.

There is really one level of information only - what DOC staff knows becomes public knowledge very quickly.

Nick Mein - There are two different needs -

(1) to inform staff
(2) to inform public

3.17 Helen Adcock- Demonstration of an existing database - WERI (Wetlands of Ecological and Representative Importance)

Helen described the history of WERI and explained that she was now ready to send conservancy parts of the database to wetlands contact people so that they can manage their parts of the database on PC and update the information in it. Periodically copies of the conservancy databases will be sent back to S & R so that the national dataset can be kept up to date, and an historical record will be kept at each update.

WERI consists of four related databases. The main database (in which most of the information is in a coded form), a plant file, an animal file, and an ecological district file.

When WERI is printed out, all codes are translated into their full meaning, ensuring the minimum amount of interpretation needed by users.

3.18 Geoff Patterson - Conservation SITE Database

Malcolm Harrison developed the SITE data model, a subset of the overview data model. Geoff Patterson wrote the SITE program from the model and sent it out to potential users, asking for feedback. He got none.

Following a meeting held by S & R to discuss the SITE database program, Geoff has added a few new screens.

<u>Discussion</u>

Scientific name versus common name

Graeme Jane - It is a good idea to enter codes for species names so that fewer mistakes are made. Ideally a program should be designed so you can enter codes if you like but it prints out full name.

It was recommended to Geoff Patterson that he programs SITE so users only have to enter six letter codes for species names but it prints full name (as in WERI, see 17. above).

<u>References</u>

David Bowie - Recommended that this section in the SITE database be restricted to photos, maps etc, i.e. non-bibliographic references. Proper bibliographic references should be stored in PAPYRUS.

Brian Sheppard - Asked why are we not using a general database package off the shelf?

Nero Panapa - Wondered if there is too much available in SITE; i.e. that it is confusing to users because it tries to be all things to all people.

Answer: It is not necessary to fill in all the fields. Those which are mandatory should be identified. There must be a code for no information.

Nick Mein - Suggested further standardising, for ease in searching.

Brian Sheppard summed up:

You can set up a database like this as a framework but you need to know what you are building the database for. It's fine to build it for whatever you like, but you should realise that it should not be used for what it wasn't set up for.

Geoff demonstrated the program to potential users.

3.19 Training Plan - Facilitator Brian Sheppard

Dave Crawford - At the beginning of DOCnet Stage I training was seen to be very important and without sufficient training the investment in hardware and software would be wasted.

Questions were asked about:

> Who were to be trained?
> Who were to be the trainers? – DOC or the vendors.

Vendor training tends to be expensive compared with using Department's Staff Training Unit.

Dave thinks training is <u>the</u> most important part of installing new systems.

You can calculate the cost/benefit.

Training allows the learning curve to be smaller.

The workshop brainstormed:

<u>Training requirements in relation to databases</u>

Training should always be tailored to user needs

E = Excellent; M = Maybe; U = Useless

Annual training (should be ongoing)	E
Ongoing	E
PC management	E
Communication with vendor	M
Decentralised training	E
Training tailored to user needs	
Relevant to the job	
General training for support people	E
Utilise full potential of system	E
Training for all staff in basics	E
Keyboard skills PC concepts	
Word processing	E
Data transfer	
Programming	
Spreadsheets	
Database training	
External as well as internal training courses	

Interpretation of data in databases	M
Spatial data compared file data	U
Software training	
Training is a priority	
Manuals/documentation	U
Self instruction courses	U
Systems management	
Use of DOC databases	
Recommendations	
Allocate training resources	

<u>Priorities for training courses related to Databases</u>

- for general DOC staff.

Mt = a management issue

C = a training course.

E, M, U, as above.

Mt	Need for ongoing training	E
C	PC management and PC concepts	E
C	Optimal use of system	E
Mt	Communication with vendor	M
Mt	Decentralised training	E
C	Generalised training for support people/systems management	E
C	Interpretation of data	M
C	Use of DOC databases	M
	Spatial data compared with file data	U
Mt	Documentation	U
C	Self instruction products	U
C	Software training including word processing, database, spreadsheets	E
C	Data transfer	U

<u>Priorities</u>

1. Training of support people - general training in systems management - for both VS and PC.

2. Introductory training/refresher training for all staff -with a proper training programme including specific courses and dates.

3. Training in the use and interpretation of DOC databases.

These are the top three priorities - other ideas recorded above.

3.20 Information Systems Policy and Strategy

Les Jones discussed policy and strategy before the workshop made its final resolutions.

Policy - Defines the framework within which information technology can be introduced into the department; guidelines, principles (high-level rules) that will be adhered to when implementing information technology.

<u>Policy does not include standards</u> but should identify <u>what</u> standards will be followed.

Strategy - Defines <u>how</u> information technology will be introduced within the department; <u>when</u> various components will be implemented.

Strategy also clearly demonstrates <u>why</u> the technology is needed, i.e. how information technology satisfies the department's needs and/or business requirements.

To develop a strategy:

(1) Get EMT commitment

(2) Define where we want to go

Identify the "nature" of the business.

Identify the information requirements needed to support those business objectives.

Identify the processes that take place to support the business objectives.

(3) Define where we are now.

Looks at the current set up.

* Now
* Where

(4) Define how we get to where we want to go.

Applying technology and constraints (budget etc) to the now to produce a path on where to go.

Responsible for policy and strategic direction.

Director General
Executive Management Team & Regional Conservators.
Directors

Central versus distributed database
PCs versus minis
Application areas
Project scope statements

When it comes to the detailed work we must first identify detailed user requirement.

Interview users to determine real requirements.

NOTE: HEAVY CONSERVANCY INVOLVEMENT IS ESSENTIAL

APPENDIX I ORIGINAL PROGRAMME FOR DEPARTMENT OF CONSERVATION NATIONAL DATABASE WORKSHOP

Note: This is the original programme, devised before the workshop took place. Several changes in the order of speakers and events took place, with some speakers unable to attend and others who did speak but were not included in this original programme.

Day 1 (Tuesday 20 March 1990 pm)

Arrival by lunch time

Reports on:

(i)	1.00 - 1.20	Existing databases: A report based on the information we have compiled on paper and existing computerised databases. (Talk: Ross Pickard)
(ii)	1.20 - 1.40	Types of Computer System: What we have now and what it can do, what are the future pathways for the expansion of the DOC computer system? (Talk: Geoff Patterson)
(iii)	1.40 - 2.45	Ways of identifying those databases that need to be computerised. (Talk: Les Jones/Geoff Patterson)
(iv)	3.15 - 3.45	Structuring data: Ways of data collection, and types of databases. (Talk: Terry Connor)
(v)	4.00 - 4.30	Information/Demonstration of existing databases: Protected Natural Areas Register. (Lindsay Daniels)
(vi)	7.30	Request for Proposals: Status of the RFP for DOCnet Stage II (MBA Consultants)

Day 2 (Wednesday 21 March 1990)

(i)	8.30 - 10.00	Workshop: Establishing Computer Database Structures - how does one database fit with the rest of the computerised DOC databases. The role of the SITE data model. (Les Jones/Malcolm Harrison)

(ii)	10.30 - 12.00	Workshop: Establishing Standards - standards are required to make sure that data can be easily interpreted. What standards should DOC be using? - What size should database fields held in common be? (e.g. address fields) - Should certain fields have a standard name? - What form of grid reference should we use? - What sort of documentation should be standard with each database? (Position papers: Duncan Cunningham, Neil Puller [LINZ])
(iii)	1.30 - 2.00	Who co-ordinates the setting up of databases? (Talk: Alan Edmonds)
(iv)	2.00 - 5.00	Demonstration of existing databases
	2.00 - 2.40	Freshwater Fish database Jody Richardson, MAFFISH
	3.10 - 3.50	Indigenous Vegetation databases – Ian Payton, MOF-FRI
	3.50 - 4.20	NZ Land Resources Inventory - Garth Harmsworth, DSIR
	4.20 - 5.00	Diana Kelly, DSIR Library Centre
	After dinner	PAPYRUS bibliographic software - David Bowie

Day 3 (Thursday 22 March 1990)

(i)	8.30 - 9.00	The Cost of Data Entry: What are the costs involved in preparing data for computerisation and how can it be entered cost effectively? (Talk: Dick Veitch)
(ii)	9.15 - 9.45	Outputting Data: What facilities are available to print, display and transport data. How can it best he done? (Talk: Ross Pickard)

Day 3 (Thursday 22 March 1990) continued

(iii)	10.15 - 11.15	Access and Interpretation: Who should be able to change or read data? Should all data be made available or should some interpret it first? (Talk: Brian Sheppard)
(iv)	11.30 - 12.00	Bulletin Boards: An easy way of passing information across a large network? (Talk: Dick Veitch)
(v)	1.30 -3.00	Training Plan: What training is currently available and what future training is required? (Facilitator: Philip Aydon)

(vi)	3.20 – 5.00	Workshop: Creating a database - an exercise in producing a database from aspects of what has been previously covered in this workshop.
(vii)	5.00	Demonstration of existing databases. WERI - Wetlands (Helen Adcock) SITE - Conservation Site Database (Geoff Patterson)

Day 4 (Friday 23 March 1990)

(i)	9.00	Summary - Richard Sadleir
(ii)		Clean-up and depart by 12 noon.

APPENDIX II STANDARDS FOR DATABASES IN THE DEPARTMENT OF CONSERVATION

Duncan M. Cunningham
Science and Research Division
Department of Conservation
P.O. Box 10-420
WELLINGTON

INTRODUCTION

1. This not an exercise in teaching people how to read maps properly and it is not my intention to do so. What I have to say assumes a certain amount of basic knowledge such as map reading and map-interpretation skills.

2. This is a plea for effective communication. If you like, talking the same language with the same accent, using the same universal expressions.

I would like to illustrate what I mean with a short story based in very small part on my own experiences as an introduced species.

I need you to use your imagination in which you are a Japanese tourist (male) who has arrived in Auckland armed with the phrase-book you bought at Tokyo airport.

You are at the bottom of Queen Street on a hot day, with a raging thirst and your tongue hanging out for a cold Kiwi beer. You find a waterfront watering-hole where it's cool and informal. What better a place to slake your thirst, try some legendary Kiwi beer, and try out your skills in the English language.

You go up to the bar, fumbling with your phrase book for the right words, hoping that you manage to communicate what you want.

While you fumble with the dog-eared pages, someone you take to be a local arrives at the bar on your right. He is in fact not a local, he is Scot, from Glasgow no less, who says to the barman "Hey Jimmy, geez a pint".

The barman, himself an expatriate Scot, calmly obliges. You're a bit thrown by this strange order and are equally puzzled when you can't find it in your phrase book. You are still fumbling through your utterly useless phrase a second local arrives on your left. You think "Ah so, better listen carefully". The local leans on the bar and says “Gizza jug mate".

You are struck dumb for several seconds before you are hit with a blinding flash of realisation. Suddenly, you know what to say because you realise that the word "Hajime", spoken by the first man, is also the word in your own language for "begin". You put two and two together, it makes perfect sense to you so you confidently say to the barman "Hajime, gizza jug mate!".

The barman looks at you with a mixture of astonishment and undisguised pity and your heart sinks when he says “Ah’m awfy sorry Jimmy, you'll have to speak English."

Apart from amusing you, I hope I've made my point about the danger of sticking to a way of saying something that you like, and makes perfect sense to you. It may not make perfect sense to someone else.

So, with that very important message firmly in mind, I would now like to discuss the best ways of expressing some fundamental pieces of map-derived information.

EXPRESSION OF MAP NUMBER

When map-based information on a floppy disk comes in to a database manager, it should ideally be completely compatible with the data in the main database. If it has to be corrected or re-arranged either before it goes into the database or before it is sorted within the database, a lot of time can be wasted. This is inefficient and defeats the purpose of using a computer-based database.

For example, in the case of single-digit map numbers, if a space or a zero is not inserted before the digit, the number e.g. 7 (007) will be treated by dBASE as 700 in a data sort. The same is also true for two-digit numbers e.g. 12 (012) which will be treated as 120. So your new records, instead of being fitted in with all your records from map S 12, will end up way down the file with records from map S 120.

This may seem obvious to you, but it is surprising how often it does occur. It also may seem trivial but time spent correcting these every time a data set comes in is time wasted which can be so easily avoided.

Why do I want zeros entered rather than blanks? The trouble with blanks is that they are so easily and so often forgotten about or ignored. If those who enter the original are in the habit of entering zeros before one and two digit numbers (because we've decided it's to be the standard), there will never be any need for corrections, and more importantly, there will never be any doubt that the number is correct. Thus we will save time, we will make the best use of databases, and at the same time, we help each other.

Summary

A standard expression of map numbers is needed to ensure uniformity throughout DOC. A problem occurs most frequently with imperial (NZMS 1) map numbers, many of which consist of one or two digits while in a database they must occupy a three digit space. Zeros should be entered before the digits to ensure consistently correct placing of the records within a database.

STANDARD FOR GRID REFERENCES

Map Series

The standard map series recommended for DOC is the metric, NZMS 260 series (1:50,000). This is the standard recommended by LINZ (Land Information New Zealand). However, whilst it is desirable to adopt the metric system as soon as possible, it may not be immediately possible for some parts of the country. There are two reasons for this: firstly because metric map coverage is not yet complete, and secondly, the cost of buying a complete new set of maps may be prohibitive for some field centres.

It is therefore important to maintain databases with both imperial and metric grid references to service DOC offices which do not have complete sets of metric maps, and the public, many of whom still have old maps. When you have fields for both imperial and metric references, your options for input and output are not restricted.

Some conservancies may have no need at all for imperial maps (or metric as yet), in which case there is no need to clutter up your databases with data which cannot be applied.

Standard grid reference

Resolution

Defining the grid reference to the nearest 100 m is the norm and should continue to be acceptable. Finer resolution is difficult to achieve with standard maps (1:50,000 or 1:63,360).

Very specialised databases e.g. the Archaeological Sites Database do need finer resolution to within 1.0 m. However, this can only be achieved with the appropriate scale maps and a thorough understanding of survey techniques.

Size

The standard six-digit (three digits east, three digits north) map reference is the long-established standard that we can, and should continue to expect from field staff and members of the public. However, it is only a basic standard so I think it is highly desirable for DOC staff who are working with <u>any</u> map-based survey material to have a broader understanding of how maps work. Like any tool, if you understand how it works, your ability to use it is greatly increased.

A database is also a tool and the more detailed the information you put into it, the more you increase the options for using that information. To be specific, if you put 5 x 5 grid references into your database, you increase the mapping and plotting capabilities of your database.

3 x 3 grid references can still be extracted from 5 x 5 references if needed but 5 x 5 references do give more information and anyway, the 3 x 3 component is neither difficult nor inconvenient to extract by eye.

Outputs from databases need not contain every piece of data and can be tailored to meet more basic needs for 3 x 3 references. I know that some people think that 5 x 5 grid references are difficult to use and prefer to stick to the familiar 3 x 3 references, but I would argue that few DOC staff would be reduced to gibbering, slobbering wrecks by the sight of a list of 5 x 5 grid references! I think we need to look to the future and encourage the output <u>users</u> within the department to accept 5 x 5 grid references.

Summary

Databases should ideally contain both imperial and metric grid references to meet the needs of users who do not have complete sets of metric maps.

Resolution of grid references should be standardised to the nearest 100 m, which is current practice. However, some very specialised databases need finer resolution.

The 3 x 3 grid reference is still the basic standard and database managers should continue to expect grid references to come in this form.

Databases must contain more detailed information if they are to be useful in the future. The 5 x 5 metric grid reference should become the standard for DOC databases, while at the same time, the imperial equivalent grid reference (4 x 4) should be retained where users are likely to need it.

STANDARD EXPRESSION OF GRID REFERENCES

Split references

One of the problems associated with managing a database is the tremendous variation in the way people present a grid reference. Most people simply present a 6-digit reference as a string of six digits. However, many people also split the reference into 3 x 3 and it is here that the confusion arises. There seems to be as many ways of expressing the split between the eastings and northings, as there are individuals.

On one occasion I received information in which an NZMS 1 series, eight digit reference was given as 5351-5412. At first reading I thought this was meant to be <u>from</u> 10,000 yard square 5354, <u>to</u> square 5412 and of course this was impossible to plot! After several attempts at trying to think like the writer of this reference, it turned out that this was his own way of expressing full eastings and northings.

To eliminate confusion caused by variation, particularly between databases, we need to adopt standard ways of communicating information which is as precise as a grid reference.

Examples of these variations are:

	6 digit	8 digit (imperial)	10 digit (metric)
1	354412	53545412	2945463650
2	354-412	5354-5412	29454-63650
3	354-412	5354-5412	29454-63650
4	354/412	5354/5412	29454/63650
5	354.412	5354.5412	29454.63650
6	354 412	5354 5412	29454 63650

We need to decide which of these options are the most acceptable and stick to them. Options 1 and 6 for six digit references are, I believe, the most acceptable as the figure is not so long that it needs to be split into 3 x 3. On the other hand, eight and 10 digit references are, because of their length, difficult to read and transcribe. Transcription errors occur with greater frequency with longer strings of digits. Dashes have other functions (see below) which, if used in this context, can cause confusion. References greater than six digits need to be split into eastings and northings in a clear and consistent manner.

I propose that option six above is the clearest and certainly the simplest. It is the most acceptable for databases as eastings and northings should be separate fields for ease of programming and machine handling. Also, the resulting printout from a database would have a blank space between separate fields.

Blanks in grid references

The next logical step in this standardising process is to establish how we express grid references which describe national map-grid squares, e.g. 1,000 m and 10,000 m (or yard) squares. A 1,000 m square from 6 above, would be 2945 6365. Similarly a 10,000 m square would be 294 636. The problem with these is that they look like something else - eight digit and six digit map references respectively. In addition, if we simply leave empty spaces where we've dropped off digits, e.g. 294   636, not only does it look odd, but the validity of the whole reference can be thrown into doubt. A database operator or output user could be forgiven for questioning whether the positions of the digits are correct.

