

Otago Biodiversity Data Management Project Report

Part 1: Questionnaire Report



Surat Bay, Catlins, New Zealand 2007

Gillian Elliot
Biodiversity Project Manager
University of Otago Library
New Zealand

October 2008



Otago Biodiversity Data Management Project Report by [University of Otago Library](http://www.library.otago.ac.nz) is licensed under a [Creative Commons Attribution-Non-Commercial-Share Alike 3.0 New Zealand License](http://creativecommons.org/licenses/by-nc-sa/3.0/). Based on a work at www.library.otago.ac.nz. Permissions beyond the scope of this license may be available at <http://www.otago.ac.nz/administration/copyright/>.

Contents

Executive summary.....	3
Questionnaire background, purpose, context & methodology.....	4
Project background	4
Project outline	5
Project documentation	5
Purpose and context of questionnaire	5
Methodology	7
Acknowledgement	8
Otago Biodiversity Project Questionnaire - responses	9
<i>A. CONSTITUENCY</i>	10
<i>B. PROJECT AIMS</i>	10
<i>C. YOUR RESEARCH DATA</i>	11
<i>D. SOURCE AND DISCIPLINE REPOSITORIES</i>	19
<i>E. METADATA</i>	21
<i>F. DATA ACCESS AND SHARING</i>	24
<i>G. SUPPORT AND FURTHER INTEREST</i>	30
Questionnaire responses – summary observations	32
Appendix A: Otago Biodiversity Project Questionnaire.....	36
Appendix B: List of questionnaire questions.....	49
Appendix C: List of respondents’ research areas.....	50
Appendix D: Summary of ‘other comments’	51

Executive summary

The 'Otago Biodiversity Data Management Project' aims to establish the framework for managing and sharing Otago biodiversity primary research data. This 12-month Library funded initiative has the support of the University of Otago Life Sciences Departments, in association with the Ecology Teaching Programme and the Ecology, Conservation and Biodiversity Research Group. It is also aligned with one of the University's 'Emerging Research Themes', addressing *Ecology, Conservation and Biodiversity in New Zealand*.

One of the main Project activities is to survey the level of interest in data management and curation among University of Otago researchers with an interest in New Zealand biodiversity, and also to learn more about the current data management practices of these researchers.

An online and print questionnaire, the *Otago Biodiversity Project Questionnaire*, was made available (between April and August 2008) to 170 Otago researchers, with an interest in biodiversity, including some non-University researchers. A total of 71 responses were received, representing a response rate of 42%.

Questionnaire findings

This document presents the findings of the *Otago Biodiversity Project Questionnaire*, including both quantitative, tabulated responses to multi-choice questions and qualitative free text comments. Some key findings from the questionnaire responses include:

There is considerable interest in formal research data management among Otago researchers (70% of respondents wish to stay in touch with this Project and further developments)

The issues of data management are not widely understood by all researchers (indicated by the high number of 'don't know' responses and also through free text comments)

The majority of Life Science researchers have relatively small amounts of data (<1TB) and this data typically has a long life (>10 years)

The majority of researchers manage their own data and apply their own metadata

Most researchers indicate interest in sharing data with others but typically do so informally only (email, personal contact and direct F2F data sharing remain popular)

Practical considerations (time/support) are barriers to sharing data (along with IP/ownership)

Questionnaire recommendations

This document summarises the findings from the questionnaire and adds further comments and observations. A second report: *Otago Biodiversity Data Management Project Part 2: Survey Report*, will add additional, qualitative data derived from interviews held between May 2007 and August 2008, with over 70 Otago researchers, with an interest in biodiversity. Recommendations resulting from both these documents¹ will be included in the final Project report - *Otago Biodiversity Data Management Project Part 4: Library Project Report*.

¹ The *Otago Biodiversity Data Management Project Part 3: Pilot Project Report* will summarise the activities and findings of a test case 'pilot project', working with a sample biodiversity data collection.

Questionnaire background, purpose, context & methodology

Project background

To gain a fuller understanding of the current Library/support needs of Otago's research community, Library staff embarked on a series of interviews with Life Scientists at the University of Otago, in May 2007. These, largely informal, interviews posed the question:

If the Library could do one thing that would better support your research what would that be?

It quickly became apparent that researchers, at least within the Life Sciences, had a need for greater support with managing a range of digital and non-digital information (including a wide range of data 'types'), which did not 'fit' easily within the parameters of what is normally managed by a university (or other) library i.e. published information sources.

In summary, researchers highlighted a need for mediated support with management and curation² of their biodiversity³ research data.

Further interviews and research highlighted the following:

- Internationally, New Zealand organisations and individuals are playing significant roles in the creation of databases and infrastructures, such as GBIF⁴, to improve management of, and access to, biodiversity information.
- Nationally, organisations such as Landcare Research and DOC are driving numerous initiatives – managing biodiversity databases and repositories and ensuring access to, and interoperability between, these resources.
- Locally, at the University of Otago, much biodiversity information is non-digital, locally-stored and difficult to access.

Why is New Zealand biodiversity important?

New Zealand's native biodiversity is unique, born of long isolation as small islands in a vast ocean. The high percentage of endemic species (those found nowhere else in the world), make New Zealand's native biodiversity both special and highly vulnerable...For example, both species of New Zealand bat are found only here in New Zealand, as are all four frogs, all 60 reptiles, more than 90 per cent of insects and marine molluscs, about 80 percent of vascular plants, and a quarter of all our bird species. By comparison, Great Britain, which separated from continental Europe only 10,000 years ago, has only two endemic species – one plant and one animal.

Why we value Biodiversity, retrieved 22_6_07: www.biodiversity.govt.nz/picture/biodiversity/why.html

Why manage biodiversity information?

While much biodiversity and ecosystem information currently exists (from a legacy of past research and inventories), and much more is collected on a daily basis, it is still not possible for all those who could benefit from having access to this information to locate, retrieve,

² *Data curation is the active and on-going management of data through its lifecycle of interest and usefulness to scholarship, science, and education. Data curation activities enable data discovery and retrieval, maintain its quality, and provide for re-use.* From Data Needs Assessment Survey, p.2, 2008, University of Illinois

³ Biodiversity: biological diversity of life

⁴ Global Biodiversity Information Facility <http://www.gbif.org/>

integrate, and apply it in any consistent fashion. In many cases, public and private funds are unknowingly spent on re-collecting information that may actually already exist in some undocumented or unavailable fashion. Much existing biodiversity and ecosystems information cannot be widely used (and may be in danger of being permanently lost) because it is not yet converted into an electronic (computerised) format. In most cases, because of different formats, conventions, or technologies, it is difficult to truly integrate information from more than one source or system.

Strategy for Biodiversity and Ecosystems Information Framework Document, Biodiversity and Ecosystems Informatics Work Group⁵, retrieved 21_6_07: www.bioeco.gov/about/nbiiframework.html

Project outline

This Library initiated project, known as the ‘Otago Biodiversity Data Management Project’ was formally established at the end of 2007 and has been funded for 12 months, 1 January to 31 December 2008. The Project aims to investigate the feasibility of managing, curating and making available, University of Otago and regional biodiversity research data to empower research, further enable collaboration and inform conservation practice.

One of the key Project Deliverables is to survey University of Otago researchers, with an interest in biodiversity (including and beyond the Life Sciences), to quantify some of the findings suggested through the informal interviews held in 2007.

Project documentation

This report is the first of four Project reports. These are:

- *The Otago Biodiversity Data Management Project Part 1: Questionnaire Report*
- *The Otago Biodiversity Data Management Project Part 2: Survey Report*
- *The Otago Biodiversity Data Management Project Part 3: Pilot Project Report*
- *The Otago Biodiversity Data Management Project Part 4: Library Project Report*

Purpose and context of questionnaire

Formal management of primary research data⁶ is (arguably) in the early stages of development in New Zealand tertiary institutions. As a result, the *creators* of the data (the researchers) are generally unaware of the issues around ‘data management and curation’ and also what data management may mean for their research. Similarly, the potential *managers* of this data (here, referring to libraries) are confronting many ‘unknowns’ surrounding data management.

