Aggregating public submissions on Conservation Management Strategies

John Pettigrew
14 Fairview Crescent
Wellington

Published by
Department of Conservation
Head Office, PO Box 10-420
Wellington, New Zealand
This report was commissioned by Science & Research Unit.

ISSN 1171-9834

© 2000 Department of Conservation, P.O. Box 10-420, Wellington, New Zealand

Reference to material in this report should be cited thus:


Keywords: Conservation Management Strategies, public consultation, database manipulation.
Abstract

Conservancies of the Department of Conservation (DOC) had sought public submissions to their regional Draft Conservation Management Strategies in mid 1995. Most conservancies used database software to summarise and analyse these submissions. A major objective of the project described in this report was to aggregate these databases and to analyse the results, to look for conservation issues of nation-wide concern. The other main objective was to develop systems to make this process easier in the future.

Nine of thirteen conservancies provided data. Differences in database formats were surmountable, but there were also very wide differences in the amount and form of the content that had been recorded. These limited the analysis of the data.

No confident identification can be made of national conservation concerns, although some tentative conclusions might offer future lines of investigation.

The associated database is designed to allow further experimentation with the data, if desired. It is likely that more useful information could be gleaned from the database, especially by reviewing the keywords of all entries. However there are over 12 000 entries.

The value of future data would be improved by using a consistent thesaurus of keywords across conservancies. The research described in this report could contribute to the development of this thesaurus.

1. Introduction

1.1 THE PROJECT

The project set out to discover whether there were conservation issues that were a common concern amongst New Zealanders as a whole, or alternatively, whether there were identifiable regional differences in conservation concerns. DOC conservancies had sought public submissions to their regional Draft Conservation Management Strategies (CMS), and most conservancies had used database software to summarise these into `comments' or `main points'. These databases therefore represented a potentially accessible source of information about regional concerns, as long as useful and comparable information could be extracted from some widely different database formats. This project would be the first attempt to aggregate this information on a national basis.
1.2 OUTPUTS

The project had three planned outputs:

1. A consolidated database, aggregating the various different databases from the conservancies into one and providing the basis for a similar exercise in the future.

The database (CMS-NZ.mdb) is in Access97 format and is available from the author, with a detailed contract report. See appendix 2 of that report for instructions on installing the file, and appendix 3 for instructions on using the database.

The main fields of the database are:

- the comments or extracts noting issues of concern (about 15,000 of them),
- keywords (about 1500), and
- categories (about 60) to which each keyword is allocated.

The database presents the aggregated information, filtered by Conservancy, Keyword, or Category. It also allows the user to easily add or change keywords, to change the category that a keyword is allocated to, and to change the list of available keywords and available categories. This allows exploration about what issues are important in which conservancies.

2. A full-text searching and analysis system, envisaged as using a demo or single-user version of a search engine like Nud*ist, Isys, or AltaVista.

This objective proved to be unworkable in the way planned, mainly because a high proportion of the text comprised abbreviations, spelling errors, and other non-standard usages (the databases were never intended for public scrutiny).

However, the supplied database allows full-text searches, using the built-in facilities of Access. The occurrences of any word or phrase can be searched for in the whole database or in a filtered sub-set.

3. A detailed contract report, providing an analysis of the data and seeking any insights on the national scale that might lie within it.

The report also describes the methodology and notes some of the problems and issues that arose. This should help the process should an update be envisaged at a later time.
2. Results

2.1 DATA

Nine of the thirteen conservancies provided data in electronic form.

The data varied widely in depth of detail. Some conservancies used their database to record detailed extracts from the submissions, whereas others recorded only the briefest of reference notes. Some used keywords, most did not. (Where they did not, I based the keywords on the chapter headings that the extracts had been allocated to in the CMS report by that conservancy.) The extracts or reference notes are referred to variously as 'comments', 'concerns 'notes' or 'issues'. In the database they are recorded in the Issues field.

2.2 WORD USE

I analysed the word frequency in the comments. A significant problem was the large number of abbreviations (which differed by conservancy), and spelling errors. In addition, many conservation concepts are best expressed as a phrase rather than as a word.

The results of this word count can be explored in the WordCount.xls spreadsheet. The first two pages are attached as appendix 6 to the full report.

2.3 KEYWORDS

Because the keywords, where used, reflected what DOC staff at the time took to be the main issue of concern, their analysis by conservancy should provide useful information. However not all conservancies used keywords, and where they were used they differed amongst conservancies.