I originally proposed that the "blank" spaces where the numbers were dropped off, be filled with zeros. It was pointed out to me that the resulting grid actually represents a specific position on a map. We have to find another way of filling in those blanks. It is in this context that I feel dashes are appropriate, for example:

1	Y15 29454 63650	= 5 x 5 grid reference to nearest 100 m
2	Y15 2945- 6365-	= grid reference for 1,000 m square
3	Y15 294– 636–	= grid reference for 10,000 m square

Information entered into a database in this manner is thus complete, leaving no doubts about its accuracy and no room for incorrect interpretation.

I propose that blanks should not exist in grid references except as a space to separate eastings and northings. Where digits have been dropped off to indicate a grid square, the resulting spaces should be occupied by dashes.

Summary

The department should adopt a standard way of expressing grid references which are split into eastings and northings. I propose that the standard expression is to leave a blank space between the two sets of digits. This is the clearest and simplest method and is exactly how a reference would appear on a printout where eastings and northings are separate fields.

Where a reference is shortened to indicate a grid square of 1,000 or 10,000 m, the resulting spaces should be occupied by dashes to indicate to users that the reference is accurate and complete.

DATABASE STRUCTURE

Field names and their order

Virtually all maps are based on either the old imperial grid or the new metric grid with a range of map series derived from each. Given then that there are only two grid systems to consider, and that we can see a need in many cases, to include grid references from both systems, do we need a field for "map series"? I put it to you that on the one hand, a database containing references from two or more map series within the same field would be a many-headed monster. It would be totally resistant to sorting and using in any logical order. On the other hand, descriptive field names for grid references can indicate the grid system used e.g. "IMPMAP", "IMPE" (Imperial Map Easting), "METN" (Metric Map Northing).

A field list would look like:

IS(LE)	IMPMAP	IMPE	IMPN	METMAP	METE	METN
X	XXX	XXXX	XXXX	XXX	XXXXX	XXXXX

e.g.

N	133	3634	3188	U21	27824	61662
S	059	5113	6317	K33	23978	58094
S	059	511-	631-	K33	2397-	5809-
S	059	51--	63--	K33	239--	580--

Multiple references

Linear survey

Where two grid references are given to describe a start and finish point along a track, river, or coast, the field list can be:

IS(LE)	METMAP	METE1	METN1	METE2	METN2
X	XXX	XXXXX	XXXXX	XXXXX	XXXXX
e.g.					
S	K33	23978	58094	23903	58116

Area survey

Where an area is described, it seems pointless to try and list all the relevant map references around an irregular shape such as a lake, particularly if it is a large one such as Lake Te Anau. In such cases there is no substitute for a picture to give an immediate impression of shape and size sketch, map, or photograph. A central reference point to indicate the location is the least complicated way of saying where something is.

Reference to the map or photograph and where it can be found can be made in the database entry in the notes or memo field. If it is a geographic feature already depicted on a standard map, a few words in the notes field to say so is all that is needed.

Summary

A "map series" field is not needed as grid references are given from, at most, two grid systems, imperial and metric. Descriptive field names can be used to indicate whether the grid reference is imperial or metric.

Eastings and northings should be in separate fields to make the database easier to read and to make the data more accessible for plotting and mapping.

"From" and "to" references should be included in separate fields where appropriate, as these can easily be fitted into most screen displays.

Lists of references to describe areas, many of which may be large and irregular, are inappropriate when a single, "Central Reference Point" accompanied by a picture (or reference to one) is far clearer and less time-consuming to use.

DATE FORMAT

Field format

This is an area where personal preference appears to take precedence over the need for a common standard. dBASE users differ considerably in opinion as to which format is the best. The two most common options are for "character" and "date" formats. Each format has a differing level of acceptance of blanks and zeros, and each performs a sort differently, according to recognition and treatment of zeros.

In addition, a "date" field, quite apart from keeping a village of Arabs from starving, will only accept dates. It will reject a date which is only a month and a year. If you have been supplied with such a date, you have to make up the rest of it (i.e. a false date) just to make it acceptable to the software. I do not find that acceptable! Someone else is going to see that date and take it at face value. Even you, who entered the date in the first place, might look at it some time later and wonder if it's correct!

Entering a date as a character field gives you far more flexibility, allowing for an honest entry, whatever it may be. It is in the expression of the character field that we need to agree.

Expression

The standard way of expressing date in the non-American world is the dd/mm/yy e.g. 28/01/88. I don't think there is any argument about that, and I would be surprised if there is any need to discuss whether we use slashes between the days, months, and years.

Examples of dates:

Whole date	Month + Year	Year only
08/01/88	/01/88	/ /88
8/1/88	/ 1/88	/ /88
	01/88	88
	00/01/88	00/00/88
	–/01/88	–/–/88
		1988

Summary

"Character" date fields allow any kind of date to be entered and are therefore more suitable for databases which receive less precise dates in which only a month and year, or only a year are given.

A consensus is needed on a standard expression for dates. Given that the standard order is dd/mm/yy, and that slashes between days, months, and years are also standard, how do we fill in the rest?

Examples of some possibilities are presented for discussion and resolution.

GEOGRAPHIC LOCATORS

For some databases it is important or useful to have a field indicating DOC Regional Conservancies or Ecological Districts. There are various ways of achieving and expressing this given two existing conditions:

Some records may be of interest to two or more conservancies. The field size would have to be wide enough to cater for this. Where more than one conservancy code is needed, codes can be separated by an oblique slash, viz WH/WK/TT. Individual codes are then selected using the search "within" ($) command.

It has been pointed out that a new system using letters should not be created as codes for conservancies already exist in the form of budget codes which are numeric. However, for the sake of user-friendliness, it may be preferable to use two-letter codes which make the conservancy immediately obvious e.g. TT = Tongariro/Taupo, OT = Otago. The purpose of this code system would be distinctly different from the budget system and in many cases be operated on completely different software.

Summary

A code indicating the conservancy which may have an interest in a given record may be useful on many databases. Information for a conservancy can be extracted very quickly by selecting the records "tagged" with their code.

Codes can be either the existing two digit numbers as used for the budget (DOCfin) system or a two letter abbreviation of conservancy names. The latter is regarded as the more user friendly.

CLOSING COMMENTS

If we are to eliminate error and doubt, we must ensure that whatever we communicate cannot possibly be interpreted any other way. Databases which contain highly specific information such as grid references must be managed in such a way as to ensure that data residing in a database, or a subset of a database, is structured and expressed in an identical way to data in another.

I believe that the key to maximising compatibility within and between databases lies in the Chinese expression of “gonghe” (which, thanks to the US Marines, we know as "gung-ho") which literally means "working together".

It is imperative that we at this workshop work together to reach agreement on database standards, and we leave here with a set of positive recommendations to be approved and implemented immediately. Let us work together, now, to set these standards, and, from here on, to maintain them.

APPENDIX III THE MAF FISHERIES FRESHWATER FISH DATABASE

Jody Richardson
MAF Fisheries
P.O. Box 6016
ROTORUA

INTRODUCTION

Many DOC staff members will already be aware the freshwater fish database exists, but I welcome the opportunity to expand your knowledge about it, as well as perhaps giving you a few ideas for improving your own databases. The freshwater fish database has been in existence since 1977 and is extensive and comprehensive, a database DOC should be consulting frequently.

The topics covered in this presentation include an evolution of the freshwater fish database why and how the database was started, and the steps we went through to get where we are today. This will lead into a presentation of the database today what information is collected and how it is stored? Input and output information includes how we and DOC can enter data and what options there are for retrieving data. Naturally, costs and charges cannot be ignored. Finally, a few tips for avoiding problems with computerised databases are presented.

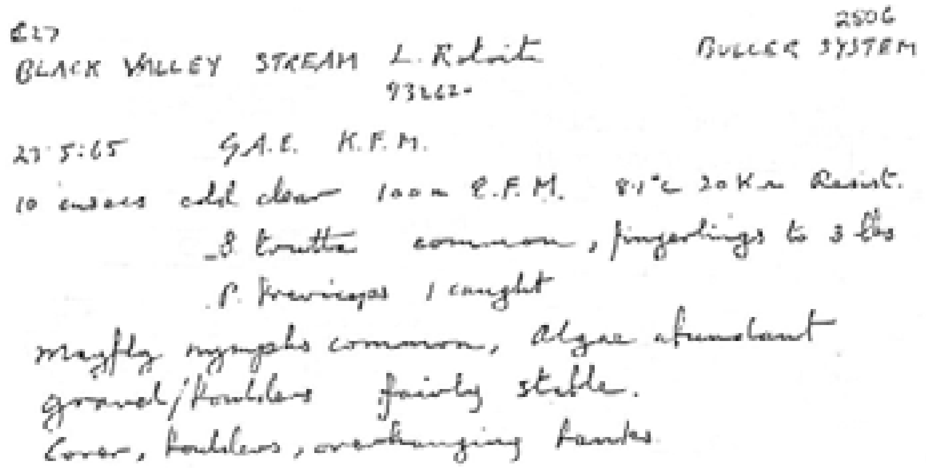
EVOLUTION AND STRUCTURE OF THE DATABASE

The freshwater database is essentially a collection of site-specific fish records - here's where we went and this is what we found. Data collection of this sort has occurring in New Zealand since 1773, when one of James Cook's naturalists collected a giant kokopu in Dusky Sound. However, it wasn't until the late 1960s to early 1970s that more systematic surveys got underway, and there are two primary reasons for this. One is the development of sophisticated and portable electric fishing equipment, virtually essential for reliably collecting New Zealand's small and cryptic native fish particularly. The other is that about that time, the fauna became well enough defined taxonomically for everyone to agree on what they were collecting, and therefore give some measure of confidence to the data.

Computers were uncommon in the late 1960s, but fisheries staff, being responsible scientists, did ensure their data was systematically recorded. Fig. 1 is an example of one of the first database records, a handwritten non-formatted card. Although *P. breviceps* has been assigned to another genus now, the card contains accurate and useful information, and is not very different from what is recorded today. There is an identification number (c29), information about where the sampling occurred, when, who, and how the data were collected, and what was found. In addition, there are also some comments about the invertebrate population, cover, the substrate composition, and the habitat type. This sort of card continued to be filled in for the next 10 years or so, and these were simply filed in some sort of order, probably alphabetically or by the card number, and manually sorted when necessary. However, during the mid to late 1970s, fisheries research began to become involved in more and more potential impact studies; what's going to happen when we build a hydro dam here, or take away half the water there? By that time, there were hundreds of these cards filed away, with much useful information, and the need for an easy and flexible way of retrieving data as background material for our studies became evident. Thus, in 1977, it was decided to computerise the database.

One of the fist things you must do in establishing a computerised database is to decide what the data will be used for. At that stage we were mainly concerned with our own needs, which were to know what species were found where. So, site-specific type data, retrievable over land or catchment areas, or by species, were determined to be satisfactory. Because we wanted to collect information from as many people in the field as possible, a standard form was considered essential to clearly set out what we wanted to record. Fig. 2 shows the form we developed in 1977. Again, still where we fished and what we found, although we have insisted on that most useful of inventions, a map reference, to pinpoint the sampling location. We also put in a catchment number so we could have some way of retrieving the data from individual catchments, and decided we wanted quantitative information about the substrate composition and habitat type rather than just written comments. Collectors were told what was the most essential information to fill in, and cards were bound into booklets of 25 with some instructions on the inside cover about what some of the categories meant.

Figure 1. An example of a pre-computerisation database card



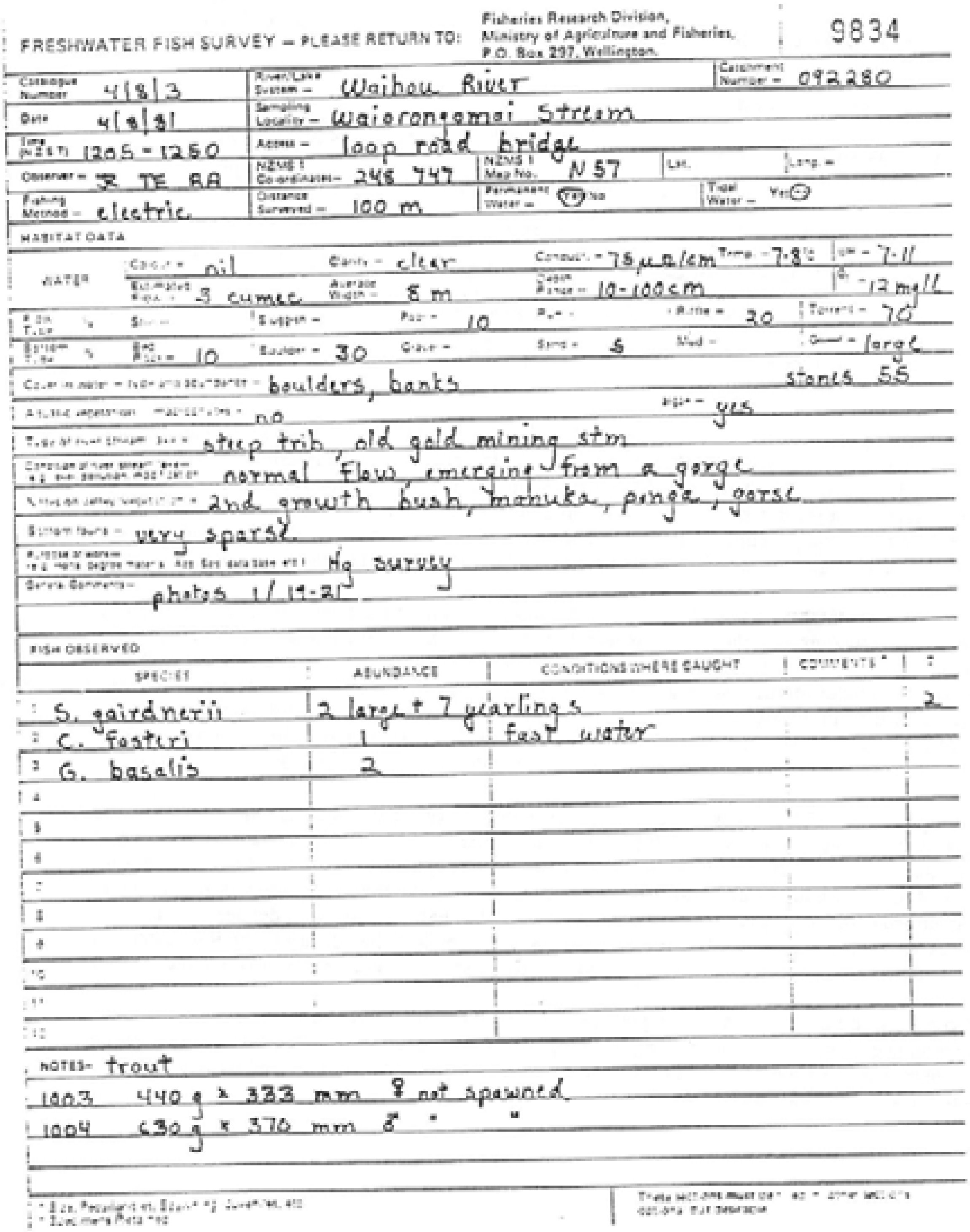
Unfortunately, the fisheries computer available in 1977, a PRIME, had no database handling packages and data were entered as text files rather than in tables. Programs written in BASIC were used to search and retrieve the information. This limited the amount of data that could actually be entered, and only 11 components from the database cards were stored on the computer files. The BASIC programs could search on three components, catchment number, map references, and species names, and printouts were produced in a single standard format containing all the information that had been entered for each card.

Soon after the database was established, all the information from the hundreds of cards that had been quietly accumulating over the years was transferred to new cards and entered, and this formed the historical resource. In 1978, all staff were issued booklets of cards and told to use them when in the field. Booklets of cards were also issued to all acclimitisation societies, wildlife service offices, regional water boards, and private consultants, many of whom elected to become contributors and users of the database. Over the next few years, quite a bit of information began to accumulate on the computer and in 1983, a brochure was published to aid users of the database (McDowall and Richardson 1983).

During 1986 an NRAC fellow, Ken Minns, undertook to review the database, which then contained about 6500 records, to see if it was still fulfilling user needs. After his review, Ken had two primary criticisms of the database. The first was that its usefulness for analytical purposes was limited by a lack of quality control on many entries due to not enough use of categorical descriptions. Categorical descriptors are predefined choices, for example on the database card, the habitat types were delineated into pools, runs, riffles, etc., but what about valley vegetation? On the example in Fig. 2, vegetation is well-defined, and all the comments could probably be categorised into 100% scrub. However, that was not always the case.

Secondly, Ken highlighted the incompleteness of the computer storage; here was a card with much quantitative information on it, but only 11 components could ever be searched for and retrieved. Following the review, a committee was formed and we decided to expand the database and designed a new card, which is the one in use today (Fig. 3).

Figure 2. An example of a database card - version one.



FRESHWATER FISH SURVEY — PLEASE RETURN TO: Fisheries Research Division, Ministry of Agriculture and Fisheries, P.O. Box 297, Wellington. 9834

Catalogue Number 4/8/3 | River/Lake System — Waihou River | Catchment Number — 092280

Date 4/8/81 | Sampling Locality — Waiorongomai Stream

Time 1205 - 1250 | Access — loop road bridge

Observer — R TE BA | NZMS 1 Co-ordinates — 248 747 | NZMS 1 Map No. N 57 | Lat. | Long.