This questionnaire attempts to achieve some or all of the following:

1. **To raise awareness** of the issues surrounding the management of research data, and also raise awareness of the current Otago Biodiversity Data Management Project, among a group of Otago researchers with an interest in New Zealand biodiversity
2. **To quantify responses to questions about research data management**, using a structured, measurable approach (building on interviews done in 2007/2008)

⁵ The BioEco Work Group was chartered in 1997 by the Committee on Environment and Natural Resources (CENR) Subcommittee on Ecological Systems (SES) to improve coordination of Federal Biodiversity and Ecosystem informatics activities and to provide a recognized, high level focal point for those activities.

⁶ Mark Euston, from ANU, neatly summarised ‘Data Management’ (2008) as “*anything outside of actually using the data.*” http://www.apsr.edu.au/data_management_plan/euston.pdf

3. **To survey Otago researchers with an interest in biodiversity:**
 - how they currently manage their research data
 - what value they place on this research data
 - whether they share their research data and access others' research data
 - what issues concern them re sharing/accessing research data
4. **To enhance understanding** of data management for discovery (including metadata) within the research community
5. **To identify researchers with a particular interest** in formally managing their own primary research data for enhanced discovery
6. **To identify additional considerations** which are important to University of Otago researchers
7. **To gauge what interest there may/may not be** in pursuing research data management as a mainstream activity of the University Library (and University)
8. **To identify potential partnerships** – both within and beyond the Otago research community. (The primary focus of the Project is the Otago region, but partnerships from outside this region are not excluded.)

Throughout the following sections the *Otago Biodiversity Data Management Project Part 1: Questionnaire Report*, summarises and reflects on the findings of the 'Otago Biodiversity Project Questionnaire'.

Some of the findings of this report will also inform part of a larger report, the *Otago Biodiversity Data Management Project Part 2: Survey Report*, which will also include additional, qualitative data derived from **interviews with 70+ Otago researchers**, with an interest in biodiversity. (Interviews took place between May 2007 and August 2008).

NOTE: Recommendations resulting from both documents will be included in the final project report: *Otago Biodiversity Data Management Project Part 4: Library Project Report*.

Audience

Both reports - *Part 1: Questionnaire Report* and *Part 2: Survey Report* – will be available to:

- a. **University of Otago Library Management**
These reports are Project 'deliverables', informing the Library (Project sponsor) of the progress of the 'Otago Biodiversity Data Management Project' and, in this current report, the results and findings of the 'Otago Biodiversity Data Questionnaire'.
- b. **Project Faculty Advisory Group:** Biodiversity Data Management Advisory Group
Appointed in March 2008, this group of academic representatives from a range of university departments are 'advisors' to this Project.
- c. **Selected Otago researchers:** questionnaire respondents, and other Otago researchers, who have expressed ongoing interest in the current 'Otago Biodiversity Data Management Project' and the findings of the 'Otago Biodiversity Project Questionnaire'.

Methodology

This questionnaire has been adapted (with approval) from three earlier surveys, from the UK (*Project StORe*, 2006), Australia (*Data Management and eResearch Practices at UQ*, 2007), and the USA (*Illinois Data Needs Assessment Survey*, 2008).

The questionnaire contains 25 multi-choice questions, and includes a 'free text' response option with each question. It was initially developed as a print questionnaire, before being converted to an online survey using Google Documents.⁷

The target survey group for this questionnaire included *any researcher at the University of Otago, who could be identified as having an interest in biodiversity*. This included all academic staff from the Life Science departments (Zoology, Botany and Marine Science) and also CSAFE⁸. Selected staff from other University departments, who were identified as having an interest in biodiversity, were also contacted. This included individuals from: Biochemistry; Chemistry; Clothing and Textiles; Design Studies; Geography; Geology; HEDC⁹; History; Information Science; Law; Management; Maori Studies; Maths; Philosophy; Political Studies; Surveying and Tourism.

Prior to distributing the questionnaire, it was shared with a small group (including non University researchers) at the 'Ecology Programme and Botany Department Seminar' (2nd April, 2008). This, and a further conversation with a local DOC manager, *widened the survey group to include others working in the area of biodiversity, within the Otago region*.

The online questionnaire was emailed to the survey group on the 7th April 2008. A reminder email was also sent out 10 weeks later, on 20th June. The online questionnaire remained available until 1st August. Between April and August, ongoing interviews with researchers provided an additional opportunity to discuss and distribute the (print) questionnaire. This method of distribution proved most effective and the majority of questionnaire responses received were print, rather than online.

In total, **170 individuals from the Otago region** were contacted and encouraged to do the questionnaire. The survey group included University of Otago researchers and independent researchers, with an interest in biodiversity, and also selected DOC¹⁰ managers.

By the end of the survey period, 71 responses (both in print and online) had been received.¹¹ This represents a questionnaire **response rate of 42%**.

The completed questionnaires contain both quantitative data (multi-choice responses) and qualitative data (free text responses). This report includes graphs of the quantitative results and also commentary on these, and the associated free text comments. The questionnaire respondents have not been identified.

⁷ Exporting the data from this online questionnaire into *Excel* (for analysis), presented a number of problems. Google Documents cannot be recommended for this type of survey in future.

⁸ Centre for the Study of Agriculture, Food and Environment: www.csafe.org.nz/

⁹ Higher Education Development Centre: <http://hedc.otago.ac.nz/hedc/>

¹⁰ DOC: New Zealand Department of Conservation: www.doc.govt.nz/

¹¹ All print responses received were added to Google Documents manually

Acknowledgement

The 'Otago Biodiversity Project Questionnaire' is based on the:

- *Project StORe survey: cross-discipline report* (2006), Edinburgh Research Archive:
www.era.lib.ed.ac.uk/handle/1842/1419

Kind thanks to Graham Pryor for his advice and access to this report.

Permission was also granted to adapt elements from two further surveys:

- *Data Needs Assessment Survey* (2008) – thanks to Melissa Cragin, University of Illinois and also the
- University of Queensland survey (2007), available via the APSR website:
www.apsr.edu.au/currentprojects/data_management_at_uq_blank_survey.pdf.

Otago Biodiversity Project Questionnaire - responses

This section presents – question by question - the 71 responses to this questionnaire. The information (included in pages 10-31) is presented as follows:

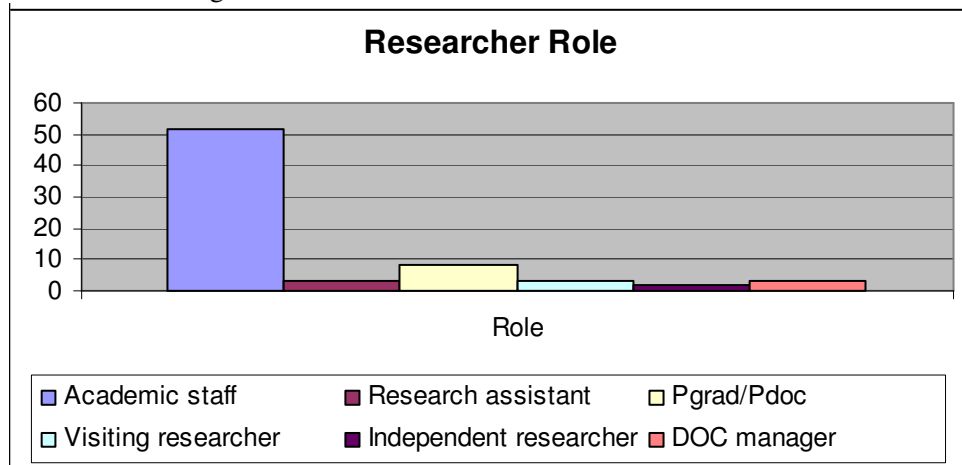
- a. **Multi-choice responses** are represented graphically, using bar and pie charts. This quantitative data is displayed both as numerical and percentage values.
- b. **Free text responses** (when available) are included, anonymously, with each question. These are presented in italics, on a grey background.
- c. **Commentary** – a short commentary is included in support of the responses to each question. The intention is to draw out some of the key findings from these responses and, as appropriate, to supply further context.