Also, both the structure of the databases used by the conservancies, and the method I used to generate keywords where the conservancies had not used them, usually led to only one keyword being applied to a comment. In many cases two or more would have been more appropriate. Thus many comments that should apply to more than one area of concern have tended to be defined too narrowly.

I took a very conservative policy towards changing the keywords that had been allocated by DOC staff, even when the meanings were fairly obviously synonymous. This led to a large number of keywords being created (about 1500), so I allocated them to 60 categories. (Both the keywords and the categories can be changed in the database, and it would be desirable to develop a better list of each.)
The results can be explored with the database. An analysis of the results is in the spreadsheet CategAnalysis.xls, supplied with the full report.

2.4 LIMITATIONS TO THE DATA

- Only nine of thirteen conservancies were represented.
- There were widely varying levels of detail in the databases - some conservancies included long summaries or extracts, others just had references.
- To some extent each conservancy had its own 'vocabulary', especially abbreviations, and typographic errors were common. So issues which were common across conservancies may not have been adequately brought together.
- Keywords allocated by DOC staff (one keyword per comment) may not have covered all the concerns represented in a comment - the issue might demand multiple keywords to reflect all the concerns arising in the submission.
- Generation of keywords from chapter headings may not have properly reflected the concerns represented in a comment.
- The results may be sensitive to the particular choices of keyword categories and category groups in the database. These lists are not the only ways of categorising the keywords, and are unlikely to be optimal.

3. Interpretation

For the above reasons, the results need to be treated with caution. In terms of the question "What do New Zealanders generally think about conservation issues?", no confident statements can be made.

The results do raise some intriguing questions, however. 'Issues, goals, and priorities' was a high-frequency category, although this is partly because it was a bit of a catch-all that could accommodate a wide range of more specific concerns.

The only other category that stands out is 'Commercial use'. This was high for Nelson, Southland, Waikato, and the West Coast, but quite low for the Bay of Plenty, Northland, and Wellington.

The data seem to show a tendency for concerns that affect people directly to rate higher than more 'altruistic' concerns, and for 'concrete' concerns to rate higher than those which are more abstract. However neither of these is clear-
cut, and they could just be the results of the particular choice of categories. (See the full report for a more detailed analysis.)

In looking at the keyword categories, some natural groupings seem to emerge. The Category Groups tables in the spreadsheet allocate the 60 categories into 11 broad groups to provide a broad-brush overview.

Figure 1 shows the percentage frequency of these 11 'groups of categories', across the country and overall. Note that this table is two levels of abstraction away from the original data, and some of its apparent message will be artifacts of categorising decisions.

4. Conclusions

So, how green are New Zealanders? This project has not provided sufficiently strong data for us to be able to answer that question. The full-text analysis is flawed because of the many potential synonyms (including abbreviations and spelling errors) for any word. The keywords in the database are a fragile basis of analysis because most of them, being derived from chapter headings, are not necessarily a good representation of the meaning of the comment. The other limitations in the data, as noted above, all reduce the level of confidence we can have in the results.

Nevertheless, part of the rationale for the project was to test the methodology and to make the process easier next time. The database provides a starting point for doing the same exercise again, and some changes could be considered to make the data more easily aggregated in the future. These changes are noted in the Recommendations section of the report.

5. Recommendations

The information obtained by this exercise has some big gaps and several potential biases. Better information will demand improvements in the initial data capture process.

The main objectives are therefore to capture useful data in the first place, to maximise consistency, and to minimise ad hoc processing of the data.

5.1 ENCOURAGE ELECTRONIC SUBMISSIONS

Over time, it will be increasingly acceptable for DOC to ask for an electronic copy of submissions as well as hard copy. This would make the whole process easier if the conservancies continued to use conventional databases to
help analyse the submissions, and would make the process very much easier if they moved to text analysis software. If all the submissions were on disk, then aggregating the process nationally could be done very effectively using software such as QSR Nud*ist4 or Nvivo. Electronic submissions would allow significantly more efficient processing of the main ideas.

5.2 RETAIN THE ORIGINAL SUBMISSIONS

For this project, scanning in the text of all typewritten submissions would have been technically possible but extremely expensive. However, the results would then have depended purely on the quality of the analysis, not on data limitations.

Scanning technology is improving rapidly. Perhaps at some future time it would be feasible to scan and analyse all submissions.

Even if this were some years away, it could provide data for longitudinal studies. It would be a pity if this source of information were lost because the submissions had been discarded.

5.3 RETAIN THE ANALYSIS FILES

The CMS submission analyses represent much work and a valuable information resource. The cost of archiving a few megabytes is trivial, while the cost of redoing them would be impractically high. I recommend that, as a policy, conservancies archive any data that could conceivably be useful in the future, and that careful thought be given before deleting costly information.