Fishing Method — electric | Distance Surveyed — 100 m | Permanent Water — Yes/No | Tidal Water — Yes/No

HABITAT DATA

WATER: Colour — nil | Clarity — clear | Conduct. — 75 µs/cm | Temp. — 7.8° | pH — 7.1

Estimated Flow — 3 cumec | Average Width — 8 m | Depth Range — 10-100 cm | — 12 mg/l

Pool — 10 | Riffle — 30 | Torrent — 70

Bed Rock — 10 | Boulder — 30 | Gravel — | Sand — 5 | Mud — | Other — large stones 55

Cover in water — boulders, banks

Aquatic vegetation — no | Algae — yes

Type of stream — steep trib, old gold mining stm

Condition of stream — normal flow, emerging from a gorge

Vegetation — 2nd growth bush, mahuka, ponga, gorse

Bottom fauna — very sparse

Purpose of sample — Hg survey

General Comments — photos 1/ 19-21

FISH OBSERVED

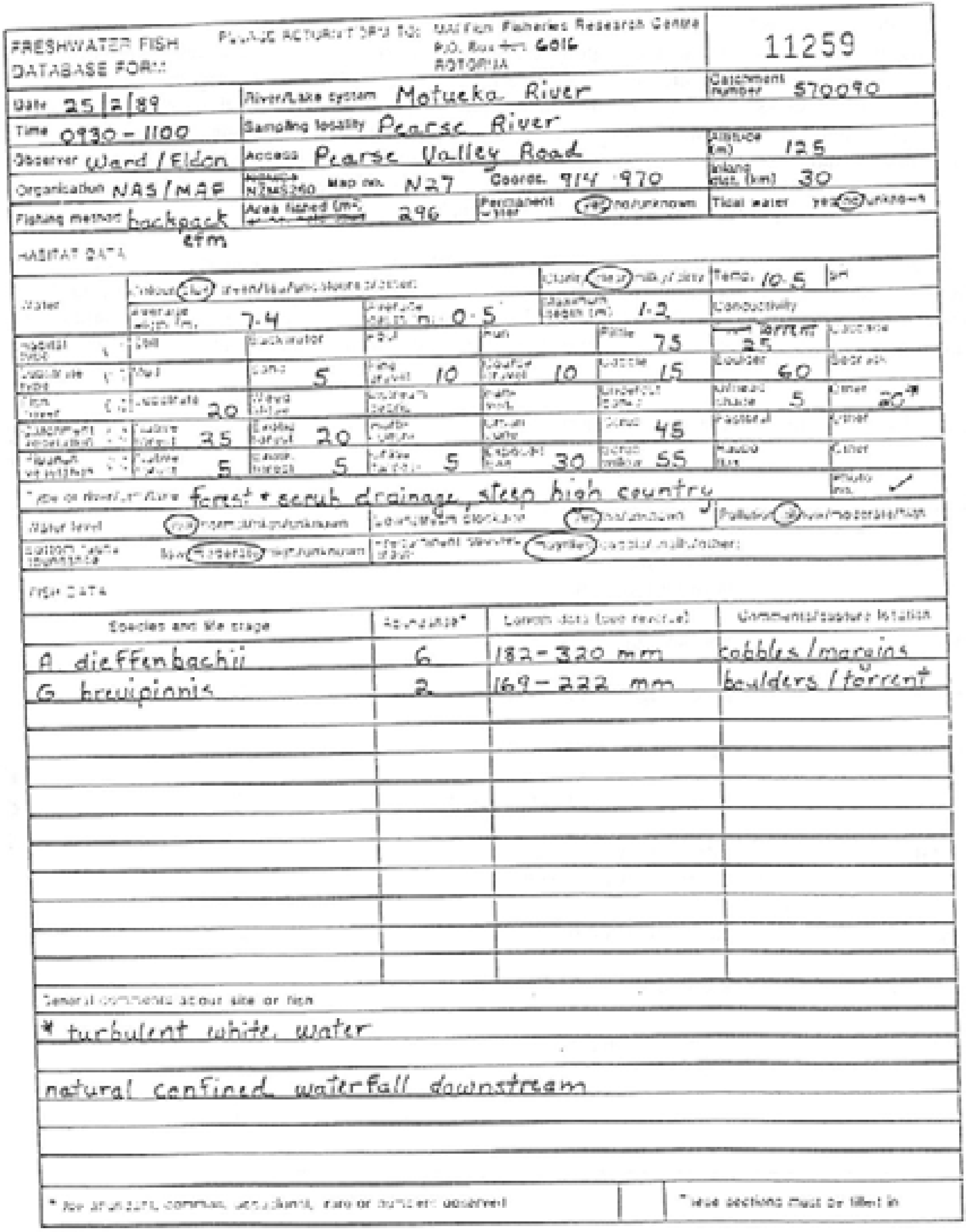
	SPECIES	ABUNDANCE	CONDITIONS WHERE CAUGHT	COMMENTS	
1	S. gairdnerii	2 large + 7 yearlings			2
2	C. fosteri	1	fast water		
3	G. basalis	2			
4					
5					
6					
7					
8					
9					
10					
11					
12					

NOTES — trout

1003 440 g x 333 mm ♀ not spawned

1004 630 g x 370 mm ♂ " "

Figure 3. An example of a database card - version two.



FRESHWATER FISH DATABASE FORM

MAF Fish Fisheries Research Centre, P.O. Box 6016, ROTORUA

11259

Date 25/2/89 | River/Lake system Motueka River | Catchment number 570090

Time 0930 – 1100 | Sampling locality Pearse River

Observer Ward / Eldon | Access Pearse Valley Road | Altitude (m) 125

Organisation NAS / MAF | NZMS260 Map no. N27 | Coords. 914 970 | Inland dist. (km) 30

Fishing method backpack efm | Area fished (m²) 296 | Permanent water yes/no/unknown | Tidal water yes/no/unknown

HABITAT DATA

Temp. 10.5

Average width (m) 7.4 | Average depth (m) 0.5 | Maximum depth (m) 1.2

Riffle 75 | Torrent 25

Sand 5 | Fine gravel 10 | Coarse gravel 10 | Cobble 15 | Boulder 60

Substrate 20 | Overhead shade 5 | Other 20*

Tussock 35 | Exotic forest 20 | Scrub 45

5 | 5 | 5 | 30 | 55

forest + scrub drainage, steep high country | Photo ✓

FISH DATA

Species and life stage	Abundance*	Length data (see reverse)	Comments/capture location
A. dieffenbachii	6	182–320 mm	cobbles/margins
G. brevipinnis	2	169–222 mm	boulders/torrent

General comments about site or fish

* turbulent white water

natural confined waterfall downstream

The new card is still basically where we fished and what we found, but many more categorical descriptors have been added to increase quality control. We also decided we wanted more quantitative information about land use practices and fish cover for numerical analysis. We have given users the option of metric or imperial map references, and inserted a few items which might help explain fish distributions; the altitude, distance from the sea, whether there is a downstream barrier to fish passage, etc. As before, the most essential information to fill in is shaded, and cards are issued in booklets of 25, with some instructions on the inside cover.

Rather fortuitously, in 1987 when the review was finally coming to a conclusion, fisheries purchased a PYRAMID 98x, which is located in Wellington, and a database package called EMPRESS. The PYRAMID is a mainframe computer, and EMPRESS is a relational database management system similar to or INGRESS. EMPRESS stores data in tables of information and allows you to search for any item in any table. The output can be produced in list form similar to our original PRIME printouts or put into tables of data ready for use with statistical packages such as SYSTAT or MINITAB. A disadvantage of EMPRESS is that it is not menu driven, so users must learn and remember commands to access data, a potential problem for computer illiterate staff.

In 1988, the database was moved from the PRIME to the PYRAMID and we began entering all the data on the new card. The data are stored in nine separate tables, each containing a similar type of data, and are linked by the unique card number (Table 1). In total, we have the capacity to store 85 separate bits of information from each card, with much more of the information coded in some way to save space on the computer.

Table 1. Freshwater fish database computer structure.

Table Name	Type of Information	No. of attributes
Records	Frequently requested information about site data	19
Species	Species found data	10
Flowdata	Percentages of habitat type	7
Landuse	Catchment vegetation type	2
Microdata	Depth, velocity, and cover data	14
Reachdata	Infrequently requested information about site	17
Ripveg	Riparian vegetation data	2
Samtab	Method and observer data	5
Substrate	Substrate composition data	9
	Total	**85**

The table "records" mainly contains information from the top shaded box on the cards (where we fished) while the table "species" consists of information from the fish data section (what we found), and these are by far the most frequently used tables. However, with a relational database package, it's possible to extract data on any attribute or combination of attributes, so it's an extremely powerful tool for retrieving data.

Cards entered since the PYRAMID transfer have had as many of the 85 attributes filled in as possible, but cards entered between 1978-87 contain only the original 11 items. This is a major, ongoing job, to complete data input for all those cards, but only about 1000 have been done up to now. Users of the database were notified about changes to the system (Richardson, 1989) and at the beginning of March 1990, the database contained just under 8400 records.

INPUT AND OUTPUT

Card filled in by staff and other contributors to the database are posted to Rotorua and I enter the new data. This is a time consuming job, but we've found the cards need to be checked over by a knowledgable fisheries person first. This is because we have many people with different levels of expertise contributing data, but also reflects inadequacies with the cards. For example, if fish size information is given, as in Fig. 3, at what size does a koaro become an adult? To be consistent, we have decided on a size, but then must ensure all the data entered conforms. Its also quite astounding how many people get their map references back to front and how many people can't add up to 100.

A BASIC program is used to enter the data, although EMPRESS will allow you to enter data directly into tables. The advantage of the BASIC program is data can be entered more or less in the order they appear on the cards, and the program automatically ensures that each bit of data ends up in the right place in the right table. The program also has many checks built in. For example, if the sum of the substrate composition doesn't add up to 100, the computer bleeps, and goes back to the start of that section. Similarly, if the species is misspelt, another bleep and prompt for re-entry. Only myself and one other person can enter or change data. Thus, inputting data for DOC staff is simple. You obtain a booklet of blank cards, fill them in, post them back, and we check and enter the data.

Anyone involved in freshwater fisheries research or management is entitled to request information from the database. All MAF Fisheries freshwater staff can use and read the database themselves, and we have developed some pre-packaged command files for those unfamiliar with EMPRESS. These simply list the EMPRESS commands you would normally have to type in, and prompts the user to enter one or more variables to search on via the keyboard. Users outside MAF, like DOC, have to channel their requests through me at present, although we are looking at other options to this.

As mentioned before, data may be selected on the basis of any of the 85 attributes or combination of attributes on the computer. Therefore, it's important to decide exactly what you want before accessing the database. The most frequent request received is for a printout of all the data within a particular river catchment or geographical area (Fig. 4). This printout for the Ngaruroro River catchment contains a card number, location, date, NZMS 1 map and the species found. The content of the printout is historical, mimicking what used to be printed out by the PRIME, but is by no means all that's available. If having additional information, such as the collector's initials, or the length of stream fished, or any of the 85 attributes which are stored would make the data more useful for your purposes, then it's up to you to decide.

Often catchment printouts are fairly lengthy and difficult to wade through, so maybe a summary is all that's needed. For example, printouts could be broken down into the species found in the mainstem, the major tributaries, minor tributaries etc. Species specific printouts are also frequent requests, and it's possible to list just the localities where rainbow trout were found in the catchment, for example.

Maps can also be produced from database output. Fig. 5 is an example of a regional distribution

Figure 4. An example of a catchment printout from the freshwater fish database.

```
PRINTOUT FROM THE FRESHWATER FISH DATABASE FOR THE
NGARURORO RIVER CATCHMENT
March 6 1990

card: 4195
catch: 231
locality: Ngaruroro River
sheet: n113
easting: 3709
northing: 3710
date: 25/03/83
commname: Rainbow trout
commname: Brown trout

card: 4196
catch: 231
locality: Ngaruroro River
sheet: n123
easting: 3727
northing: 3589
date: 25/03/83
commname: Rainbow trout
commname: Brown trout

card: 4197
catch: 231
locality: Ngaruroro River
sheet: n133
easting: 3789
northing: 3383
date: 25/03/83
commname: Rainbow trout
commname: Brown trout
commname: Black flounder
commname: Longfinned eel
commname: Common smelt

card: 8369
catch: 231
locality: Raoraoroa Stream
sheet: n123
easting: 3698
northing: 3563
date: 19/01/87
commname: Longfinned eel
commname: Rainbow trout
```

Figure 5. An example of a regional distribution map from the freshwater fish database.

Galaxias argenteus

map - giant kokopu locations in the DOC Southern Region. Maps can also be plotted for the whole of the North or South Islands, for individual river catchments, or for other geographical regions like national parks. Maps can only be drawn to fit on an A4 sheet of paper and thus it is not possible to scale the maps to any NZMS map series, such as 1:250,000.

The freshwater fish database is not directly linked to any other databases. For example, if you wanted to know whether the giant kokopu locations shown in Fig. 5 were linked to a specific rainfall pattern or soil type, then you would have to access another database separately and link the data. However, because the freshwater database has geographical references, I think this would be possible.

Generally, we don't do much interpretation of the data when it's sent out, although this depends on the client's background, what they ask for, and how much they are willing to pay for. Information from the database has been used as background material for catchment inventories and protected natural area surveys, as evidence in water right and water conservation order hearings, and in environmental impact statements and Freshwater Fisheries Reports. The data has not been used for analytical purposes, mainly because quantitative information has only been entered over the past two years. However, this is an area MAF Fisheries has identified as high priority research.

COSTS AND CHARGES

What does it cost to retrieve data? Charges are based primarily on the number of records retrieved in each instance, with a bit of extra for processing. As mentioned before, agencies other than MAF Fisheries contribute to the database. To protect these agencies from incurring costs when retrieving information, a two-tiered system of charging is applied. Essentially, contributors to the database get credit for cards they send in to offset charges for extracts at a ration of 3:1. If you send in three cards, for example, then you are entitled to retrieve nine at no cost. Non-contributors are first charged a substantial access fee and then up to five times as much for each record retrieved. Details about charging are set out in the policy statement attached as Appendix A. Obviously for heavy users of the database, it pays to become a contributor.

Each DOC region is considered separately as a contributor or not to the database. Of the 14 DOC regions, six are contributors; Waikato, Bay of Plenty, East Cape, Tongariro/Taupo and Canterbury. Other agencies who use the database, such as acclimatisation societies, private consultants, regional councils, etc., are treated in the same manner.

What does it cost to maintain the freshwater fish database? Based on last year's figures, about $10,000 a year, of which 66% is salary costs. The total also includes a couple of trips to Wellington to sort out any problems we are having, plus allows for the purchase of maps, computer paper, and the printing of blank database cards. Income from requests covers about 10-30% of the maintenance costs.

AVOIDING PROBLEMS WITH COMPUTERISED DATABASES

Since this is a database workshop, I thought I would draw on my experience with computerised databases and point out a few things that can go wrong with them. These fall into two problem areas; one is that the database is under-utilised and the other is that the output is misinterpreted (Table 2).

Computerised databases are wonderful tools, allowing vast funds of knowledge to be easily tapped. But they are also expensive and time consuming to maintain, especially when data input is an ongoing process as with the freshwater fish database. It is therefore important that they are used to their potential. Reasons why a database may not be fully utilised could be its potential was overestimated and that computerisation was unnecessary. However, it's more likely the database is too difficult or expensive to access or use, or even that people are unaware of its existence. Maybe you aren't collecting or storing all the information users need, as we found when the freshwater fish database was reviewed in 1986.

Table 2. Avoiding computerised database problems.

Problem area	How to avoid
Under-utilisation	Know your audience Conduct periodic reviews Store what you collect Publicise your database Make database user friendly Network the database Keep cost as low as possible
Misinterpretation	Use a single form to collect data Quality control data entry - delete doubtful language - check and double check data entry - limit entry access Learn the database language Know data limitations Know clients limitations

Ways to avoid these problems include first and foremost knowing your audience - who is going to use the database and what information do they want? It's likely this will change over time, so periodic reviews of the database are also required. Store as much information as you collect. Data collection is expensive compared to computer storage, and presumably if you are collecting a certain type of information, then it has a foreseeable end use. Publicise your database and make retrieving data as easy as possible using menus or pre-packaged command files. Giving users "hands on" access to the data will also increase use, as will keeping charges as low as possible.

Errors that can cause data misinterpretation might arise from relatively simple things such as typographical errors during data entry, data of doubtful quality, or misidentification of species. Avoid these by using a single form so contributors clearly know what information is required, and then check and double check the data as they're entered. Limit the number of people who can enter or change data.

It is also important to be sure you know what you are asking the computer to extract. For example, with EMPRESS if you neglect to add certain conditional statements during a select, then it's possible to end up with lots more data than you should. Understanding the foibles of your database language and developing fool-proof pre-packaged command files should help.

Knowing the inherent limitations of your data is also essential. Site specific presence/absence data like the freshwater fish database are open to misinterpretation. A species might not be recorded at a particular site on a particular occasion, but may occur at sites up or downstream, or seasonally. With data of this sort, all you can be really sure of is that a species was definitely found at a site at a particular time, not that it was absolutely absent.

Know the limitations of your clients as well. Basic biological knowledge is usually necessary for interpreting output. For example, most native freshwater fish migrate to or from the sea during various stages of their life cycles. The absence of a particular species from a site may not be due to unsuitable habitat, but to the fact that it's the wrong time of year or that there is a blockage to migration downstream. Interpret for users if you have any doubts about their subject knowledge or ability to interpret output. I hope these tips will help you avoid major problems with your databases.

To conclude, I would like to issue a plea to all the remaining DOC regions to become database contributors. DOC is one of the heaviest users of the freshwater fish database and it would be desirable for MAF not to have to charge the regions for information. Booklets of blank cards and information from the database can be obtained from me in Rotorua.

REFERENCES

McDowall, R.M. and Richardson, J. 1983. The New Zealand freshwater fish survey: guide to input and output. Fisheries Research Division *Information Leaflet No. 12*. 15 p.

Richardson, J. 1989. The all-new freshwater fish database. *Freshwater Catch 41*: 20-21.

APPENDIX A FRESHWATER FISH DATABASE POLICY STATEMENT

Policy statement on data handling, retrieval, and charging for services.

<u>Background</u> MAF must now charge clients for retrieving data from the freshwater fish database. However, confusion has arisen over ownership and distribution of data that non-MAF Fisheries agencies contribute. There have also been questions about the confidentiality of data sent in by non-MAF Fisheries agencies, as well as inconsistencies in charging for services. This policy statement will help clarify matters for agencies contributing and requesting data, and MAF Fisheries staff using and servicing the database.

<u>Contributing data</u> Only agencies who regularly send in cards will be regarded as contributors. Regularly means at least once a calendar year. This policy will apply to the DOC regions, acclimatisation societies, catchment authorities, etc. separately.

All data sent into MAF Fisheries will be available to all database users unless the contributing agency specifically requests that the data be kept confidential. Data will only be kept confidential for a maximum period of three years after which it will be made available.