A. CONSTITUENCY

1. Please identify your own role by selecting from the following list

While ‘University of Otago (OU) researchers with an interest in biodiversity’ were the primary target group for this questionnaire, a number of non-OU researchers expressed interest in this Project and the associated survey. 7% of responses received were from Otago DOC managers and independent researchers.

74% of respondents identified themselves as ‘University academic staff’ i.e. established researchers at Otago.



Each respondent selected a single response to this question.

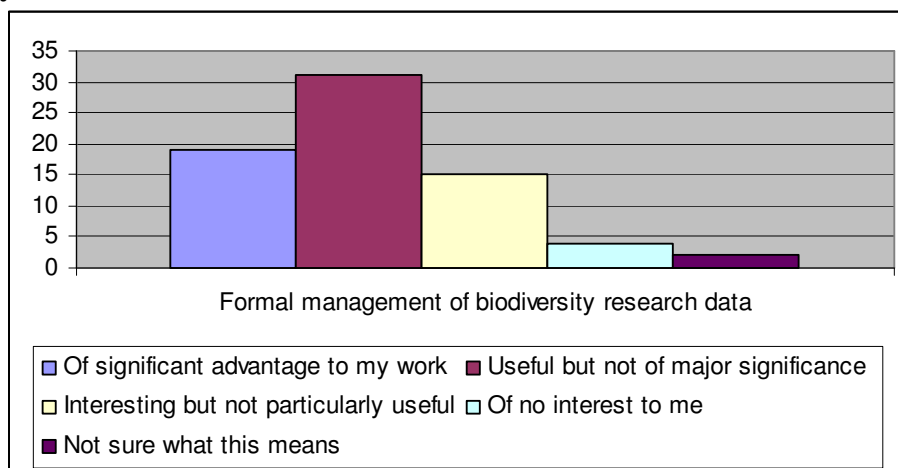
[All bar charts indicate the **number of responses** on the ‘y’ axis (rather than % of responses.)]

2. Please indicate your area of research

Only one respondent chose not to include information about their particular area of interest. A complete list of research topics is included in *Appendix C*.

B. PROJECT AIMS

3. How would formal management of biodiversity research data be useful to you?



Each respondent selected a single response to this question.

70% of respondents selected either ‘Of significant advantage to my work’ or ‘Useful but not of major significance’. Of those selecting the latter, associated free text comments raise some pertinent concerns:

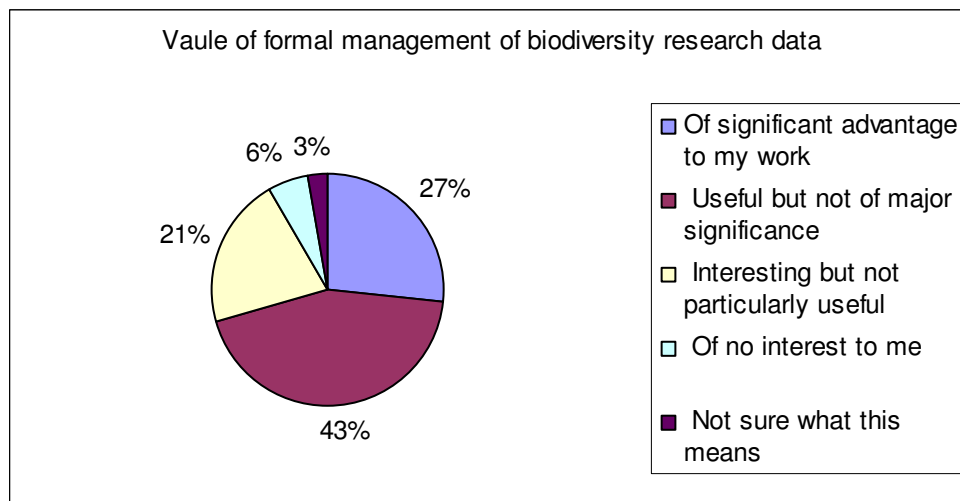
“It really depends on how it is managed. It could make things much better or be a real hindrance.”

“[X] are consumers rather than producers of data, but it is very interesting to know what others are doing. However, the tradition in [X] is to cite only published data.”

Only a small number of respondents (6%) expressed no interest in formal management of biodiversity research data. Two of the supporting free text comments suggest this is dependant on the Project approach, rather than formal management of research data:

“Little of my research concerns biodiversity at present.”

“Although I recognise biodiversity, this scope of work is far too narrow for dealing with the issue.”



Two respondents were ‘Not sure what this means’, and one added a further comment:

“The more readily accessible data is the better for ongoing management esp with the ongoing loss of the corporate memory in the department.”

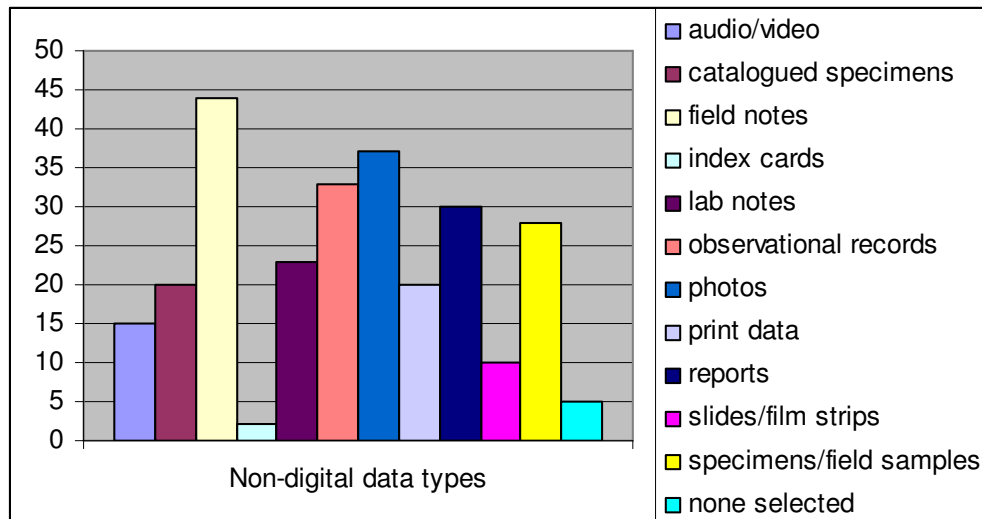
C. YOUR RESEARCH DATA

4. What kinds of non-digital data do you generate or collect for your research (or have you generated or collected in the past)?

‘Field notes’ were the most selected non-digital data type, selected by 62% of respondents, followed by ‘Photographs’ (52%), ‘Observational records’ (46%) and ‘Reports’ (42%). Just 7% of respondents reported having no ‘non-digital data’ (or chose not to select any of the options). Other non-digital data formats suggested by respondents included:

“Published papers”, “Genotypic data on organisms” and “Questionnaires”.