5.4 STANDARDISE ABBREVIATIONS

A full-text index of all the words used in CMS analyses could be informative, but is less so than it could be because of the variety of abbreviations and formatting used. Simple things like whether or not to use a period after an abbreviation, and whether or not to capitalise, make a significant difference. Also, different abbreviations are used for common words and concepts (e.g. DOC, DoC, and D.O.C.).

Given the pressure to do the main task - production of the CMS document - it may be difficult to achieve total DOC-wide standardisation. However, in many cases the choice of an abbreviation is completely arbitrary, and it would be as easy to use the standard one as to invent a new one. I recommend the development of a standard list of common abbreviations.
5.5 USE KEYWORDS

When DOC staff first read the submissions, they take a view about the issues raised. Keywords allocated at this point would be useful when later writing the CMS document, as well as for other purposes such as a consolidation exercise like this one.

I recommend that, as much as possible, notes about issues should have keywords allocated to them.

5.6 HAVE A STANDARD LIST OF KEYWORD CATEGORIES

The keywords may vary with different conservancies, because of specific local issues and concerns. Keyword categories, however, can be more consistent, and this may not require much more than agreeing on a standard terminology. The list of categories used in the database should be seen as a starting point only, used to get the database operational. It needs refinement by DOC staff.

5.7 PROVIDE FOR MULTIPLE KEYWORDS

The existing Access database used by many regions has a Keywords field - a single column - in the main table. This tends to discourage the use of more than one keyword for a particular entry, and makes keyword searching more complicated if more than one keyword is entered.

It is preferable to allow any arbitrary number of keywords to be allocated to one "Issue" entry (i.e. a comment or note). Thus an issue relating to coastal ecosystems should be keyworded under both "Coastal" and "Ecosystems", so it would be picked up in a search for either.

This requires the keywords to be in a separate, linked, ‘keywords’ table rather than being entered into a column in the main table. The database supplied with this report is designed in this way. However, if conservancies prefer to use the Access database they have used in the past, I recommend that this change be made to it (the change is reasonably straightforward).
### Category Groups – Percentage frequency

<table>
<thead>
<tr>
<th>Category Groups</th>
<th>Bay</th>
<th>Cant</th>
<th>NeIg</th>
<th>North</th>
<th>Otag</th>
<th>South</th>
<th>WaiK</th>
<th>Wellington</th>
<th>West</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Management</td>
<td>10</td>
<td>5</td>
<td>7</td>
<td>11</td>
<td>12</td>
<td>15</td>
<td>11</td>
<td>13</td>
<td>15</td>
<td>7</td>
</tr>
<tr>
<td>Facilities for people</td>
<td>12</td>
<td>11</td>
<td>15</td>
<td>11</td>
<td>7</td>
<td>13</td>
<td>13</td>
<td>12</td>
<td>13</td>
<td>7</td>
</tr>
<tr>
<td>Habits</td>
<td>1</td>
<td>11</td>
<td>12</td>
<td>6</td>
<td>7</td>
<td>13</td>
<td>13</td>
<td>12</td>
<td>13</td>
<td>7</td>
</tr>
<tr>
<td>Inhabitants</td>
<td>8</td>
<td>12</td>
<td>11</td>
<td>7</td>
<td>11</td>
<td>12</td>
<td>11</td>
<td>8</td>
<td>11</td>
<td>7</td>
</tr>
<tr>
<td>Human impact</td>
<td>10</td>
<td>7</td>
<td>12</td>
<td>5</td>
<td>7</td>
<td>14</td>
<td>13</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Commerce</td>
<td>5</td>
<td>7</td>
<td>14</td>
<td>3</td>
<td>11</td>
<td>3</td>
<td>13</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>Cultural</td>
<td>15</td>
<td>6</td>
<td>4</td>
<td>15</td>
<td>6</td>
<td>1</td>
<td>9</td>
<td>6</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>Information</td>
<td>4</td>
<td>8</td>
<td>6</td>
<td>5</td>
<td>1</td>
<td>7</td>
<td>6</td>
<td>8</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Stewardship</td>
<td>2</td>
<td>7</td>
<td>5</td>
<td>6</td>
<td>8</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>Systems</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>5</td>
<td>5</td>
<td>13</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Impact of technology</td>
<td>1</td>
<td>4</td>
<td>6</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>6</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

Figure 1. Percentage frequency of Category Groups.