<u>Retrieving Data</u> Only legitimate and specific requests for data will be processed; printouts covering all the data over vast land areas will not be issued. By making requests at the appropriate time when an issue becomes current, clients are assured of having the most up-to-date information held in the database.

All MAF Fisheries freshwater staff have free access to the database and can process their own requests. They are <u>not</u> to process any other requests.

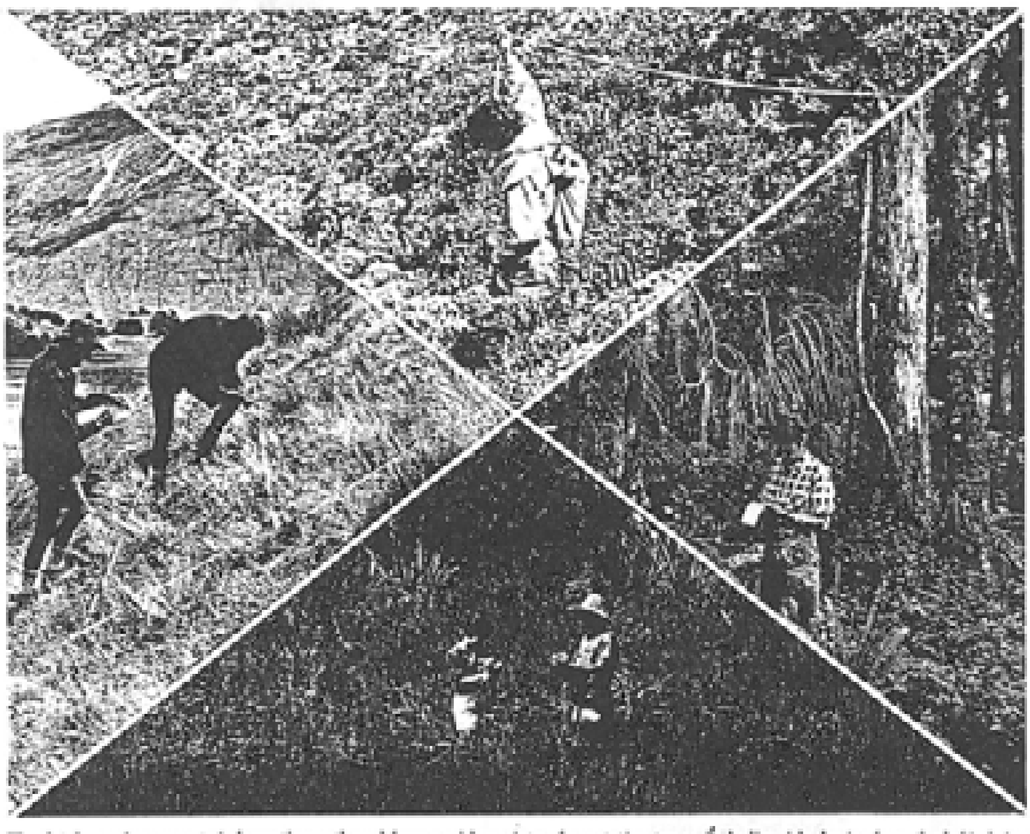
<u>Charging for Services</u> Contributors who wish to retrieve data are charged for this service. Charges to contributors are on a "per record retrieved" basis of $1 per entry, plus a processing fee of $10 per request. Charges to contributors will be offset by credit gained for cards sent in (at the rate of $3 per card) as well as credit gained if another contributor accesses their data (at the rate of $1 per card). This credit may only be used to offset costs of retrieving data for legitimate requests and should not be regarded as payment for contributing to the database.

Non-contributors are charged on a "per-record retrieved" basis of $5 per record for entries less than 2 years old, and $2 per record for all others, plus an access fee. At present the access fee is $100. Where a large amount of data is accessed, fees acquired from non-contributor requests will be paid to the originator of the data on the basis they were charged for. Access and processing fees will be paid to MAF Fisheries.

What's New in Forest Research

No. 175 1989

Databases for New Zealand's indigenous vegetation



The databases incorporate information gathered from a wide variety of vegetation types, including (clockwise from the left) alpine grassland, subalpine scrub, lowland forest, and wetlands.

Despite major changes to New Zealand's forests and grasslands since man settled here, large areas of indigenous vegetation remain. Recent estimates put the total area of indigenous forests at 6.2 million ha, with shrubland, wetlands, unimproved grassland, and the alpine zone accounting for a further 9.3 million ha. This land is under increasing pressure from both recreational and commercial users. For it to be managed wisely, decision-makers need to know the nature of the vegetation in each area and how it is changing.

FOREST RESEARCH INSTITUTE

Private Bag 3020, Rotorua, New Zealand. P.O. Box 31-011, Christchurch, New Zealanc

Since the 1940s, New Zealand Forest Service Conservancies, FRI, and (since 1987) the Department of Conservation, have been involved in collecting data on New Zealand's indigenous vegetation. By 1988 over 300 surveys, involving more than 70 000 plots, had been carried out in lowland and high country areas (Fig. 1). To make the accumulated data available to managers and researchers, staff of the Forest and Grassland Ecology group (FRI, Christchurch) and the Indigenous Forest Management group (FRI, Rotorua) have compiled two databases — the National Indigenous Vegetation Survey (NIVS) Database, and the National Forest Survey (NFS) Database. Both databases run on the Ministry of Forestry's VAX computer network. Specially written computer programs enable users to retrieve, analyse, and summarise information.

Fig. 1 – Surveyed areas of indigenous vegetation which have provided data for the NIVS Database (black areas) and the NFS Database (coloured areas).

Vegetation surveys

Vegetation surveys are carried out for specific purposes which dictate what type of data is collected and how (*see* Table 1). For example, the NFS Database contains data that were intended to assess the extent of New Zealand's native timber resources. The data, some of the earliest collected by the New Zealand Forest Service, included information on tree size and density, timber volume, and regeneration. In contrast, a recent survey of northern Fiordland by FRI, Christchurch, focussed on habitat use, and resulting damage to vegetation, by deer and possums. For this survey, canopy and understory density and composition, site characteristics, and degree of animal browse were some of the measurements recorded.

In each survey unit (e.g., a river catchment) representative plots are selected to provide estimates of the plants present. The vegetation and site characteristics are described and/or measured, and recorded on plot sheets. Unrecognised plants are taken back to the laboratory for identification. The 'raw' field data are then entered into the computer, checked for errors and missing components, and added to the database. The data can be listed or specially written programs can present them as summaries or analyse them further. The results can be viewed on screen or as computer print-outs, including maps.

TABLE 1 - Survey methods for which data are currently held on the indigenous vegetation databases

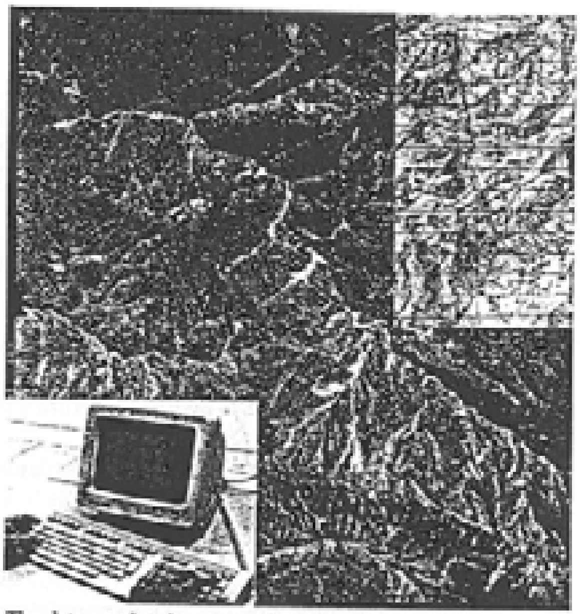
Database	Survey method	Forest	Shrubland	Grassland	Wetland
NFS	5 chain by 2 chain plots	+			
NIVS	Cruciform *	+			
	Recce (descriptive)	+	+	+	+
	Seedling	+	+		
	Quadrat	+			
	Variable area	+			
	Point height intercept *		+		
	Frequency			+	
	Photocentre *			+	
	Ring transect			+	
	Miscellaneous	+	+	+	

* not currently recommended

The NFS Database

The NFS Database contains data collected from two surveys: the National Forest Survey (NFS) of over 15 000 plots, collected during 1947–56; and the North Island Forest Ecological Survey (Ecosurvey) of over 4000 plots, collected during 1956–67 to extend coverage to the North Island forests not covered by the NFS. Large amounts of information are involved; only certain components of the raw data have been entered into the computer: density of trees (and poles, Ecosurvey only) by species and size class; a list of the minor species recorded for each plot; location, using metric and imperial map co-ordinates; altitude; aspect; slope; topography; drainage; and canopy density. Timber volumes and seedling data have not been included in the computer files, but these are held with all the other raw data, which include plot sheets, maps, and aerial photographs. By February 1989, c.92% of the relevant survey data had been entered into the computer.

Data can be extracted for plots within a specified area or containing specified species. Output can be as printed summary tables of species composition and site characteristics, and/or as maps showing plot locations which can be overlaid with information on site characteristics (e.g., altitude).



The data on plot sheets, aerial photographs, and maps are transferred to computer files.

The NIVS Database

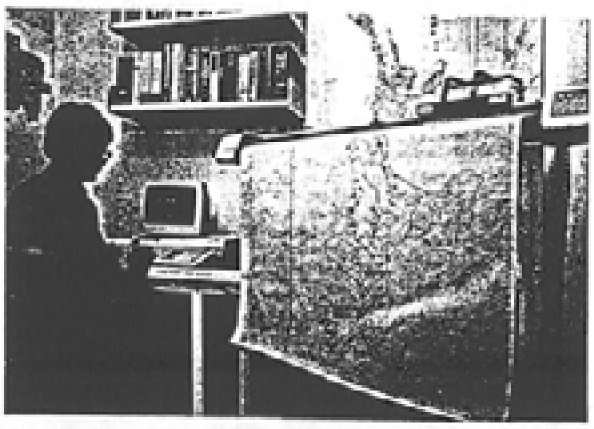
The NIVS Database contains data from more than 50 000 plots that have been collected during vegetation surveys carried out by the New Zealand Forest Service (now the Ministry of Forestry and the Department of Conservation), other than the NFS and Ecosurvey data. As for the NFS Database, all plot sheets, maps, and photographs for the surveys are held in storage. However, unlike the NFS Database, all the data on the vegetation and site characteristics in the NIVS Database are being entered into the computer. By February 1989, 427 computer data files had been included in the database, and it is estimated that most data entry will be completed by 1991.

Because the NIVS Database is a collection of the data from many surveys, the database contains two directories: a survey directory listing the surveys' agency, organisers, relevant maps, and the data recorded, and a directory of the computer files.

The computer data files are checked for errors and missing data using programs written by FRI staff. The descriptive data can be summarised to detail for any species, on one or more plots: presence; frequency (i.e., number of plots on which the species occurs); and degree of cover. Site characteristics, such as altitude and aspect, can be summarised at the same time. Other programs summarise data collected on stem density, diameter, and basal area, and distributions for individual species, plots, or forest types.

Comparing plant communities

A range of data analysis programs enables the user to describe the vegetation from a specified area in a number of ways. Many of these programs use sophisticated mathematics, made possible only by access to powerful computing facilities. For example, plots can be sorted into groups with similar species composition using either divisive ("splitting") or agglomerative ("clumping") classification procedures. Or, plots can be arranged along gradients based either on measured environmental factors (e.g., altitude) or on their species composition. Other programs enable users to identify indicator species of different communities or to compare their species composition and sites factors.



An example of vegetation plot data overlain with a physical data set.

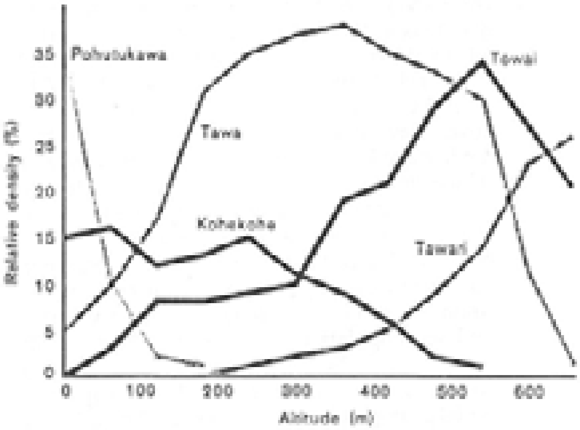
Typical uses of the databases

Both databases, and the additional programs for summarising and analysing them, are valuable for helping to answer a range of management and research questions. Their nationwide coverage lends them to broad-scale regional analyses.

For example, the NFS Database has been used by the Department of Conservation as part of its Protected Natural Area (PNA) survey programme on the Coromandel Peninsula. Tree data for that region were extracted and analysed, producing eleven major forest classes. Those involved in the PNA programme were then able to concentrate on areas not covered by the NFS Database.

Staff of FRI and the Department of Conservation are currently the principal users of the NIVS Database, extracting print-outs of various data files or analysing sets of survey data. One use has been to compare changes in vegetation over time, in an attempt to determine whether various populations of introduced browsing animals (as defined by the Wild Animal Control Act) are at "acceptable" levels.

Use of the databases can be arranged for anyone that is linked to the Ministry of Forestry's VAX computer network either directly, or indirectly such as through PACNET. For other clients FRI staff can access the databases on their behalf.



A graph of the relationship between the relative density of five tree species and a site factor, extracted from a database.

Future possibilities

The flexible nature of the databases means that they can be modified to meet the needs of their users. The Ministry of Forestry and the Department of Conservation are keen to promote wide use of the information in both databases.

Data from future vegetation surveys will be added to the NIVS Database, as they become available. Also, since the original plot records are held, permanently marked plots can be relocated and remeasured.

The set of programs written specifically for the NIVS Database is still growing. Scientists are designing a package to analyse quantitative data collected from grassland plots and another that will extract relevant data for any particular area (e.g., an ecological region or district). Also, programs such as those already available for the NFS Database will be added that will overlay vegetation information with that on site characteristics. However, these programs require map co-ordinates for each plot, which the NIVS Database currently lacks. Inclusion of these will be a major on-going task.

Commercial information systems are also being assessed for their usefulness to the databases and whether using them to manipulate information will improve our understanding of natural systems. Geographical information systems (GIS) are packages for analysing information about a wide range of geographical and biological components collected from plots and currently are under scrutiny.

Summary

Data collected during many surveys of indigenous forests and grasslands have been entered into databases on the Ministry of Forestry's VAX computer network. The two resulting databases have been designed for data analysis, to help answer a range of management and research questions. Future research will investigate the likely benefits of expanding this system to include information about other biological and geographical components, using recently developed geographical information systems.

This article is based on the work of :
I.J. Payton, L. Burrows, C. Hall, J. Hunt
Forest Research Institute
P.O. Box 31-011,
CHRISTCHURCH

J.R. Leathwick
Forest Research Institute
Private Bag 3020,
ROTORUA

ISSN 0110–1048

PRINTED BY ROTORUA PRINTERS , ROTORUA - 1989

APPENDIX V THE NZLRI - COMPUTERISED PHYSICAL RESOURCE DATABASE

G.R. Harmsworth
DSIR Land Resources
Private Bag
PALMERSTON NORTH

INTRODUCTION

The New Zealand Land Resource Inventory (NZLRI) is a nationwide inventory of the physical characteristics of our land. Five dominant physical factors are recorded: rock, soil, slope, erosion, and vegetation. It is the only database (available as printed worksheets or as a computer database) in New Zealand which classifies land according to its suitability for sustainable use (i.e. Land Use Capability), using an eight class system consistently and on a national basis. Since mapping began in 1973 (originally at a scale of 1:63,360 and since 1985 at 1:50,000), the New Zealand landscape has been divided into approximately 100,000 individual map units or management units and each assessed according to the Land Use Capability (LUC) system of land classification.

Approximately 800 LUC units (significantly different types of land) have been defined in the NZLRI within eleven distinct regions, each with its own regional classification. LUC units from each North Island region have been correlated.

LAND USE CAPABILITY

The word "capability" introduces the concept of "sustainability". In the NZLRI, capability is used to define the lands capacity for permanent sustained agricultural production. In this context agriculture is defined as being land uses such as cropping, grazing and forestry. It means that the prime thought in the mind of the mapper when classifying land is both the "capability" and the "sustainability" of each piece or parcel of land, not just the physical and chemical properties of factors such as soil and rock. This immediately brings a management component into the mapping. It involves thinking about the land's present versatility, the risk or hazard to the land by introducing certain land uses or types of management, and if the land is degraded or physically limited, the potential for improvement to the soil resource, or at least to maintain it in it's present condition. The prime concern is to classify and supply information for planning purposes, so as to avoid or mitigate degradation of the land resource (e.g. soil or rock type), so that the land resource is at least maintained for future generations. Unsuitable land use practices, for example pastoral farming on marginal lands, usually have far reaching and long term effects on the environment. The capability ranking (i.e. LUC Classes I to VIII) indicates the degree of physical limitation, thereby affecting versatility. It can also indicate the level of degradation (from its original or pristine state).

Examples of unsustainable land uses are seen in the Taranaki and East Coast regions, where present pastoral land use on marginal hill and steepland country has exacerbated erosion, increased the amount of sediment entering streams and rivers, increased the frequency and magnitude of flooding, and the resulting deposition has ruined lowland areas and reduced both the areal extent of productive land and the potential for cropping. The steep hill country areas would have been better left in forest. In most cases erosion and vegetation are inextricably linked. Removal of a forest or scrub cover on marginal lands almost always results in increased erosion and run-off and increased flooding and sedimentation downstream.

Unsustainable management in New Zealand is not only restricted to steep hill country areas but is frequently evident on our high producing arable (or cropping) lands. Unwise cultivation above certain slope angles often leads to excessive erosion (e.g. on the Bombay hills south of Auckland), cultivation during dry seasons on some soils often leads to serious wind erosion problems stripping the soil resource (e.g. parts of the Canterbury Plains and North Otago), and where the soil is unable to cope with the sustained pressure of repeated cultivation - ploughing, compaction, cropping - many areas are showing signs of serious soil degradation (e.g. in the Manawatu on some of the best cropping soils in New Zealand).