Non-digital data types collected

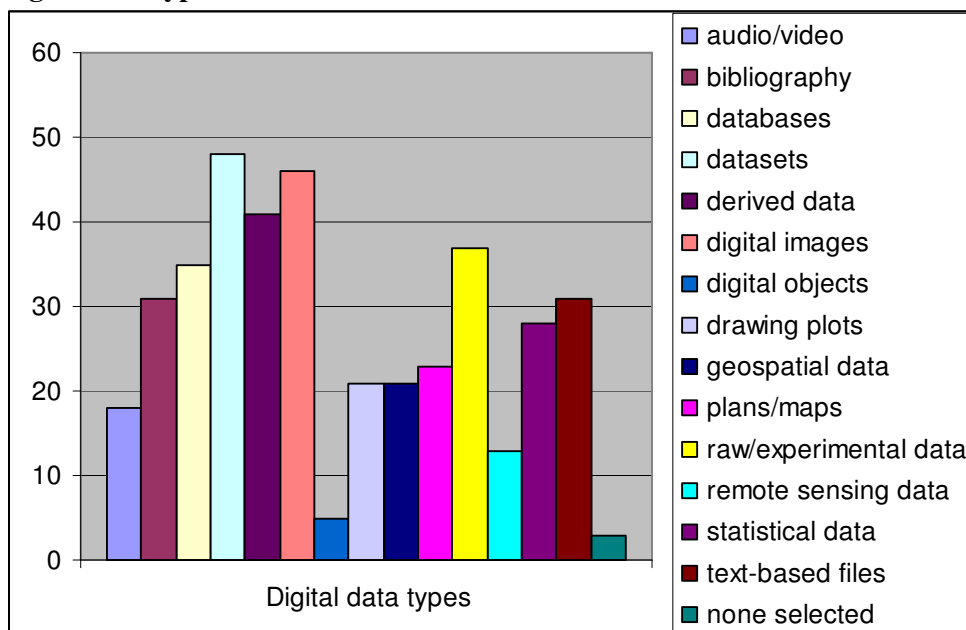


Respondents selected multiple responses to this question.

5. What kinds of digital data do you generate or collect for your research?

The highest proportion of respondents held 'Datasets' (68%). The most reported digital data types were 'Datasets', 'Digital images' (65%) and 'Derived data' (58%). Almost all researchers (apart from those selecting 'none') had multiple kinds of digital data.

Digital data types collected



Respondents selected multiple responses to this question.

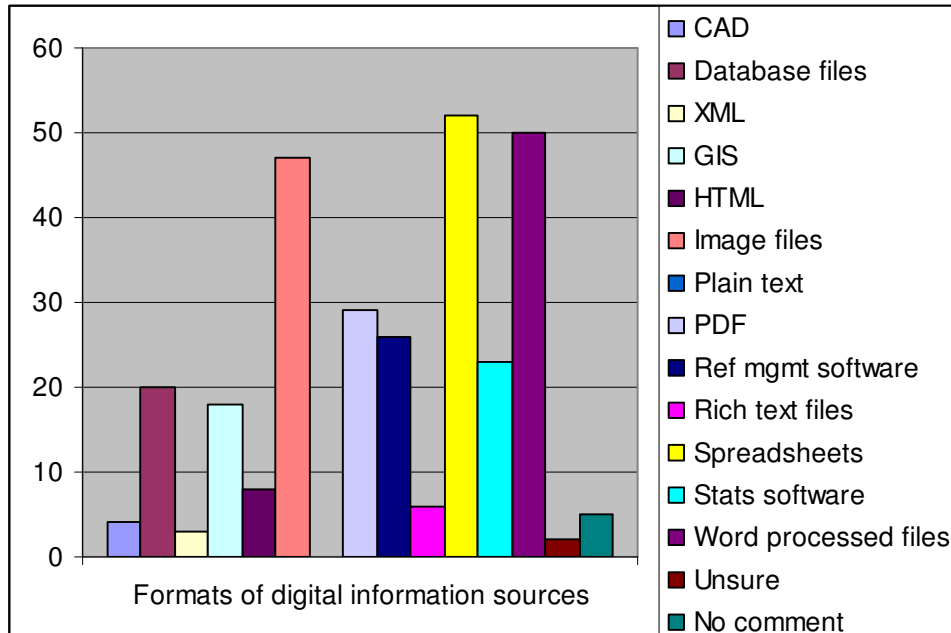
Other digital data types suggested by respondents included:

“GIS data”, “Field Guide - published & digital”, “Published papers (in journals)”, “GPS data (Ground Control Points)”, “Databases...(Paradox, MS ACCESS)”, “Observational records ie. blog, Field notes”, “We have footage (digital) of a lot of the otago area...”, “Climate data, Banding returns”.

6. In what formats are these digital information sources held?

'Spreadsheets (e.g. Excel/.xls)', 'Word processed files (e.g. Word/.doc)' and 'Image files (e.g. .jpg, .tif, .bmp, .gif)' were the most highly selected formats, chosen by 73%, 70% and 66% of researchers respectively. A small number of researchers, just 4%, were 'Unsure of the format' or made no comment.

Formats of digital information sources

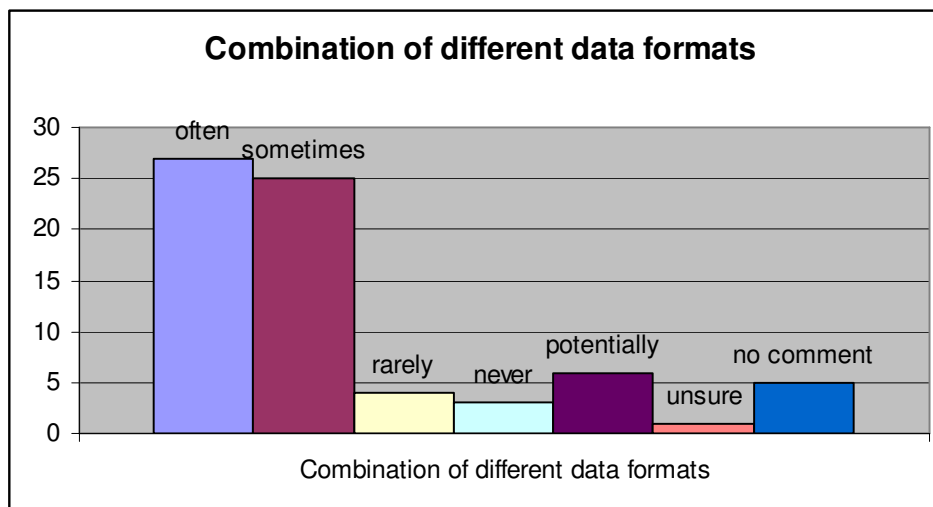


Respondents selected multiple responses to this question.

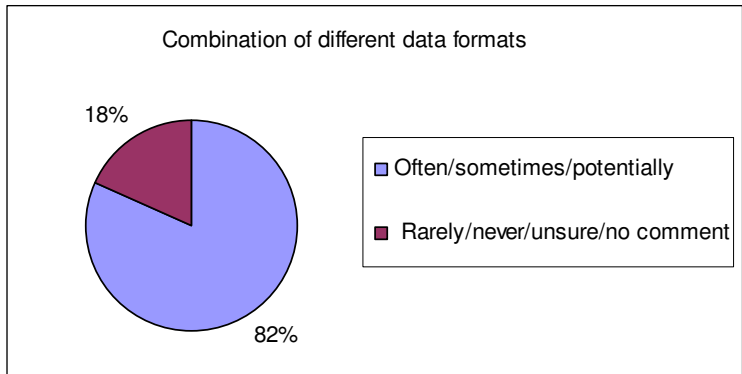
Other digital information source formats, suggested in the free text comments, include:

"Latex, Bibtex, Matlab data files", "RINEX, Tide Gauge Records", "Mindmaps (free Mind format) which can be exported as XML", "DV tapes...", "mov files (video)", "In house software data files (spreadsheet-like), Ocean Data View, Matlab" and "MP3".