### Appendix 1: Available files

- CMS-NZ.mdb (MS Access 97 database).
- WordCount.xls (Excel 97 spreadsheet, word count of Issues field.)
- CategAnalysis.xls (Excel 97 spreadsheet, analysis of Keywords and Categories fields.)
- HowGreen.doc (The full report in MSWord.)
- HowGreenShort.doc (This report in MSWord.)
Appendix 2: Groups, Categories, and Keywords

This shows all the Keywords and Categories in the database, and the Groups in the analysis spreadsheet, listed as a hierarchy. The portion headed ‘No Group/No Category’ comprises keywords which have not been allocated to a category or group. This source of this appendix is the report GroupsCategoriesKeywordsExport, in the database. Longer keyword phrases are truncated to save space.

Group
Category
Keyword

[No Group]
[No Category]

Integrated Management
Integrative management
Intellectual property rights
International Significance
International Market
Interim rec facilities
Introduction to Geographical Overview
ISSUES, signs
Kai Tahu
Kaka - Methods
Kaka - Results
Kaka Invasive Threats
Kaka, Invasive Threats
Kaka, special ranking
Kapahau - Volcano
Kapapa Anawhata
Kapapa Anawhata Strategy
Keros
Key Visitor Attractions
LAC
Landcorp, Epi centre
Landings
Limits of Acceptable Change
Link areas
Linkages
Local botanical expertise
Long term plans
Meltings ka
Maintaining a Range of Opportunities.
Marginal Sizes Explanation
Marine Stewardship
Memorandum of Understanding
Methods
Mole cricket
MORP
MTB Tracks
Natural quiet
Natural resources
Navigation aids
Network of reserves
Network Utilities
Night Sky
Non Commercial sites
Non-statutory plans
Notes
Numbers
OGP and CMP
On-site Interpretation
Open Space Covenants
Other Uses
Overview
Park Addition Stewardship
Patterns of Protection
Physical and Ecological
Physical description
PLACE OBJECTIVES AND
Place general Place sections
PNA ACCESS
PNA definition
PNA programme
PNA surveys
PNAP
PNAP in Otago
Pollution
PR in AFPO
Prevention
Private & Public
Pronouns
Protected Area Network
PRR
Ratans
RAP
RAP definition
RAPs
Rationale
Rationale for Restrictions
Reclassification
References
Remnant plains plants
Reserves
Resource and estate use
Resource Use
Resources
Responsibilities
RESR
Restoration
Restoration funding
Restoration map
Restrictions
Results Sought
Review
RHA
RHA, WARs
RHAs
Rock art
Rock drawings
Roles, ‘background’
Rocks
ROS
ROS classes
Rusunga & CRC
Rural back-country drive in
Salt Marsh
Sanctuary
SIR
Significance
Social
Social & Economic Well Being
Social implications
Southern
Special Places
Sports fish and game bird management
STA
Staffing
Stewardship
Survey, wildlife/green corridors
Table
Tables General
Taupou
Taupou maunga
Te Rusunga O Raipa
Te Taumata Rusunga
Teachers
Telecommunication Facilities
Telecommunications
Telecommunications Imp
Thames
Threat Seeker
Tomanga
Town Belt
UCL
Umbrella
UNCED
Update
Upgrading
Urban fringe

Use
Use generally
Uses Not Otherwise or Fully Provided For
Uses of Areas Managed by the Department
Using Raipa knowledge
Vandalism
Veteral assurance by staff
WAC Staff
Water Based Activities
WCD
Wildlife Encounters
World Heritage Site

Commerce

Commercial use
Commercial activities
Comrun EEL Fishing Access commercial
Commercial Accommodation
Commercial and Other Uses
Commercial Concessions
Commercial Enterprise
Commercial fishing
Commercial Freshwater Fishing
Commercial hunting
Commercial Operations
Commercial Recreation
Commercial Recreation and Tourism
Commercial Sites
Commercial Sphagnum Harvesting
Commercial Structures and Public Works
Commercial Use
Commercial Use of Land
Commercial use/cultural use
Commercial v rec fish and game
Commercial vs rec hunting
Communique and Liaison
Community and Liaison
Concession Application Criteria
Concessions
Concessions - Raipa
Concessions and Commercial Activities
Easements
Easements and Other Uses
East Fishery
East Fishing
Electricity generation
Engineering
Farming
Filming
Filming & Sphagnum
Forestry
Forests Amendment Act 1993
Graceing
Graceing Compliance 396
Graceing Loen File Release
Graceing Licences
Graceing Licence and leases
Guiding
Hydro
Hydro power
Hydro-electric Power Development
Hydro-Electric Generation
Leases
Leases, Licences & Concessions other than
Limits of Effect of Concession Activities
Logging