The Land Use Capability system of land classification is designed principally to recognise all the lands physical limitations (e.g. erosion), and the word "capability" emphasises the importance of classifying land as though it was being viewed in the present and in the future (e.g. forested or scrub covered arable land is assessed as though it could be in cultivation). Therefore this systematic approach classifies land according to long term effects and implications of different types of land use and management.

The multi-factor approach used in the NZLRI (i.e. recording the five dominant factors: rock, soil, slope, erosion, vegetation) allows a more holistic approach to be employed for land use and/or environmental planning than that developed from single factor mapping systems. Integration of separately derived single factor data is often a problem. Therefore the multi-factor approach and land use capability assessment makes the NZLRI database unique, and a useful tool for land use and environmental planning.

THE NZLRI MAPPING SYSTEM

Mapping comprises subdividing landforms into areas which are physically similar. The subdivision of landforms into map units is based mainly on five main inventory factors: rock, soil, slope, erosion and vegetation which are recorded in each map unit on the NZLRI worksheets. Where any significant difference occurs in one of these physical factors a new map unit is formed. Primary map units are based on rock, soil and slope. Further subdivision into secondary map units is based on erosion and vegetation. A Land Use Capability (LUC) assessment is then made for each map unit, following final compilation of each individual map unit. The assessment not only takes into account the five dominant physical factors mentioned above but is also based on factors such as climate and effect of past land use (e.g. historical evidence which references any misuse or abuse of the land).

The LUC units are more than just subdivided landform units - the mapper takes into account factors such as climate, altitude, the potential for erosion, wetness or flood hazard, critical slopes for inducing erosion, and past land use and management. These factors, if significant, can often result in different inventory map unit boundaries being drawn to those which would result from straight landform analysis. The major difference between a LUC unit and a landform unit is that the LUC unit is defined on a land management basis taking into consideration factors such as climate, altitude, potential erosion, and potential effects of management. The LUC unit is therefore more akin to a management unit.

Single factor datasets, such as soil or geology maps on their own, very seldom consider management factors, or their implications for long term land use. Also, they are designed with many different purposes or objectives in mind, and are produced at a variety of scales (e.g. time-stratigraphic geology maps). Therefore it is often difficult to integrate all this information together for soil and water conservation planning. Most single factor datasets are primarily concerned with recording the present characteristics of physical factors such as soil, rock, or vegetation. Geology maps are mainly concerned with recording the age, structure, and stratigraphy of the rocks. Soil maps usually classify land solely on the basis of the soil parameter, such as maps showing suitability for cropping.

A Land Use Suitability map is very different from a Land Use Capability map, in that the latter takes into account versatility of land and risk to the physical environment. In a LUC assessment, the rock factor (e.g alluvium) or soil (e.g. organic soils such as peats) showing similar characteristics could be further subdivided on the basis of flood hazard or continuing wetness and further delineation would be at the class level (e.g. Class VII, VI, IV, III). The same applies to subdividing hill country areas, where with LUC mapping a critical slope angle may be regarded as most important in terms of erosion risk. This is why the multi-factor approach is so important for planning, because the emphasis on each factor can change from area to area, region to region. The purpose of each study will also affect the way certain factors from the NZLRI are emphasised or prioritised.

BACKGROUND

Early discussion on Land Classification in New Zealand began principally in response to the widespread mismanagement of land following agricultural development. It also followed in the wake of storm damage and flooding which is a relatively common occurrence throughout this country. Early work; such as Taylor's 1938 paper on Land Deterioration in Heavier Rainfall Districts of NZ received considerable political attention. Taylor's principal recommendation was that the land should be classified and areas unsuited to pasture should be put into other land uses. He stressed at the time that farmers and society had a common interest in preventing further deterioration, Taylor was commissioned by Government to head a "Committee of enquiry into the maintenance of vegetative cover in NZ with special reference to land erosion". Government had asked for reliable information "on the extent to which soil deterioration was taking place, as a prerequisite for taking measures to prevent further losses in soil fertility". By June 1939 the report was complete. It recommended that "statutory and administrative measures should be taken to inaugurate a programme to handle serious soil erosion, soil conservation, and land utilisation problems. Such a programme should involve the co-operation and collaboration of foresters, botanists, agriculturalists, engineers, and soil technologists". War delayed the publication of this report until 1945.

Other major work at this time were Vic Zotov's surveys of the Tussock grasslands of the South Islands in 1939. He also stressed the seriousness of the soil erosion problems. Gibbs and Raeside finally published their survey of soil erosion in the South Island High country in 1945. Continual lobbying and hard work by a small group of dedicated New Zealanders finally led to the Soil and Rivers control Act becoming law in 1941, followed by the establishment of the Soil Conservation and Rivers Control Council, and the setting up of Regional Catchment Authorities. These regional authorities were phased out in late 1989. Leaders in the Soil conservation field in these early days (1930s-1950s) included Doug Campbell, Lance McCaskill and Kenneth Cumberland.

In an attempt to better co-ordinate soil conservation activities between 1955 and 1966 several soil conservation staff were moved into the Department of Agriculture, though still working on programmes determined by the Soil Conservation and Rivers Control Council. But this actually separated work on rivers from that on land. The consequent lack of an integrated approach to soil conservation created major difficulties. Another report by D.A. Williams (Chief of the US Soil Conservation Service) commissioned in 1964 determined that an integrated programme of research and policy at both national and regional levels was needed and hence the idea of an agency involved in Water and Soil research and policy was conceived. Finally after much debate a Water and Soil Division was established within the Ministry of Works. In 1967 the Water and Soil Conservation Act was passed which finally linked management of water and land. However, soil and water research continued to be somewhat uncoordinated between a number of organisations. Another report prepared in 1973 by E.G. Dunford found there were difficulties in the focus of research in the areas of water management and soil conservation. He recommended that a problem-oriented research programme be carried out for the Soil Conservation and River Control Council and Catchment authorities, by a single unit from within the MWD.

At this time there were difficulties however, particularly in areas concerning responsibility of the land and water resource. This concern led to the National Water and Soil Conservation Authority (NWASCA) being formed in 1983 from the merger of the SCRCC and a parallel body, the Water Resources Council. But in the late 1980s, NWASCA was disbanded as part of government restructuring. During this restructuring the Water and Soil Division of the MWD was also disbanded, with the old Water and Soil research centres along with the NZLRI database becoming a part of the DSIR and the policy side of the Division going to MFE.

With the establishment of regional councils and the onset of the new environmental legislation in NZ, the problem of focusing on soil and water conservation issues and accompanying research will undoubtedly again need to be addressed.

BACKGROUND OF TIIE NZLRI

The NZLRI was established to provide national standards and clear guidelines for both LUC assessments and erosion severity assessments. It was primarily set up as a national resource base for soil and water management in this country. Up until completion of the NZLRI survey, there were varying standards for land resource evaluation, making government subsidised erosion control grants difficult to administer.

It is important that we continue to use a whole catchment approach for soil conservation and water management in NZ. The multi-factor system in the NZLRI has been highly suited to this type of approach. Many environmental studies require that the data be considered using a holistic approach, looking at a number of factors from whole areas whether they be at the catchment, district, or regional levels. It will also be important in the future to interpret the NZLRI data accurately, and to be aware of its limitations, such as scale. It is therefore very important to understand the Land Use Capability system when making interpretations of the NZLRI data.

APPLICATIONS

When the NZLRI mapping commenced in 1973 it was not envisaged that the application of the NZLRI data would be so wide and varied. It's usefulness for planning purposes is mainly attributed to the system being multi-factor based and it's focus on management oriented land units. It is now recognised that land use and environmental planning relies on the evaluation and interpretation of a large number of key environmental factors, often from a wide area. When discussing soil and water management, it is necessary to think about those environmental factors within the area of interest and furthermore, implications for offsite or downstream areas often outside the area of interest. With regional, district, or catchment level planning it is short sighted to study environmentally linked areas in isolation.

ENQUIRIES FOR NZLRI DATA

The multi-factor approach and the capability assessment in the NZLRI has allowed a wide variety of specific purpose studies to be carried out utilising a number of key land resource factors. This generally involves selection of one or all of the inventory factors or LUC parameters. Productivity information is also frequently required.

Enquiries may range from simple to complex. For example, the enquirer may simply request presentation of just one factor, say the NZLRI vegetation factor (on a map unit by map unit basis), or may specify vegetation data with additional interpretation (e.g. re-classifying the data into specific groups), or ultimately complex interpretation (where almost new data is presented - based on the original NZLRI data). Equally the enquirer may wish to be presented with maps and tables of a combination of inventory factors (e.g. soil, slope, erosion, vegetation) and LUC, or have a more complex interpretation of these combined factors carried out. Further interpretation of original NZLRI data often involves further analysis - such as photo-interpretation, interpreting tables or computer plots, evaluation of certain physical factors, or specialised projects involving remotely sensed techniques such as satellite imagery.

DATA PRESENTATION

The data is available in a variety of forms, but most commonly as computer plots and tables with accompanying reports. Because the NZLRI has been mapped at a scale of 1:63360 and more recently at a scale of 1:50000, it is more acceptable to produce computer plots at similar scales or smaller scales than those of the original NZLRI mapping scales. But this depends entirely on the type of study and the degree of interpretation.

Data at more detailed (larger than and 1:63360 and 1:50000) scales is usually presented following further analysis or modification to the original NZLRI For example, using subsequent interpretation to re-draw map unit boundaries at larger scales, based on existing NZLRI data, or re-classifying an inventory factor such as vegetation onto a larger scale base. Therefore where increased detail or analysis is required, further interpretation is always necessary to ensure that the data is portrayed as accurately as possible.

EXAMPLES OF WORK CARRIED OUT FOR THE DEPARTMENT OF CONSERVATION

Aotuhia District - East of Stratford in Taranaki

The Government decided to redevelop the Aotuhia block in the 1970s and control was handed over to the old Lands and Survey Department. During restructuring of Government Departments in 1988 the Government announced that Crown lands were to be re-allocated. The Department of Conservation (DOC),Wanganui, requested information from the NZLRI for land use decisions.

Subsequently this exercise involved both extraction of data from the computer and photo-interpretation.

DOC requested a computer plot showing three categories of pastoral potential for a number of development blocks in the Aotuhia district. The categories were:

a) land with long term pastoral potential
b) land with marginal pastoral potential and requiring intensive soil and water inputs
c) land unsuitable for pastoral use

In addition, a computer plot of vegetation was also requested. This required further interpretation using photographs so that vegetation could be categorised according to the Department of Conservation's vegetation classification.

Plots at 1:50000 were produced. Pastoral plots were produced by grouping certain LUC units together. For the category c) - unsuitable land - further information was used from earlier research (unrelated to this exercise) in the Taranaki-Wanganui areas, which had been carried out by the Water and Soil Division based at Aokautere, Palmerston North. The results from this earlier work further defined the interpretations derived from the NZLRI database. It had been found that slopes above 28^{o} were particularly susceptible to landslides. This corresponded with the F slope class (26-35^{o}) in the NZLRI and indicated that most Class VII land was unsuitable for pasture. Existing NZLRI map unit boundaries were then used as a basis for further analysis and delineation using aerial photo interpretation.

A combination of the Pastoral map and the Vegetation map was then used to identify the areas which were regarded as not being sustainable under present land uses, and therefore land which should be allocated to the DOC estate.

Matemateonga Ecological district

This study was carried out for DOC, Stratford office, in 1987 as part of a protected natural areas evaluation. The first part of the project involved digitising the Ecological district boundary. The data was then presented as computer plots of Land Use Capability (LUC), vegetation and slope within district boundaries. Accompanying tables of LUC showing areas was also produced. Vegetation plots showed dominant vegetation within map unit areas, and slope was re-grouped into four categories.

Manawatu Ecological region

A plot of land systems within the Manawatu Ecological district was required. The first stage for this work was to draw the ecological district boundary on screen and extract the area of interest. All LUC data within the district boundaries was then plotted. LUC units in the plot then needed to be grouped and re-classified into the appropriate land system categories requested (e.g. sand country, alluvial plains).

REFERENCES

Baumgart, LL. 1989: "Soil Conservation and Land Use Planning". New Zealand Soil News 37. No. % pp. 111-115.

NWASCO 1979: Our Land Resources. National Water and Soil Conservation Organisation 1979: Wellington, New Zealand 79 p.

Soil Conservation and Rivers Control Council 1971: "Land Use Capability Survey Handbook" (2nd Ed.). Water and Soil Division, Ministry of Works and Development Wellington, New Zealand. 139p.

APPENDIX VI THE NZLRI - AND MANAGEMENT OF THE DOC ESTATE

G.O. Eyles
DSIR Land Resources
Private Bag
PALMERSTON NORTH

INTRODUCTION

In this paper I will explain some of the ways the NZLRI can be of use to DOC staff. The maintenance of this data set is supported by loop funding from DOC indicating its historical use for conservation purposes. Hopefully its value to DOC will be seen as being sufficiently important to warrant supporting its maintenance in the future!

We have found over the years that the NZLRI data set has many more uses than could have been imagined when we initiated the programme in 1973. At that time it was established to provide national standards for both LUC assessments and erosion severity assessments. (In those days of government subsidised erosion control works, many Catchment Boards had the "worst" erosion in NZ within their borders, hence they hoped to get Government money grants - this problem ceased once the survey was completed!). It was also to be used as the basis of regional planning, a concept which was just coming in then.

BACKGROUND

The history of agriculture in New Zealand shows that much of the early development was followed by land degradation, increased sediment transport and flooding. By the mid-1930s environmental damage had become so obvious Government established the Soil Conservation and Rivers Control Council (SCRCC) to control flooding and minimise soil erosion. It was apparent to the Council that the extent and severity of the soil erosion problem had to be assessed and that planning of erosion control programmes and wise land use had to be nationally coordinated.

The initial approach to planning wise land use by the Council was to use soil surveys (3) but these did not provide sufficient physical information. A more holistic approach was chosen so that the mapping units became more management orientated. This involved mapping the physical information and recording data in a clear, user friendly manner that enabled rapid interpretations. The system needed to be sufficiently practical for soil conservators who had no special training on earth sciences or agronomy, to undertake both the mapping and the interpretations and for farmers to understand these interpretations.

Adopting the LUC System

In 1952, the USDA Land Use Capability system was adopted by the SCRCC as the physical basis for soil conservation planning. For the following 35 years Government subsidies for erosion control were based on this land use capability data. Soil and water management plans (introduced in 1956), and catchment control schemes relied on the LUC system for physical data, with the result that by 1970 some 70% of New Zealand had been covered by reconnaissance scale LUC mapping, and 15% at detailed scales. However, both the scale and the standards of these surveys varied. As subsidies were in part based on erosion severity, Catchment Boards sometimes tended to "exaggerate" severity, while a lack of national standards made quality control difficult.

The New Zealand Land Resource Inventory (NZLRI)

To obtain consistent LUC assessments, the SCRCC, in 1970, directed the Land Use Capability mapping team from the Water and Soil Division of the Ministry of Works to prepare a 1:63360 scale LUC map of New Zealand. This survey, carried out between 1973 and 1979 (and now being maintained) became the New Zealand Land Resource Inventory. The objectives of this national survey were twofold, to provide national LUC and erosion assessment standards and to provide a physical land resource data base for regional and national planning. The survey was undertaken by a multi-disciplinary team comprising up to 13 soil conservators and scientists mapping on a regional basis, using a team approach. Each person was responsible for completing a 110000 ha map each 10 to 12 weeks. The maps were published as low cost, two colour worksheets, while documentation of the inventory and regional LUC classifications were published as extended legends and bulletins.

The survey classifies New Zealand into 90000 map units, each containing five sets of inventory information; rock, soil, slope, erosion and vegetation. These map units are grouped into a total of 662 correlated LUC units within 11 regional LUC classifications. The NZLRI data is available as published "worksheets" or as computer based information.

Data analyses of any significant area was a problem. The data is recorded on 320 printed worksheets. This was overcome by storing all data in a GIS. Initially the GIS software (LADEDA) was developed by our own staff, but due to difficulties in maintaining and developing this in an environment of reducing staff numbers, a change was made in 1988 to commercial GIS software. The GIS chosen was a state of the art package, which has both PC and mainframe versions. Updating and upgrading of the NZLRI continues.

The Land Use Capability System

The New Zealand LUC classification system differed from the USDA approach for many years in that it is a ranking of LAND according to its versatility and limitations. The USDA system, by contrast, is a ranking of SOILS according to the same criteria. This major difference reflects a recognition that soil is only one of the environmental parameters controlling sustainable land use. In many cases the soil factor is insignificant when compared with the stability of near surface rock, flooding potential or climate. The LAND approach allows any one or more physical factors which are important in controlling land use, to be emphasised.

The LUC system uses a multi-factor mapping approach. It has two components within each map unit; a physical resource inventory is recorded and from this (together with local knowledge of climate and the effects of past land use), the sustainability for sustained use is assessed. This is the Land Use Capability Assessment.

The Physical Land Resource Inventory

The inventory records five physical factors which are known to be important for planning sustained land use. Each factor is recorded on a dominance basis with each land unit. In New Zealand these factors are; rock, soil, slope, erosion and vegetation. The classifications use criteria that can be identified in the field or from photo interpretation. Each of the inventory factors is briefly discussed in the following section.

The standard inventory code is illustrated in Figure 1.

Rock Type

A rock type classification has been developed specifically to suit the needs of LUC. This classification groups rocks with similar erosion susceptibilities and characteristics, and concentrates on those rocks which directly influence surface morphology and land use. The rocks are described in terms understandable by planners and land managers. As part of this exercise, separate rock type classifications for the North Island (5) and South Island (6) were developed for the NZLRI. These are now being amalgamated into one.

Table 1 illustrates the grouping of the sedimentary mudstone/sandstone sequences. These were subdivided according to grain size, bedding and pattern as these affect morphology, stability and fertility.