7. Are the research data you generate sometimes a combination of different data formats?



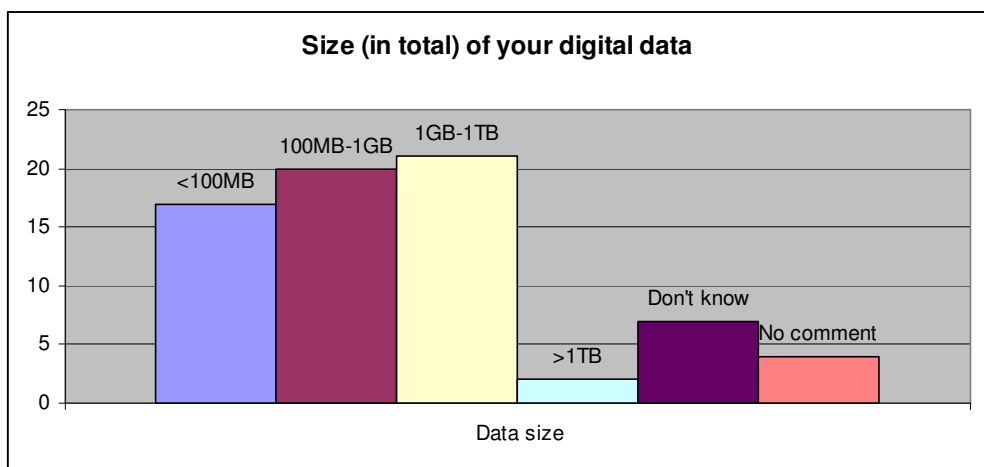
Each respondent selected a single response to this question.



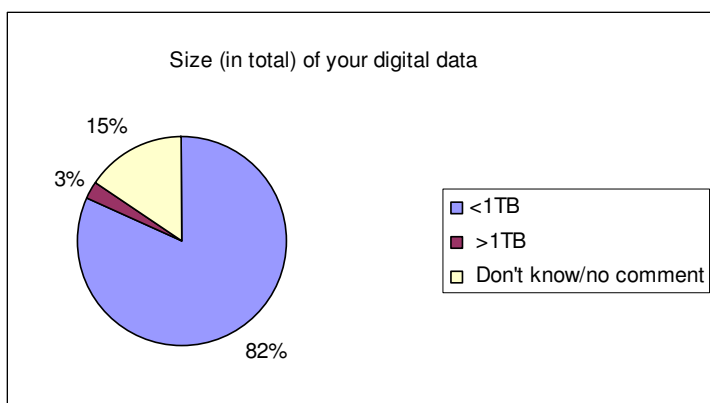
38% of respondents agreed that the data they generate is 'Often' a combination of different data formats and 35% agreed that this is 'Sometimes' the case.

A further 8% agreed that data is 'Potentially' a combination of different data formats.

8. How large (in total) are your digital research data? Please estimate.



Each respondent selected a single response to this question.



52% of respondents (individually) held less than 1GB of data and 82% held less than 1TB of data.

The number of researchers who selected 'don't know' or 'no comment' is also significant, 15% of total respondents.

Only two researchers claimed to have data in excess of 1 TB.

This question generated a number of free text comments, possibly as a result of the uncertainty around the question of data size:

- "Don't know - pretty big as they are massive image files through GIS"*
- "Not sure what counts"*
- "Not sure. There is lots of data stored across many different computers."*
- "Photos and maps are large files - data and text are less than 100MB"*

“Have no idea - lots!”

“Depends on size of dataset - GIS and images are very large”

“This is hard to estimate but probably <1 terabyte in total”

“But not really sure”

“ie. per project” [checked 100MB-1GB]

“Maybe 2 Gig”

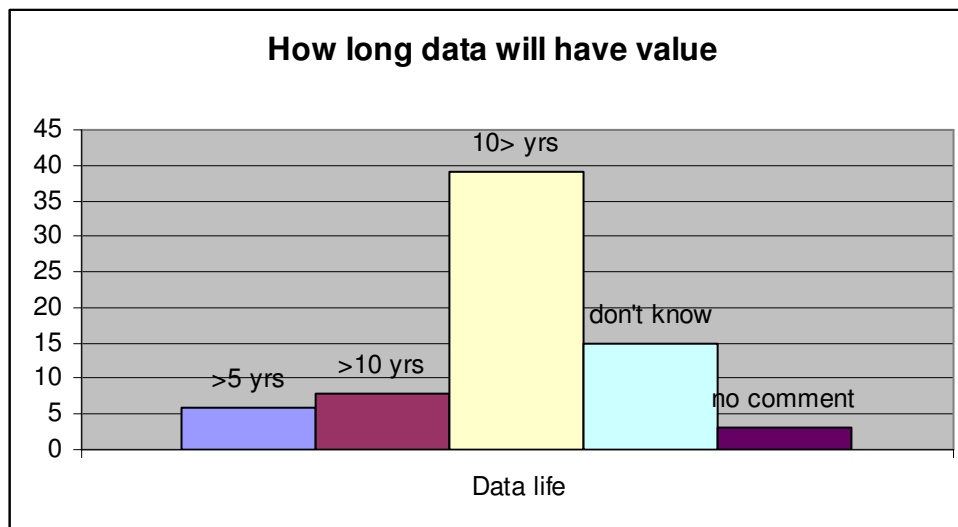
“How far back do you want to go - if you include digital photos it's huge but I'd only be guessing”

“Hmm not sure I can guess at this - probably less than 100 MG”

There was also one comment relating to data obsolescence and ongoing availability of data:

“Some of my digital research data is now inaccessible owing to changes in software (statistical packages now obsolete; drawing, graphics and word-processing packages now obsolete)”

9. How long do you think your research data will have value?



Each respondent selected a single response to this question.

55% of respondents indicated that their data will have value for more than 10 years. The next highest response category was 'don't know' (21%), indicating considerable uncertainty in knowing the value of research data over time. Again, a number of researchers included (quite discipline specific) supporting free text comments:

“Personal photographs of specific parts of NZ, eg the Milford Track, Karori Wildlife Sanctuary or Orokonui Ecosanctuary, provide an historical record of the change with time.”

“Depends on the projects I am involved, most of them up to 10 year, another one more than ten years as it is a long term monitoring study.”

“NZ reptiles v. long lived - typically, several decades. So data for individual animals remains of value to future workers beyond my expected career completion.”

“The films are watched for many years and used as a teaching tool too....”

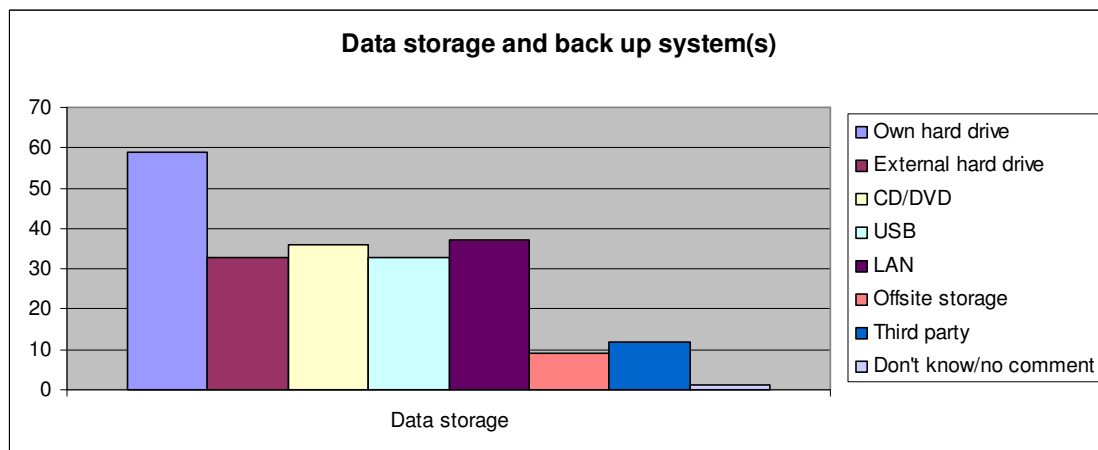
“Some of the data is used by the lay community, so don't know how long”

“Because significant biological events are often determined by ENSO climate cycle and because much biota in NZ is very long lived, Multiples of ten years are useful periods to collect and maintain data.”