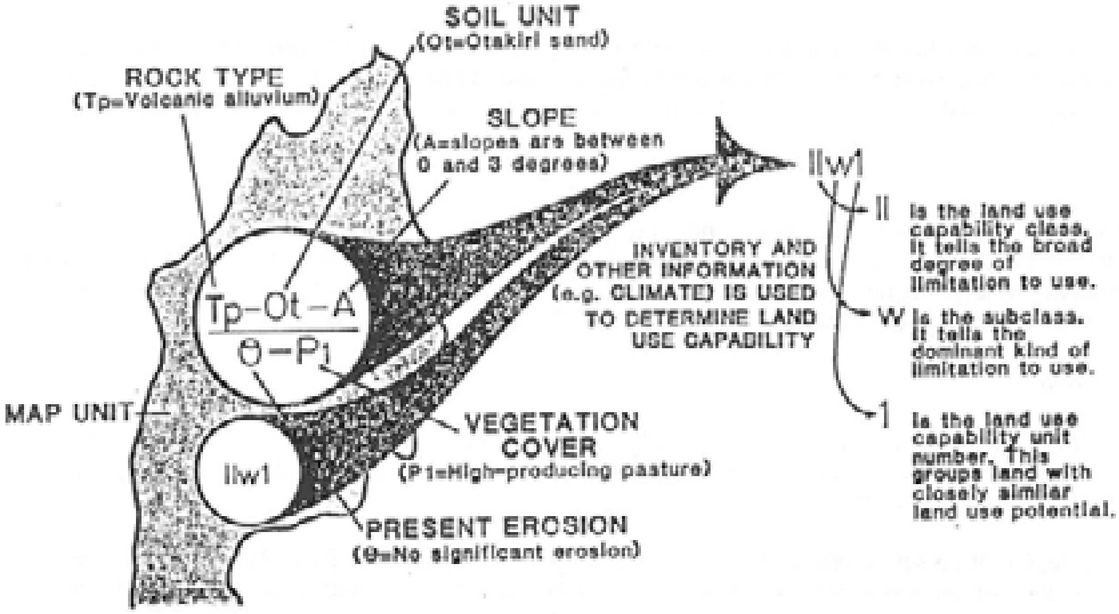


Figure 1. The inventory code and assessed LUC used in the New Zealand Land Use Capability system (adapted from (2)).

Table 1. Rock type groupings for the sedimentary mudstone, sandstone sequences indicating the importance of erosion in the classification.

Mudstones and fine siltstones	Erosion	Erosion Control measures	Land Use Capability class
Mm Massive	Soil slip erosion on colluvium	Space planting	III to VII
Mb Banded	Soil slip on scarp slopes. Earthflow on dip slopes and colluvium	Space and block planting.	VI to VII
Mj Jointed	Deep and shallow earthflow, soil slip, gully	Space and block planting, recontouring, diversion banking, debris dams and pair plantings	VI to VII
Mc Bentonitic	Severe deep earthflow, gully control forestry	All of the above and erosion	VI to VIII
Coarse siltstones and sandstones			
Sm Massive	Sheet Soil slip Tunnel gullies	Oversowing Space planting Planting in tunnel gullies	III to VII
Sb Banded	Sheet Soil Slip	Oversowing Space planting	III to VII

Soil

Where possible soil information is obtained from existing soil surveys with field checking to ensure the information is correct. Where data is not available at the appropriate scale, a physiographic analysis technique is used to re-interpret small scale soil information to the 1:63360 or 1:50000 scale.

Slope

The slope classification (7) groups slopes in degrees into seven classes; A (0-3), B (4-7), C (8-15), D (16-20), E (21-25), F (26-35), G (greater than 35). These groupings are based on broad management criteria. At larger scales the groupings can be further subdivided to include criteria such as aspect, position on slope and exposure.

Erosion

Both type and degree of erosion is recorded. In the NZLRI thirteen types were recognised: soil slip, earth slip, debris avalanche, slump, earthflow, scree, sheet, wind, gully, rill, tunnel gully, streambank as well as deposition.

The degree of present erosion was expressed as a six part ranking (see Table 2). For sheet, wind and scree, the ranking was a visual estimate of the percent bare ground within each map unit (8) affected by that erosion type. There is no measure of rate of soil loss but in large scale LUC mapping the degree of soil profile loss can be recorded.

For fluvial and mass movement erosion types, the degree of present erosion is based on seriousness, taking into account area affected, technical difficulty and cost of repair.

Table 2. Erosion Severity Rankings used in the New Zealand Land Resource Inventory Survey

Erosion Ranking	Surface Erosion (estimated % of bare ground)	Mass movement and Fluvial erosion (Seriousness)
0	<1	none
1	1-10	slight
2	11-20	moderate
3	21-40	severe
4	41-60	very severe
5	>60	extreme

Vegetation

Emphasis is placed on identifying agriculturally important species and associations rather than providing a botanical classification. Usually five vegetation groups are recorded; Grassland, Cropland, Scrubland, Forest and Miscellaneous (9) using codes. At large scales dominant species are often recorded using codes.

The Land Use Capability Assessment

Using the inventory, together with a knowledge of the local climate and effects of past land use, map units are grouped into land use capability units according to their physical limitations and degree of versatility. This LUC classification uses a simple eight class classification illustrated in Table 3.

Table 3. The eight LUC classes are arranged in order of increasing limitation and decreasing versatility.

Class	Cropping Suitability	General Pastoral & Production Forestry Suitability	General Suitability
I	High	High	Multiple use land
II			
III	Medium		
IV	Low		
V	Unsuitable		Pastoral or Forestry land
VI		Medium	
VII		Low	
VIII		Unsuitable	Catchment protection land

increasing limitation to use ↓

decreasing versatility ↓

The land use capability classification has three levels, the LUC class, subclass and the LUC unit (10).

The LUC CLASS identified the overall level of limitation and versatility.

There are eight classes in the New Zealand system which conforms to the international approach. However the emphasis has changed for some of the classes.

The SUBCLASS which identifies the type of physical limitation. Only four subclasses are used;

- e erosion
- w wetness
- s rooting zone limitation (soil)
- c climate

A mapper is required to choose only one subclass.
This approach ensures that only the dominant limitation is recorded.

The LUC UNIT group inventory units that respond similarly to the same management, are adapted to the same kind of crops, pastures, or forest species, have about same potential yield and require the application of the same soil conservation measures.

The LUC unit is the management portion of the classification and is emphasised rather than the subclass.

The degree of detail in the definition of a unit depends on scale. At farm plan scales (e.g. the definition should be sufficient to enable erosion control measures to be planned, located and costed. At smaller scales, such as (1 mile to 1 inch) the definition is broader and this allows the types of soil conservation to be planned but without their exact locations being identified.

The LUC regional classifications are described in Regional LUC Extended Legends (11) and in regional bulletins (12). The extended legends LUC units in order of decreasing capability, but the bulletins group them into suites. Suites combine those LUC units which occur on the same landform system.

In the NZLRI there are 662 separate LUC units covering New Zealand. These provide the basis for adding further interpretive data sets. For instance cooperative exercises with advisory staff from the Ministry of Agriculture and Fisheries and the Ministry of Forestry have built up national data sets for three levels of stock carrying capacity (present average, top farmer and attainable potential) and site index values for Pinus radiata (13).

The use of GIS has allowed more data to be collected for each map unit. It must be emphasised, however, that recording data takes time (and money), so while it is possible to record many items of data it is more appropriate to record only that data which is necessary!

The LUC classification is ranked according to use risk. However, each LUC unit identifies physically similar types of land -which at the mapping scale can be taken as uniform.

The LUC units can therefore be ranked in ways other than that listed in the legend. For instance, in a recreational classification LUC class 8 land may be class 1 land!

In other words, the LUC classification should be looked upon as a grouping of areas which are physically similar.

Updating the NZLRI

With any national data set it is important that it be maintained and developed. We intend maintaining it but, of course, this is dependent on adequate finance being available. Figure 2 shows the percent status of the NZLRI and areas currently being updated.

Updating is at scale 1:50000 which enables the smallest area to be delineated to be 15ha and overall will provide an exceptional planning base.

The data is being supported by classification documentation in the form of bulletins and reports,

The new Division and the Division to come provides opportunities for the NZLRI to be expanded further. Enabling the NZLRI soil names to be connected to the NZ Soil Data Base means soil characteristics (chemical and physical) will be able to be plotted and interpreted nationally.

DOC Applications

I do not intend to list what all the applications are, instead a few examples will be given and then I expect suggested applications to come from the floor.

National Applications

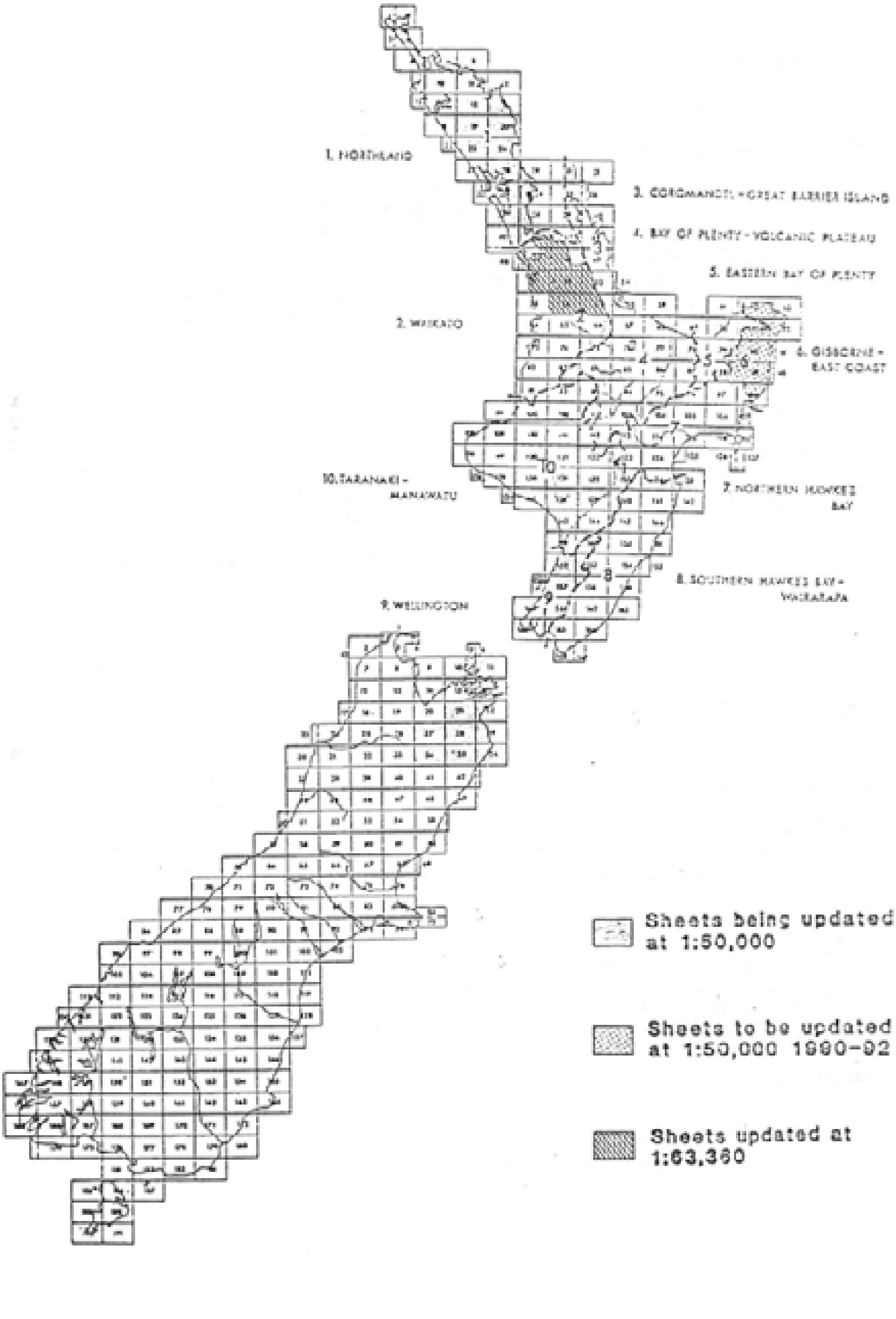
Interrogation of the national data base allows nationally consistent interpretations to be made. These can range from the simple, such as identifying land suited only for protection use to the complex such as developing a sustainable land use model.

To illustrate these:

1. A plot of LUC class VIII land indicates land with physical limitations such that it is only suited to protection purposes.

2. Susceptibility to scrub reversion. By assessing the potential rate of scrub reversion on each of the 660 LUC units in New Zealand a map of the areas most prone to reversion can be prepared.

Figure 2. Percent status of the NZLRI and areas currently being updated



3. By modelling a series of physical factors (such as erosion hazard LUC capability and reversion potential) a plot indicating the highest level of sustainable use could be prepared. This would have significant impacts on generating national land use policies.

Regional Applications

Regional Planning Using the National LUC Data Set

1. Land development options

In this example staff of the Department of Conservation wished to know which land in a particular development block was suitable for sustained grazing. In addition, the Department wished to know areas in indigenous forest, scrub and pasture. The first requirement was obtained by grouping the LUC units in the NZLRI which covered the block according to whether or not they were suited to sustainable grazing. However, as the worksheets had been prepared 9 years before the enquiry, it was necessary to undertake a brief field inspection to update the vegetation section of the inventory before the second part of the enquiry could be actioned. The updated vegetation information was incorporated in the NZLRI data base, allowing both parts of the enquiry to be presented as coloured plots of the development block.

2. Ranking a region according to rabbit hazard

The development of a new rabbit control policy has been underway in New Zealand for some time. During this exercise a regional study to rank areas prone to rabbit infestation in Central Otago was undertaken using the NZLRI as the physical data base for interpretations. To obtain a hazard ranking the study team grouped map units according to the named soil. Each soil was then ranked according to an assessed rabbit hazard as allocated by the team, based on field studies. This ranking was then modified by the type of vegetation cover. By modelling this combination in the NZLRI data set, plots of relative rabbit hazard were able to be generated.

3. Identification of potential reserves

Reserves should represent type areas of land within a region. The NZLRI can provide the initial sieve for this assessment of type. From this information further selection can be made.

PNA Programme

The NZLRI is the ideal base upon which to store the PNA data. Inputting PNA information would allow the basic relationships to be established between the potential priority areas and their extent in the Ecological District. The NZLRI data is generally not appropriate to identify the quality of the vegetation in these areas but it would identify the extent e.g. the area of snow tussock in an ecological region allowing the representativeness of the priority areas of say undisturbed snow tussock assessed.

Ecological Districts

Ecological Districts have been stored in the NZLRI data set. However, I have no indication of their having been used. If they have not it is a great pity as the NZLRI would be very effective in characterising an Ecological District in terms of landforms, general vegetation patterns etc. This for instance, would provide an effective starting point for further studies such as selecting PNA areas.

Additional data sets

Any data sets that DOC has which can be related to the NZ Grid can be added to the NZLRI and associated data sets. Each additional data set increases the versatility of the overall system. Join us and reap the benefits of building NZ's major physical resource data base.

SUMMARY

This paper provides a brief overview of the NZLRI and provides examples of how this data set can be of use to DOC staff. Our experience, however, is that the uses are dependent on what we need, with the result we keep finding new uses for the data. If you USE the NZLRI you will have the same experience -one of finding more and more uses for it.

REFERENCES

Klingebiel, A.A., Montgomery, P.M. 1961. Agricultural Handbook 210. Soil Conservation Service, US Department of Agriculture.

National Water and Soil Conservation Organisation. 1979. Wellington, New Zealand. 78p.

Grange, L.I., Gibbs, H.S. 1947. New Zealand Soil Bureau Bulletin 1.

Jessen, M.R. 1987. Water and Soil Misc. Publ. No. 15. 74 p.

Crippen, T.F., Eyles, G.O. 1985. Water and Soil Misc. Publ. No. 72.

Lynn, I.H. 1985. Water and Soil Misc. Publ. No. 73. 66 p.

Fletcher, J.R. 1988. Soil Survey and Land Evaluation V8 (3), 131-140.

Eyles, G.O. 1985. Water and Soil Misc. Publ. No. 85. 61 p.

Hunter, G.G., Blaschke, P.M. 1986. Water and Soil Misc. Publ. No. 101. 92 p.

Soil Conservation and Rivers Control Council, 1971. Wellington, New Zealand.

Noble, K.E. 1979. National Water and Soil Conservation Organisation, Wellington, New Zealand.

Fletcher, J.R. 1987. Water and Soil Misc. Publ. No. 110. 227 p.

APPENDIX VII DSIR DATABASES

Diana Kelly
DSIR Library Centre
Private Bag 13
PETONE

DATABASES ON DSIR DATABASE NETWORK

Bibliographic suite	SIRIS (or SCITEC) - DSIR Information System LIBRA DSIR Online Catalogue ADOS Book ordering ASKS Serials management
Other databases:	Information management - for files Interloan of books and journal articles Circulation of books and serials to staff Issues to borrowers
	Botany Division herbarium - CHIRP Entomology Division insect database - BUGS

SOFTWARE - use BASIS SOFTWARE on VAX NETWORK

20 libraries in the network (from Auckland to Christchurch), all contributing.

All data checked by mainly automated processes at DSIR Library Centre.

LIBRA DATABASE - DSIR online library catalogue

Bibliographic database only

Size: 57,800 records

Coverage: All DSIR Libraries contribute, 1982 onwards

Older material in some disciplines

Types of records: (see Fig. 1)

Fields (see Fig. 2)

Formats can be individually tailored

Holdings all 20 DSIR libraries contribute.
Each has a library symbol on record

Who will search it for you and get you a copy?

Any DSIR library - see Appendix A for addresses of DSIR divisions
Information can be faxed or interloaned to you.

SIRIS DATABASE - DSIR

Information System

Bibliographic items only

Size: 45,000 records (see Fig. 3)

Coverage: DSIR publications, material published by DSIR scientists anywhere, or considered to be of use to them.

Either published or unpublished.

Virtually comprehensive for DSIR from 1980 onwards (Geology exception)

A considerable number of pre-1980 biological and conservation records are on the database, mainly due to the efforts of Ecology Division.

Fields: (see Fig. 4 and 5)

Searching DSIR database:

Who will search these databases for you, provide printouts of searches, and get you copies of references that look useful?

SIRIS ONLY:

1. Your Head Office Librarian (David Bowie) through National Library database SciTec (another version of SIRIS).

FOR A COMPEHENSIVE SEARCH:

2. Contact your nearest DSIR Division or

3. DSIR Library Centre, Box 13, Petone.
   Phone (04) 690362 (direct line).

Charges: Database searches are charged on a rate of $70 per hour. Most searches are quick – ONLY A MINUTE OR SO.

An example of SIRIS search is appended (Appendix B).
An example of LIBRA search is appended (Appendix C).

References useful to the Department of Conservation:

AGRICULTURE
- Weeds
- Forestry protection
- Animal protection
- Fisheries protection ....600 records

BIOLOGY
- Ecology
- Populations
- Biological conservation ....5340 records

EARTH SCIENCES
- Land classification and land use
- Environmental conservation
- Environmental pollution
- Mapping and surveying
- Regional geology
- Geomorphology....................4200 records

GENERAL
- History and Biography
- Science policy
- Science management .... 1817 records

Figure 1. A typical record:

```
Item 6

CNTRE NO.:   XEC8800124
INDEXING:    UD
LEVEL:       A
STATS:       N8808 S8809LS
ACC LISTS:   MSWI8804
LIBRARIES:   ^^^^^^^^^MSWI ^
             ^^^^^^^^^^
             ^
TITLE:       Planting trials for the revegetation of Mana Island
             / Susan M. Timmins, Ian A.E. Atkinson and Colin Ogle
DATE:        1988
COLLATION:   iii, 11 p. ; 30 cm
SERIES:      Science and Research internal report ; 3
NOTES:       "This is an unpublished report ... Permission for use of any of its
             contents in print must be obtained from the Director, Science and
             Research"
AUTHOR:      Timmins, Susan M.
             Atkinson, Ian A.E.
             Ogle, Colin
CORP:        New Zealand. Department of Conservation. Science and Research
             Directorate
UDC:         581.9 (931.268.81:581.524.3
DESCR:       MANA ISLAND
             VEGETATION
             REVEGETATION
WSWI:        SERIAL New Zealand. Department of Conservation
STACK:       PLANTING TRIALS FOR THE REVEGETATION OF MANA ISLAND
DATABASE:    LIBRA
LAST PRSN:   LSC7
LAST UPDT:   19880909
BASIS NO:    101367
```

Figure 2. Fields:

SIRIS File

NAME	NUMBER	DISPLAY	SEARCH	TYPE OF INDEX
AB	106	ABST:	AB=	KEYWORD
AN	1	CNTRE NO.:	AN=	EACH CODE
AU	36	AUTHOR:	AU=	KEYWORD, LINKS
BL	16	LEVEL:	BL=	EACH CODE
BN	700	BASIS NO:	BN=	EACH NUMBER
CC	4	COUNTRY:	CC=	EACH CODE
CL	55	COLLATION:		
CN	80	CONFRENCE:	CN=	KEYWORD, LINKS
CO	79	CORP:	CO=	KEYWORD, LINKS
COUNT	725	No. UPDTS:	COUNT=	TOTAL
DA	52	DATE:	DA=	EACH DATE
DB	730	FILE:	DB=	EACH FILENAME
DE	94	DESCR:	DE=	EACH DESCR.
			DEK=	KEYWORD
EN	42	EDITION:		
FM	10	FORMAT:	FM=	EACH CODE
IB	31	ISBN:	IB=	EACH ISBN
ID	710	LAST PRSN:	ID=	LAST USERID
IS	28	ISSN:	IS=	EACH ISSN
MJ	95	MAJ DESCR:	MJ=	EACH DESCR.
NT	64	NOTES:		
OC	23	OWNER CDE:	OC=	EACH CODE
PL	46	PLACE:		
PU	49	PUBLISHER:	PU=	KEYWORD
RN	3	NZSA No.:	RN=	EACH NUMBER
RP	61	REPORT:	RP=	EACH CODE
RR	2	RELATD RN:		
RS	740	SECURITY:	RS=	EACH CODE
SC	93	SUBJ CATG:	SC=	EACH CODE
SE	58	SERIES:	SE=	KEYWORD, LINKS
SOAU	150	SRCE AUTH:	SOAU=	KEYWORD, LINKS
SOCL	158	SRCE COLL:		
SOCO	153	SRCE CORP:	SOCO=	KEYWORD, LINKS
SOEN	159	SRCE EDTN:		
SONT	177	SRCE NOTE:		
SOPL	162	SRCE PLCE:		
SOPU	165	SRCE PUBL:	SOPU=	KEYWORD
SORP	174	SRCE RPRT:	SORP=	EACH CODE
SOSE	171	SRCE SER:	SOSE=	KEYWORD, LINKS
SOTI	156	SRCE TITL:	SOTI=	KEYWORD
TIM	34	TITLE:	TIM=	KEYWORD
TP	7	BIB TYPE:	TP=	EACH CODE
UPDAT	720	LAST UPDT:	UPDAT=	LAST DATE

Figure 3. If you break the SIRIS database into subject categories:

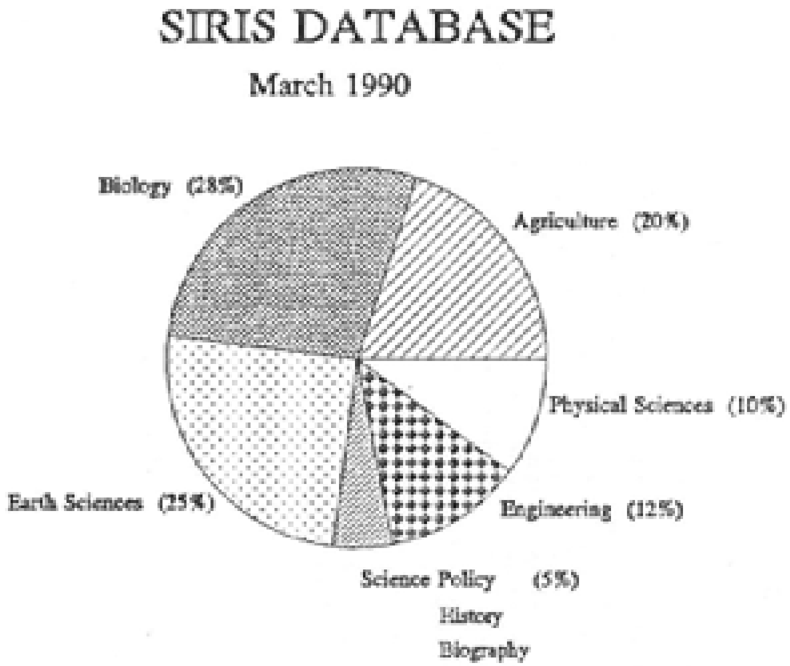


Figure 4. Fields:

NAME	NUMBER	DISPLAY	SEARCH	TYPE OF INDEX
AB	106	ABST:	AB=	KEYWORD
AC	22	ACC LISTS:	AC=	EACH CODE
AN	1	CNTRE NO.:	AN=	EACH CODE
AU	76	AUTHOR:	AU=	KEYWORD, LINKS
BL	16	LEVEL:	BL=	EACH CODE
BN	700	BASIS NO:	BN=	EACH NUMBER
CC	4	COUNTRY:	CC=	EACH CODE
CK	112	CODE Nos.:		
CL	55	COLLATION:		
CO	79	CORP:	CO=	KEYWORD, LINKS
COUNT	725	No. UPDTS:	COUNT=	TOTAL
DA	52	DATE:	DA=	EACH DATE
DB	730	FILE:	DB=	EACH FILENAME
DE	94	DESCR:	DE=	EACH DESCR.
			DEK=	KEYWORD
EN	42	EDITION:		
FM	10	FORMAT:	FM=	EACH CODE
FR	25	FREQ:	FR=	EACH CODE
FT	67	FORMER:		
IB	31	ISBN:	IB=	EACH ISBN
ID	710	LAST PRSN:	ID=	LAST USERID
IN	13	INDEXING:	IN=	EACH CODE
IS	28	ISSN:	IS=	EACH ISSN
LC	97	LC:	LC=	EACH (SUB)DIV.
LO	200	DBN HOLDS:		
LT	70	LATER:		
ME	88	M/E:		
NO	43	NUMBERING:		
NT	64	NOTES:		
OC	23	OWNER CDE:	OC=	EACH CODE
PL	46	PLACE:		
PREF	802	PREF:		
PU	49	PUBLISHER:	PU=	KEYWORD
RA	103	SEE ALSO:		
RF	100	SEE:		
RID	109	RID:	RID=	EACH CODE
RP	61	REPORT:	RP=	EACH CODE
RS	740	SECURITY:	RS=	EACH CODE
SE	58	SERIES:	SE=	KEYWORD, LINKS
SOCL	168	SRCE COLL:		
SOEN	159	SRCE EDTN:		
SOPL	162	SRCE PLCE:		
SOPU	165	SRCE PUBL:	SOPU=	KEYWORD
SOSE	171	SRCE SER:	SOSE=	KEYWORD, LINKS
SOTI	155	SRCE TITL:	SOTI=	KEYWORD
SS	19	STATS:	SS=	EACH CODE
STACK	800	STACK:		
TEMP	900	DATA IS	TEMP=	"UNCHECKED"
TIM	34	TITLE:	TIM=	KEYWORD
TIS	37	TITLE:	TIS=	KEYWORD
TL	40			
TP	7	BIB TYPE:	TP=	EACH CODE
UD	91	UDC:	UD=	EACH NUMBER
UDCCODE	800	UDCCODE:		
UK	600	UDKEYWRD:	UK=	KEYWORD
UPDAT	720	LAST UPDT:	UPDAT=	LAST DATE
UR	605	RELATOR:		
US	603	SCOPE:		

Figure 5. A typical record:

```
Item 45

CNTRE NO.:   AS9000042
COUNTRY:     NZ
BIB TYPE:    J
LEVEL:       AS
ISSN:        0067-0464
BOOK ORDR:   UPG AS 9001
TITLE:       The excavation of sites R11/887, R11/888 and R11/899, Tamaki,
             Auckland
DATE:        1989
COLLATION:   p. 1-24
NOTES:       39 refs; 19 figs; 2 tables
AUTHOR:      Foster, R. (Department of Conservation. Auckland)
             Sewell, B. (Department of Conservation. Auckland)
SRCE TITL:   Records of the Auckland Institute and Museum. 26
SUBJ CATG:   B47
DESCR:       TAMAKI
             AUCKLAND
             ARTEFACTS
             EXCAVATIONS
             TE APUNGA O TAINUI
             ARCHEOLOGY
             NEW ZEALAND MAORI
             CROP STORAGE
             ARCHEOLOGICAL SITES
             NZMS260R11
MAJ DESCR:   AUCKLAND
             EXCAVATIONS
             ARCHEOLOGY
ABST:        The excavation of adjacent sites R11/887, 888 and 899 in Tamaki are
             described. Together the sites formed part of an extensive open
             settlement occupied in the sixteenth century. The rescue nature of
             the excavations provided a valuable opportunity to examine a
             greater area of an open settlement than is usually possible.
             Specific activities were identified at R11/887, including cooking,
             living, storage and stoneworking areas. Site R11/888 has been
             interpreted as a specialist site with huts for storage of equipment
             and a living floor of laid shells. At R11/899 there were several
             superimposed houses and a group of deep storage pits. The layout of
             the sites presented an impressive illustration of an unfortified
             hamlet, probably associated with occupation of the pa at Te Apunga
             o Tainui. (auths)
DATABASE:    SIRIS
SECURITY:    0
LAST UPDT:   19900201
BASIS NO:    168576
```

APPENDIX A DSIR ADDRESSES

DSIR Antarctic
Union Centre Building
214a Oxford Terrace
P.O. Box 13-247, Christchurch
Telephone (03) 791-540
Facsimile (03) 791-545
Manager: Mr Hugh Logan

DSIR ASEAN Centre
214 Pandan Loop, Singapore 0512
Telephone (65) 777-8868
Facsimile (65) 778-4400
Science Liaison Officer
Mr Norman Lodge

DSIR Chemistry
Gracefield Road, Lower Hutt
Private Bag, Petone
Telephone (04) 666-919
Facsimile (04) 694-500
Director: Dr Gordon Leary

DSIR Crop Research
Ellesmere Junction Road, Lincoln
Private Bag, Christchurch
Telephone (03) 252-511
Facsimile (03) 252-074
Director: Dr Mike Dunbier

DSIR Fruit and Trees
Fitzherbert West
Private Bag, Palmerston North
Telephone (063) 68-019
Facsimile (063) 61-130
Director: Dr James McWha

DSIR Geology and Geophysics
State Insurance Building
Andrews Avenue
P.O. Box 30-368, Lower Hutt
Telephone (04) 699-059
Facsimile (04) 695-016
Director: Dr Ian Speden

DSIR Grasslands
Fitzherbert West
Private Bag, Palmerston North
Telephone (063) 68-019
Facsimile (063) 61-130
Director: Mr John Lancashire

DSIR Industrial Development
Brooke House
24 Balfour Road, Parnell
P.O. Box 2225, Auckland
Telephone (09) 303-4116
Facsimile (09) 370-618
Director: Dr Ashley Wilson

DSIR Land Resources
Eastern Hutt Road, Taita
Private Bag, Lower Hutt
Telephone (04) 673-119
Facsimile (04) 673-114
Director: Dr Derek Milne

DSIR Library Centre
Gracefield Road, Lower Hutt
Private Bag 13, Petone
Telephone (04) 666-919
Facsimile (04) 694-500
Chief Librarian: Miss Monica Hissink

DSIR Marine and Freshwater
Greta Point, 310 Evans Bay Road
Private Bag, Kilbirnie, Wellington
Telephone (04) 861-189
Facsimile (04) 690-117
Director: Dr Bill Robinson

DSIR Physical Sciences
Gracefield Road
P.O. Box 31-313, Lower Hutt
Telephone (04) 666-919
Facsimile (04) 690-117
Director: Dr Bill Robinson

DSIR Plant Protection
Mt Albert Research Centre
120 Mt Albert Road
Private Bag, Auckland
Telephone (09) 893-660
Facsimile (09) 863-330
Director: Mr John Longworth

DSIR Publishing
16 Kent Terrace
P.O. Box 9741, Wellington
Telephone (04) 858-939
Facsimile (04) 850-631
Manager: Dr Norman Hawcroft

DSIR Social Science
27 Creyke Road
P.O. Box 29-181, Christchurch
Telephone (03) 351-6019
Facsimile (03) 351-9923
Leader: Dr Roberta Hill

APPENDIX B SEARCH OF SIRIS DATABASE

SEARCH of SIRIS Database

I A E Atkinson Rats and islands

1. Hamilton, W.M. (DSIR. Wellington) ; Atkinson, I.A. (DSIR. Wellington)
Vegetation
In: Little Barrier Island (Hauturu). - Wellington
: DSIR, 1961. - 198 p. - (DSIR bulletin ; 137). -
P. 87-121

2. Atkinson, I.A.E. (DSIR, Botany Division. Taita)
The flora, and soils of Middle and Green Islands, Mercury Islands
Group
In: New Zealand journal of botany. 2(4), 1964:
385-402

3. Atkinson, I.A.E. (DSIR, Botany Division. Lincoln) ; Campbell, D.J.
(DSIR, Ecology Division. Lower Hutt)
Habitat factors affecting saddlebacks on Hen Island. - (ECOL-Pub--141)
In: Proceedings of the New Zealand Ecological
Society. 13, 1966: 35-40

4. Atkinson, I.A.E. (DSIR, Botany Division. Taita)
An ecological reconnaissance of Coppermine Island, Hen and Chicken
Group
In: New Zealand journal of botany. 6(3), 1968:
285-294

5. Atkinson, I.A.E. (DSIR, Botany Division. Taita)
Vegetation and flora of Sail Rock, Hen and Chicken Islands
In: New Zealand journal of botany. 10(2), 1972:
545-558

6. Atkinson, I.A.E. (DSIR, Botany Division. Taita)
Protection and use of the islands in Hauraki Gulf Maritime Park
In: Proceedings of the New Zealand Ecological
Society. 20, 1973: 103-114

7. Bell, B.D. (Department of Internal Affairs, Wildlife Service.
Wellington) ; Atkinson, I.A.E. (DSIR, Botany Division. Taita)
Report on Rangatira Swamp, Kapiti Island. - 1975. - 1 p.. -
(BD-VR--084)

8. Atkinson, I.A.E. (DSIR, Botany Division. Taita)
Vegetation surveillance for Kapiti opossum study, August 1975 -
February 1977. - 1977. - 5 p.. - (BD-VR--166)

9. Atkinson, I.A.E. (DSIR, Botany Division. Taita)
Kapiti opossum study: vegetation surveillance in August 1978. - 1978.
- 11 p.. - (BD-VR--240)

10. Gibb, J.A. (DSIR, Botany Division. Taita) ; Atkinson, I.A.E. (DSIR,
Botany Division. Taita)
Precautions to be taken against rats reaching Mana Island. - 1978. -
2 p.. - (BD-VR--254)

11. Atkinson, I.A.E. (DSIR, Botany Division. Taita)
Kapiti opossum study: surveillance of northern rata in August 1979. -
1979. - 7 p.. - (BD-VR--291)

APPENDIX C SEARCH OF LIBRA DATABASE

SEARCH ON RAT* ISLAND* ON LIBRA DATABASE

1.
Title: Asia-Pacific interchange, June 17-27 1968 : proceedings : rodents as factors in disease and economic cost
Corporate Name: Asia - Pacific Interchange (1968 : Honolulu, Hawaii); Institute for Technical Interchange (East-West Center)
Place: Honolulu
Publisher: Center for Cultural and Technical Interchange between East & West
Date: [c1968]
Collation: xii, 285 p. : ill. ; 28 cm
Location(s): WSWI 599.32:632.693.2 ASI ^ C.1 LH, C.2 NS
Basis No.: 132704

2.
Title: Fiordland
Corporate Name: New Zealand National Parks Centennial Commission; Television New Zealand; Broadcasting Corporation of New Zealand
Publisher: Broadcasting Corporation of New Zealand
Date: 1987
Collation: 1 videocassette (60 mins) : sd., col., 23.39 mm/sec ; 12 mm
Series: Journeys in National Parks
Location(s): WSWI AUDIOVISUAL NO. 7
Basis No.: 158408

3.
Title: Hauraki Gulf
Corporate Name: New Zealand National Parks Centennial Commission; Television New Zealand; Broadcasting Corporation of New Zealand
Publisher: Broadcasting Corporation of New Zealand
Date: 1987
Collation: 1 videocassette (60 mins) : sd., col., 23.39 mm/sec ; 12 mm
Series: Journeys in National Parks
Location(s): WSWI AUDIOVISUAL NO. 7
Basis No.: 158410

6.
Personal Name: Dick, A.M.P. (Andrew Mark Philip)
Title: Rats on Kapiti Island, New Zealand : coexistence and diet of Rattus norvegicus Berkenhout and Rattus exulans Peale
Place: Palmerston North
Publisher: Massey University
Date: 1985
Collation: 2 microfiches ; 10 x 15 cm
Location(s): WSWI MICROFICHE 599.323.4(931.41) DIC
Basis No.: 135190

7.
Personal Name: Hay, J.R.
Title: The North Island kokako (Callaeas cinerea wilsoni) on northern Great Barrier Island
Source: Journal of the Royal Society of New Zealand. 15(3), 1985: 291-293
Location(s): WSWI REPRINTS Hay, J.R.; Douglas, M.E.; Bellingham, P. 1985
Basis No.: 23455

APPENDIX VIII LESSONS FROM AN ARCHAEOLOGICAL SITE COMPUTER FILE

Brian Sheppard
Science and Research Division
Department of Conservation
P.O. Box 10-420
Wellington

(A paper which appeared in New Zealand Geographer 46, 1, 1990, 40-42)

A computer-based Geographical Information System (GIS) is often perceived as a tool that will give a land manager instant access to the facts needed to make an informed management decision. In many cases it will suggest totally misleading information which a user will be tempted to accept without question. It is the author's strong belief that a well planned GIS can be an invaluable research and management tool, but only when the significance and limitations of the component layers of data are properly understood. Many sets of data can only be used reliably by the close involvement of people with specialist knowledge. Many users, however, will be tempted by the speed of access and apparent similicity of GIS overlays to act on the basis of their perceived analysis rather than incurring further costs and consequent delays by involving several specialists. This paper describes some common data limitations that contribute to interpretational problems but which are poorly understood. It distinguishes locational limitations, which should be accommodated within the design of a GIS, from those of data significance, which must be given expert interpretation.