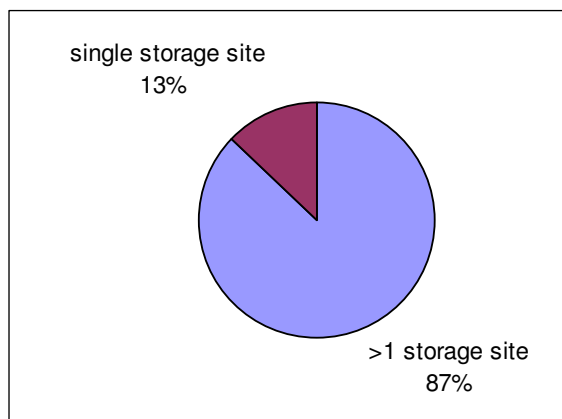
“depends on what research will follow on national/ international scale and on the project, my work on stress in [X] species might have value for the next 5 years, the community based research with Maori might have value for longer, should it be useful for local fisheries management”

“Our specimens will be of value as long as they last. Sequence data will be valid in 10 years but I can't tell if it will be useful or not. But I would say 10 year timeframe is not unreasonable. However that sequence data is already curated by Genbank.”

10. What data storage and back up system(s) do you currently have in place?



Respondents selected multiple responses to this question.



Single v Multiple Storage Sites

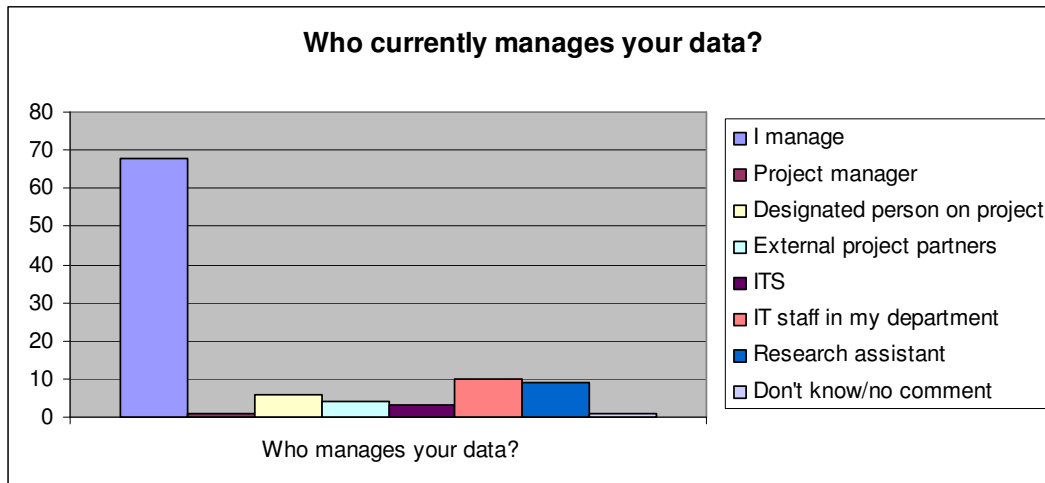
While 82% of respondents indicated their research data was stored on their ‘Own computer hard drive’, all but nine researchers also stored their data in two or more locations.

Those who indicated just one storage site, were using either their ‘Local area network’ (LAN) or ‘Offsite storage’.

These statistics are encouraging, suggesting that researchers appreciate the importance

of data storage and back up systems.

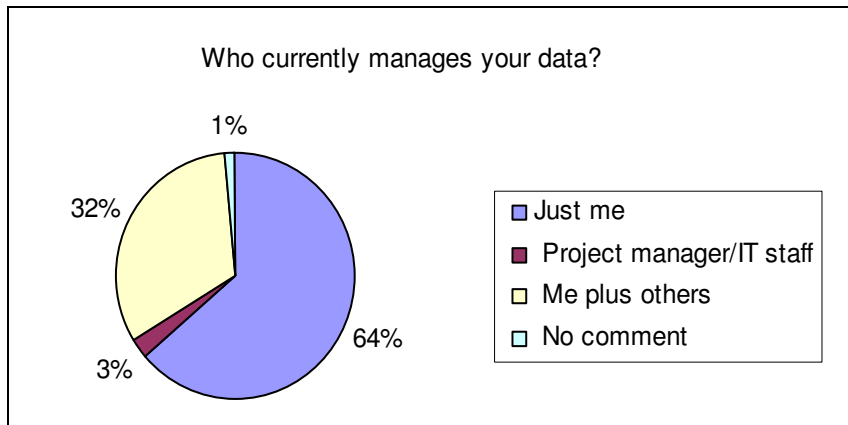
11. Who currently manages your data?



Respondents selected multiple responses to this question.

96% of researchers selected 'I manage my own data'. One researcher checked 'no comment', and the remaining two researchers checked either 'Research project manager' or 'IT staff within my department, centre or research institute'.

64% of all respondents indicated they are the **only** ones to manage their own data.



Free text comments indicate that postgraduate students play a role in managing at least some of the researchers' data:

"Much is generated and initially held by my postgraduate students"

"Generally a [collaboration] between the supervisor, research assistants and a group of postgrad students."

"And students (postgraduate) also manage data for their projects. I ask for copies (digital - CD) of all data sets on thesis completion."

"Postgrad research students"

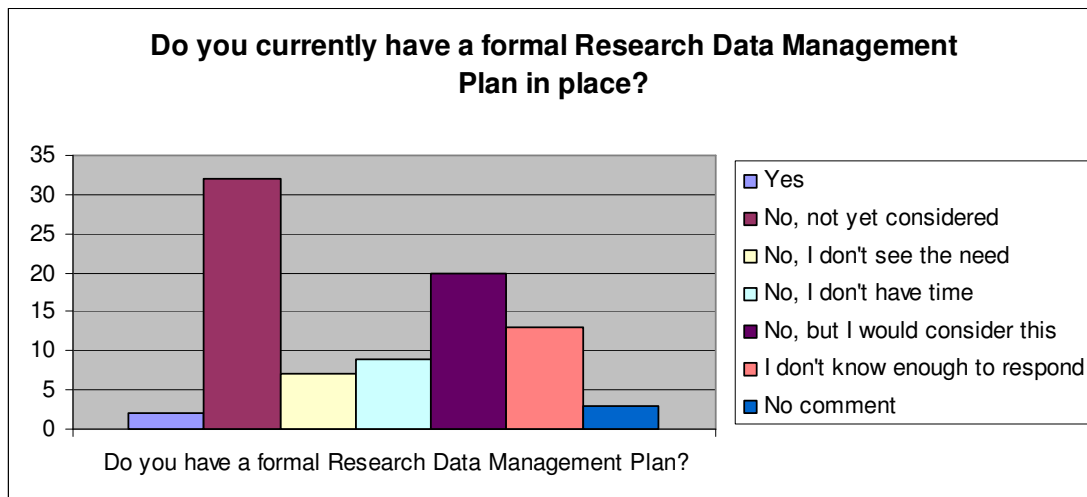
"My research students (postgraduate)"

"...is also ongoing data exchange with postgrad students..."

Several researchers also mentioned external organizations or databases:

“National Vegetation database”, “I work with NIWA...”, and “NZ Plant Databases [and] NZ Plant Conservation Network (collaborations arranged but not yet activated)”

12. Do you currently have a formal Research Data Management Plan in place and, if not, please indicate the reasons.



Respondents selected multiple responses to this question.

When asked about a formal Research Data Management Plan, 45% of researchers responded ‘No, I haven’t yet considered this’. The second largest response (28%) suggested interest in ‘developing a formal plan in future (given time and support)’.

Only two researchers reported having a ‘Data Management Plan currently in place’. A significant proportion, 15%, did not ‘know enough about this topic to respond’.

Interestingly 17% of researchers checked more than one response to this question. Many chose a combination of ‘No, I haven’t yet consider this’, and ‘No, but I would consider developing a formal plan in future (given time and support)’.

Some of the free text comments give varying insights into why the researchers gave the responses they did:

“I can see a need for it [Research Data Management Plan] for larger projects, but do not feel it worth the investment of time for my present projects”

“I send copies of my data files to others (post-doc, overseas collaborators, etc) as appropriate, but I don't know what they do with it.”

“Has not been enough time to give much thought to this issue.”

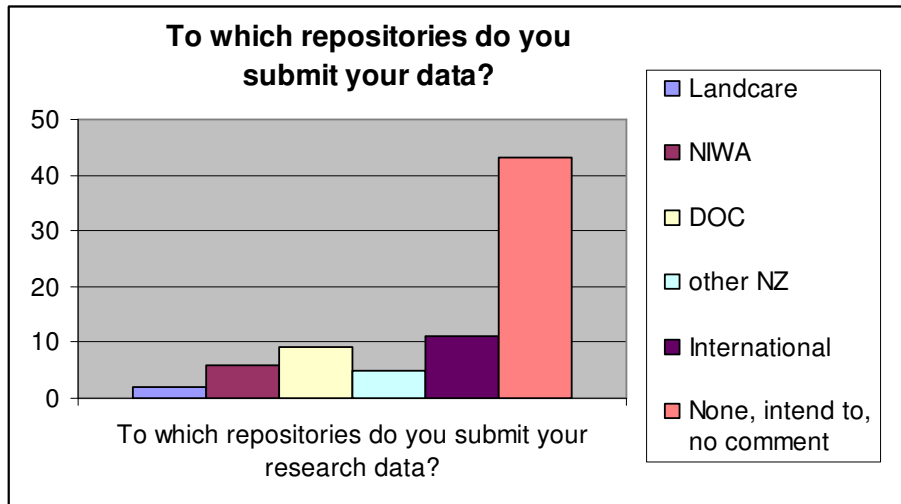
“I have considered this, but have not been able to find any support for this sort of action”

“My management plan is to transfer everything that I have collected since 1970, and saved from destruction during the Govt department destruction of the 1980s, to appropriate permanent archives [details provided]”

D. SOURCE AND DISCIPLINE REPOSITORIES

13. To which repositories – and how often - do you submit your research data?

a. To which repositories do you submit your research data?



Respondents selected multiple responses to this question.

A large percentage of the researchers surveyed do not currently submit their research data to any repository, either locally, within New Zealand, or internationally. 61% of respondents indicated that they either did not submit their data to any repository, did not do so currently, but intended to do so in future, or made not comment.

Of those who did currently submit to a repository, 15% submitted data to international repositories, such as GLORIA¹² and Genbank (both mentioned more than once). Those who submitted data to New Zealand repositories selected: DOC (13%); NIWA (8%); Landcare (3%) and other New Zealand repositories (7%).

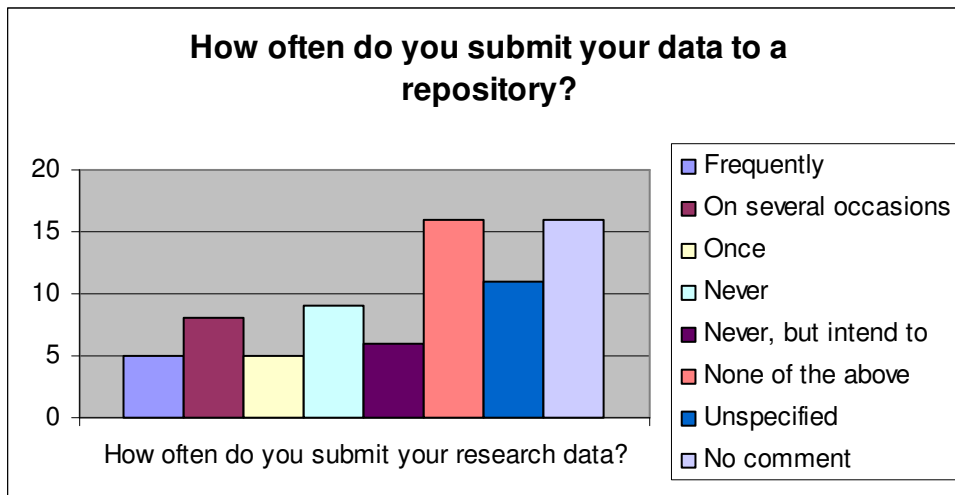
Only 5 researchers deposited in more than one repository and all but one of these deposited in both New Zealand and international repositories.

b. How often do you submit your research data to a repository?

7% of respondents indicated that they deposit their data into a repository (or repositories) 'frequently'. A slightly larger number, 11%, indicated they had done so 'on several occasions' and 7% had deposited data just 'once'. 15% indicated that they deposited data, but did not indicate how frequently they did so (unspecified).

The two highest responses were for 'None of the above' and 'no comment', each of which were selected by 23% of respondents.

¹² GLORIA: GLocal Observation Research Initiative in Alpine Environments



Respondents selected multiple responses to this question.

A number of free text comments, in support of this question (parts a. and b.), provide further insight into why a particular repository has been selected (or not), or suggest other repositories:

“Some information is submitted to Antarctica NZ from Antarctic projects”

“My data (such as it is) stays on the dOC system.”

*“Presumably you mean electronic & not published repositories
NZ Plant Database... is Nga Tipu o Aotearoa: NZ Plants the same as NZPD?”*

“I have only been back in NZ for 7 months. I suspect some data for a current project will go back to land care.”

“Never submitted data to any database. It does not apply to my data.”

“Have not need to yet (personally) on any of the projects I've been involved in, but may need to in future.”

“info may get in to the databases via other people that i share with but I rarely put data in myself.”

“never submitted any data to any of the repositories and very rarely use them”

“Have never submitted data to the above and am unlikely to do so”

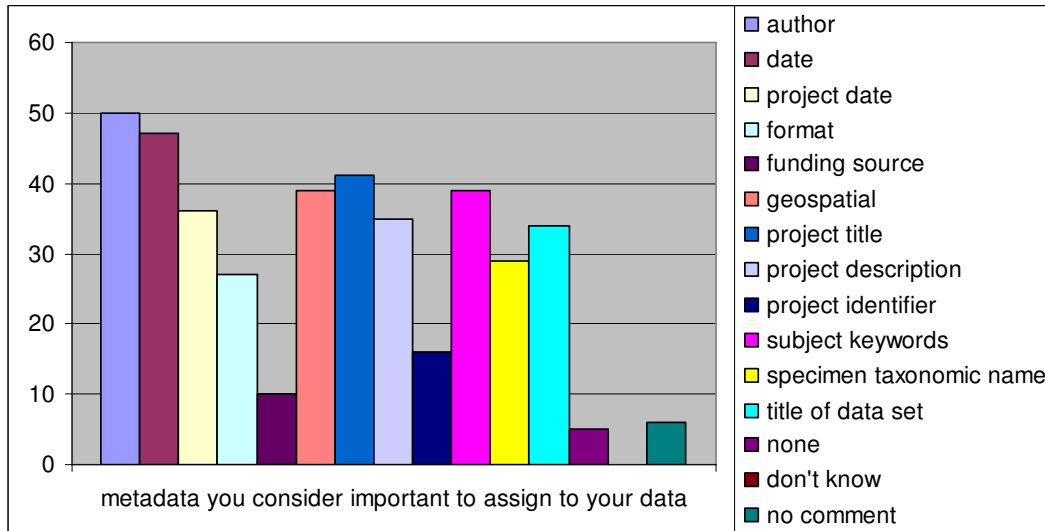
Other free text comments raise a couple of significant issues:

“The [X] sounded great but issues of access for external (university) users has never been resolved (despite promises).”

“Data is not in a convenient enough format to submit”

E. METADATA

14. Can you please indicate what types of metadata you consider important to assign to your data.



Respondents selected multiple responses to this question.

Most of the metadata options suggested were selected by approximately half (or more) of the respondents, with the exception of ‘Funding source’ and ‘Project reference number/identifiers’ (selected by just 14% and 23% respectively). ‘Don’t know’ was unchecked although 7% of respondents also chose not to check any box, and only one of these researchers included a comment:

Generally we use accepted International standards & metadata.