HISTORICAL BACKGROUND

New Zealand has the rare distinction of operating a single archaeological site recording scheme, in which its amateur and professional archaeologists record their observations (Daniels, 1979). The scheme was conceived, and is still operated, by the New Zealand Archaeological Association (NZAA) and is highly regarded as a research and management tool by archaeologists and planners alike. Although started nearly 30 years ago, with no computer application in sight, its structure is well suited to computer handling. It is a map-based system in which archaeological site records are accessed by a separate numbering sequence for each topographical (NZMS1 or NZMS260) map sheet. Each site is located by grid reference and recorded in a standard format that briefly describes its location and physical form.

The site recording scheme received a boost in 1976 with the passing of the Historic Places Amendment Act which, for the first time, applied a statutory protection against unnecessary damage to any archaeological site. The New Zealand Historic Places Trust, as the body responsible for administering the Act, saw that it would be the largest user of the site records and so contributed both to the gathering of information and to the development of a computer system to analyse the data held in the records.

During the reorganisation of Government departments in 1987, the department providing staff and administrative support for the Trust and its Act was changed from Internal Affairs to the new Department of Conservation. It is this Department that now maintains the central paper file of, and computer index to, the Association's archaeological site records.

The resources necessary to place a large data file on a computer normally require careful justification, underpinned by a list of uses to which the new system could be put. This list would probably include tantalising new applications for the data, supporting a persuasive argument for its purchase. The thinking behind this justification is crucial, not only to assess the benefits of capital expenditure and operating costs, but also the credibility of the agencies associated with its eventual operation. Hindsight makes many of the lessons learned seem obvious, but they may not be so apparent at the early planning stages, as the following examples show.

The first justification that might be given for setting up a computer file could be that it would replace cabinets of paper records. While this could happen, it may not be practicable. For instance, a large and constantly growing paper file with lengthy records could create insurmountable problems for data input and for the complexity and speed of subsequent analysis. Replacement of paper records was an initial aim of the Trust's computer file, but it was one that had to be radically revised because new data was being acquired faster than it could be entered, and because the records also contained diagrams. Even if site recording had stopped until the records had been entered, a computer file could not replace the paper version because diagrams could not be held on the type of system available then, or that might be affordable in the foreseeable future.

The problems were eventually overcome by identifying the data essential to the most useful analyses and then using the computer to provide a sophisticated cross index to, and summaries of, the paper records. This allowed the task to be reduced to a manageable level, and for the system to be made operational within a realistic period. What had been lacking in the original planning was the experience needed to recognise the implications of goals set at the design stage.

PROBLEMS OF MAP-BASED INFORMATION SYSTEMS

The meaning of a grid reference and the limitations of a map's precision are fundamental to much of the following discussion but are widely misunderstood. To summarise the main points that will be made below:

1. A grid reference (a) is to a square on the map grid and must identify the size of that square; and (b) identifies the southwest corner of that square, not its centre.

2. The most precise location obtainable from an NZMS1 map sheet is that of a 100 yard grid square (100 metre grid square from an NZMS260).

Map-based information systems allow field locations to be revisited or compared at a later date. While the precision of a location should be obvious to a field recorder, it may require explanation for another user.

A named location needs the least explanation (particularly when expressed in such terms as land title, street address or territorial local authority) because the boundaries are unambiguously defined elsewhere. The location might alternatively be recorded as an area of land defined by the positions of its corners, thereby representing the area as a polygon. Alternatively, the same location might be recorded as being within, or surrounding, a specified map grid square, the location of which is defined by its grid reference.

Locations defined by polygons or grid reference present more of a problem than named locations because the user has to be given an explanation, either of their precision or of the method by which they were derived. The reason for this is most easily understood by reference to examples, first of a very precise method and then one of more typical precision. If measured by accurate survey, a location might be defined to within a matter of millimetres.

More frequently, however, it will be derived by reference to the grid of a suitable map. It is obvious that large scale maps will allow greater precision than smaller scale ones, but the limitations of that precision have to be known, both to the

---

Brian Sheppard is Senior Archaeologist in the Science and Research Division, Department of Conservation.

New Zealand Geographer 46, 1, 1990, 40-42.

recorder and to subsequent users. These limitations are very important but not always appreciated. They can be considered here by examining the significance of a grid reference, the limitations of its precision and the effect of transferring data from one map edition to another.

Map grid

The map grids most commonly used for field recording in New Zealand are those on which the inch to the mile NZMS1 and its metric equivalent NZMS260 series are based. New Zealand was mapped for the NZMS1 series in two parts, with one grid placed over the North Island and the other over the South Island. They were of the same Transverse Mercator projection but skewed slightly in order to accommodate the shape of New Zealand. They are known as the North Island and South Island Map Grids. By contrast, the later NZMS260 (metric) map series uses a single grid, the New Zealand Map Grid, to cover the whole country.

The grid lines, in either map series, are identified by numbers printed in each margin. In the corners of each map those numbers can be seen as part of longer references, indicating distances east or north of the grid's 'false' origin, at its southwest corner. The distances are measured in yards for the North and South Island Map Grids and metres for the New Zealand Map Grid. They provide the coordinates, expressed as an 'easting' and 'northing', by which any point on the grid can be unambiguously defined.

Grid reference

A grid reference describes the location of a square on the map grid, by specifying the coordinates of its *southwest* corner. The small squares printed on the maps considered here have sides equivalent to 1,000 yards (NZMS1) or 1,000 metres (NZMS260). (Thicker grid lines are also included, for ease of identifying the 10,000 yard or 10,000 metre grid squares). Smaller grid squares can be used, by subdividing the sides of a 1,000 yard or metre square into ten equal parts. These represent 100 yard or metre squares and, for reasons described below, are the limit of resolution that the maps allow.

The normal presentation of a grid reference, described on each map sheet, is a concise code. It begins with the number of the map sheet, to identify the part of the map grid being described, and omits the leading digit (NZMS1) or two digits (NZMS260) of each coordinate, shown as the smaller numbers in the map corners. These are represented by the small digits $_{26}$ and $_{59}$ in the examples below. In this convention, the remaining digits identify both the location and size of the grid square being referenced. To reference a 1,000 yard or metre square, only the two digits printed in either margin are used. For a 100 yard or metre square, the third, measured, digits are added. In either case, the coordinates are those of the western and southern sides of the grid square. Examples are:

REFERENCE	SQUARE	SW CORNER AT	
R27 5988	1 000m	$E_{26}59_{000}$	$N_{59}88_{000}$
R27 591887	100m	$E_{26}591_{00}$	$N_{59}88_{700}$
R27 590880	100m	$E_{26}590_{00}$	$N_{59}88_{000}$

While the abbreviated reference is very convenient for manual use, it does not easily lend itself to computer searches of multiple map sheets. For this type of application, it is much simpler to search references to the whole grid. In other words, the leading digits of the easting and northing have to be included.

Limitation of precision

A field recorder derives a grid reference from a map by comparing the surrounding landscape features with those shown on the map. While a photographic enlargement of the map would allow an apparently finer resolution of the detail, it would be beyond the accuracy to which the features had been drawn. This is obviously true where they are shown schematically. A road, for example, that may be 20 metres wide would be represented by a thick line, the scale width of which might be 60 metres. This means that an observation located with respect to the edge of such a feature would be misplaced by the same distance. To compound the problem, even if the landscape features were precisely positioned and drawn, it would be difficult for a field recorder to locate observations with any greater certainty in areas where little landscape detail was represented.

These factors limit the certainty of a location, derived from an NZMS1 or NZMS260 map to that of a 100 yard or 100 metre grid square which, if drawn on the map, would have sides of 1.6mm (NZMS1) or 2.0mm (NZMS260).

Transfer of data between maps

Information recorded on one map edition or version will inevitably need to be transferred to another at a later date. Consider first the transfer to a map of similar scale. Some early editions of NZMS1 maps were prepared in great haste and include landscape detail only sketched from plane table surveys. In some cases, later editions of the same map repeated that detail while in others it was redrawn. For this reason, recording schemes require the map edition to be noted with a grid reference.

A grid reference originally derived from a map with incorrect landscape detail, would be perfectly usable on that map but would indicate another location on a corrected version. Experience shows that this type of error is usually of acceptable magnitude. Regular use of the large file of over 40,000 archaeological site records has shown that converted grid references are usually to the correct, or adjacent, 100 metre square. But even this can be enough to suggest a location on the wrong side of a road or river. It has to be noted, however, that landscape detail has occasionally been wrongly positioned by a much greater distance. Comparison of the NZMS1 N168-269 (Palliser) sheet with its NZMS260 equivalent shows a shift in landscape detail, presumably erroneously positioned on the NZMS1 sheet, by some 400 yards.

Modern mapping techniques offer the user of NZMS260 maps benefits of greater reliability and accuracy than was available in the NZMS1 series, but their slightly larger scale (27 percent) does emphasise the inadequacies of some grid references transformed from references to earlier maps.

The relationship of equivalent positions on the North or South Island Map Grid to the New Zealand Map Grid is not a simple one. Apart from the different alignments of the three grids, the New Zealand Map Grid is drawn on a different conformal projection from the Transverse Mercator used for the North and South Island (Yard) Grids (Lands and Survey, 1973). The mathematical transformation is expressed by a formula that uses the complex arithmetic of imaginary numbers, one component of which is the square root of minus one. These transformations are capable of an accuracy of location to a matter of centimetres. Other formulae have been derived, for the New Zealand Historic Places Trust, that present the relationship in terms of polynomial equations. They were developed for use with the NZAA files in order to use program languages and equipment that could not handle complex numbers. The resultant programs were still capable of transforming easting and northing coordinates to an accuracy of ±2 metres. Fortunately, this is well within the limitations previously described and so has no noticable effect on the result's precision.

At this point, it may be helpful to consider the process of calculating equivalent grid references. It effectively has three stages, the first taking place when a field recorder determines a grid reference from a map.

Stage 1: Field recording

By convention of map use, a field recorder gives a grid reference to the southwest corner of the grid square in which the observation is reported. This is expressed by rounding down the coordinates of a point on the map (perceived to be the centre of the feature being reported) to the nearest 100 or 1,000 yards

Stage 2: Transformation of coordinates

The mathematical relationship of equivalent points on the or metres, as appropriate for the grid square being defined.

North or South Island Map Grids to the New Zealand Map Grid is well known (Lands and Survey, 1973) and forms the basis for calculating equivalent grid references. These calculations, however, only provide a transformation of the coordinates of an equivalent point on the respective grids. In the conversion of grid references, that point is the southwest corner of the grid square being defined.

Stage 3: Conversion to an equivalent grid reference

The equivalent grid reference to that defined by the field recorder is found first by identifying the equivalent size of grid square being referenced (100 metres for 100 yards). The transformed coordinate is then expressed as an equivalent grid reference by rounding up, or down, the calculated easting and northing to the nearest 100 or 1,000 yards or metres, as appropriate to the size of grid square to be defined.

With that information in mind, it is possible to consider data transfer between maps of different scale, or with locations derived from other means, such as ground survey measurement. Grid references derived from large scale maps with the same map grid, can legitimately define smaller grid squares but the converse is not true. For example, if an observation recorded on an NZMS260 sheet is transferred to a 1:10,000 cadastral sheet, it would still have to be represented as a 100 metre square, even though that square would have 10mm sides.

Finally, the limit of precision for any comparison of locations is determined by the least precise component. In other words, if a file contains a mixture of recordings located to precisions of ten and 100 metre grid squares, its resolution would be limited to that of a 100 metre grid square.

TOWARDS A SOLUTION

A knowledge of locational precision is essential in the interpretation of geographically defined data but there are two other fundamental aspects that must be considered. One relates to the significance of the data and the other to the significance of the observations represented. In other words, the data being depicted does not, in itself, explain what is being portrayed: for instance, whether one plotted grid square represents one or many observations and, conversely, whether the voids between recorded locations still await examination or have none of the attributes being examined. Neither does it indicate, without description, the comparative importance that should be placed on the sites or occurrences that are being analysed. These ambiguities demonstrate that a file is one of *uninterpreted data*.

The limitations described above are common to most applications of a geographic information system, regardless of whether it is operated from a computer or from paper maps and files. Computer assistance can lead to a false sense of confidence in the data by providing a simple and clear representation of the data, particularly when overlaid by relevant information from other databases. Because the usual application of GIS programs is the recording of urban features, such as buildings, roads, pipes and cables, which have been located by accurate measurement on the ground, programmers have recognised little need to flag an unacceptably high magnification of a component layer. Clearly, an understanding of the locational accuracy of geographical data is necessary for any interpretation. Ecological and geological databases, for instance, can rarely define boundaries to the same level of precision as those of urban street furniture. Even coasts, rivers, and in the present state of transition, territorial local authority boundaries will usually have been digitised from maps or aerial photographs and should therefore be limited to applications with output at no larger scale than that of their source maps or photographs. It must therefore be both feasible and desirable to incorporate these factors in the visual output of any geographic information system.

The problem currently appears to be considered by software designers as one of a training need for their customers. This is partly, but not entirely, true. Misinterpretation from the uninformed use of commercial GIS software would be greatly reduced by incorporating, with each layer of data, information that defines the maximum scale at which it is to be output. These factors could be compared by the software to warn a user when one layer is being displayed at a scale beyond that at which it was defined. This simple enhancement would be a valuable feature for most GIS applications.

The programs that the author developed for use with the NZAA files are much less sophisticated than those that are commercially available, but are tailored to meet the principal needs and limitations of the archaeological data. They show the location of an archaeological site as a square, the location and size of which represents the 100 metre or yard square in which the site was recorded. These programs calculate the size of square appropriate to the scale at which the site distribution is to be drawn. They calculate its correct location by centering the square at a scale distance of 50 yards or metres east and north of the grid reference coordinates. If the references were to 1,000 yard or metre grid squares, this offset would be changed to the scale equivalent of 500 yards or metres.

CONCLUSIONS

This paper began by warning of the danger in defining unrealistic goals when setting up a computer file. The worst penalty for unrealistic goals is a loss of credibility for the people or agency involved when the system proves to be too cumbersome, or even impossible, to use. But what if it is properly designed and developed, to a level that leads to requests for direct access by a wide range of users? Clearly, an inexperienced user would not be aware of the significance or limitations of the data, and is likely to accept the results at face value.

Software writers could provide assistance to overcome one part of the problem. GIS software should incorporate the means by which locational accuracy of any data set is defined. It should be used to warn a user of consequent limitations of output scale to any combination of data assembled for output. The other part of the problem requires a distinction to be made between data where interpretation is not needed and *raw* data, which must be subjected to expert interpretation.

Some files, such as address mailing lists, may require no interpretation. They might be set up with fields that clearly indicate the type of mailout for which each record is to be used. If so, they can be considered as interpreted data files but the types of file considered in this paper contain *raw* data. These are observations which provide *information* only when studied with associated knowledge, such as recording methods, areas studied, or the significance to be attached to different types of site or observation.

If these data files record places, or species that require some form of protection, then the objective of creating the files as a management tool will be defeated if they are improperly used. This is the most difficult, and potentially dangerous, aspect of the unskilled analysis of raw data. Unless the interpretation is made by people who understand their significance, analyses will inevitably lead to wrong conclusions. These could be disastrous to the subjects which were meant to be protected as well as undermining the credibility of the agency concerned.

Used properly, however, they have an invaluable contribution to make. The ability of a computer to select, from large numbers of records, only those that meet certain specifications, and to sort them into an order suited to a particular form of study, opens up countless possibilities for expert interpretation. Used in this way, the understanding that can be achieved, and the consequent contribution that can be made, by expert users is increased dramatically. Like so many modern tools, a map-based information system can be an essential aid to productivity or a mechanism for unwitting deception: a tool or time bomb.

The thoughts and observations expressed in this paper are based on the author's experience of developing this system and similar ones overseas. They are not necessarily shared by the Departments, the Trust or the Association who have jointly contributed to the initiation and development of this system.