Other free text comments included the following metadata suggestions:

“Catalogue numbers (e.g., OUxxxxx)”

“identifier, altitude, habit notes”

“Band no. of marked birds”

“Methods of data collection”

“Good question...”

“Date (eg of data creation) ... of collection”

“Data description, production algorithm; Spatial Resolution (Image); Geographic Coordination System; Geographic extent; Version; Time period; Filename convention; Citation; Quality assessment”

“All data specified under ISOTC211 or FGDC Standards for Geospatial Metadata”

“Exactly which combination will vary with the data set in question”

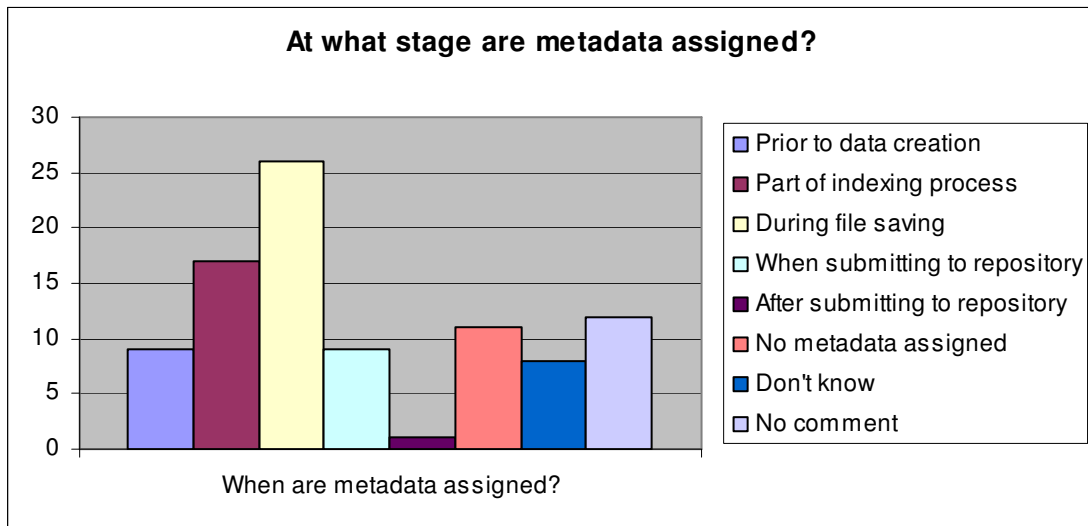
“I generally regard all this data as public information.”

“copyright issues (for film, etc)”

“which individuals/communities/ maori entities were involved”

“The more the better!”

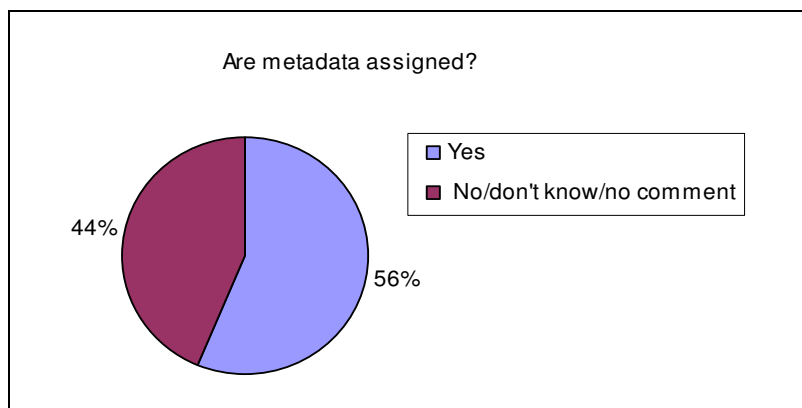
15. At what stage are metadata assigned to your research data?



Respondents selected multiple responses to this question.

The most preferred stage to assign metadata to the research data is ‘During file saving’ - 37% of researchers chose this option and 17% chose **only** this option. The second most preferred option, ‘As part of the indexing process for research data files’, was selected by 24% of researchers.

A significant number of researchers indicated some uncertainty about when, or even if, metadata is assigned – 44% of researchers indicated either ‘No metadata are assigned’, ‘Don’t know’ or gave no comment.



Some of the free text comments give further indication of how metadata are/are not assigned:

“This depends on the project”

“When data are transcribed from field notes into electronic form”

“I complete metadata according to evolution of dataset”

“During data creation”

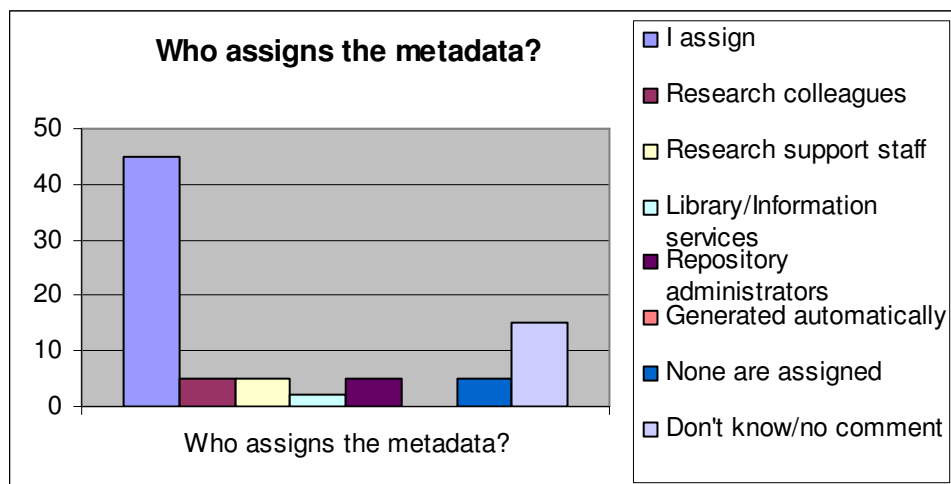
“A mix happens. Usually relates to my filing system.”

“during write up”

“When drafting a research grant application, contract or report”

“It really depends what it is, and it's patchy. While metadata are a great idea I don't always have time to assign all the files for photographs, for instance”

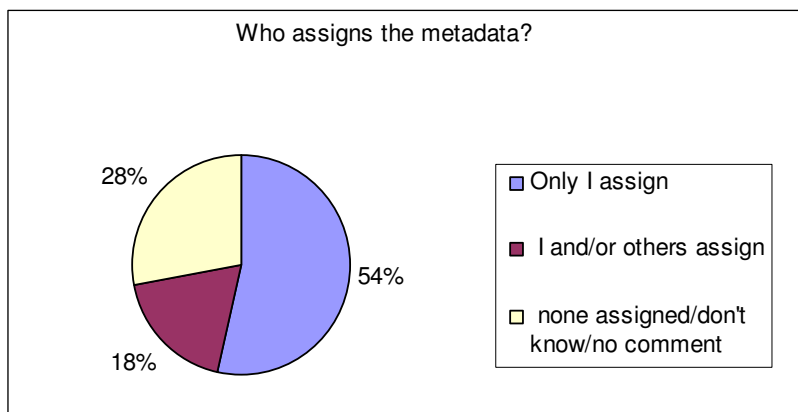
16. Who assigns metadata to your research data?



Respondents selected multiple responses to this question.

Most respondents (89%) selected only one response to this question. Of those who chose more than one response, all but one selected a combination of ‘I decide which terms to use and I assign them’ together with one (or more) other options.

Of the total responses, the most preferred response was ‘I decide which terms to use and I assign them’ (63%). As with the previous question, a significant number of researchers (28%) assigned either no metadata or were uncertain about who assigned the metadata (or made no comment).



54% of researchers selected **only** ‘I decide which terms to use and I assign them’ although the free text comments, associated with this response, contradict this statement somewhat:

“My research students”

“but get checked by team”

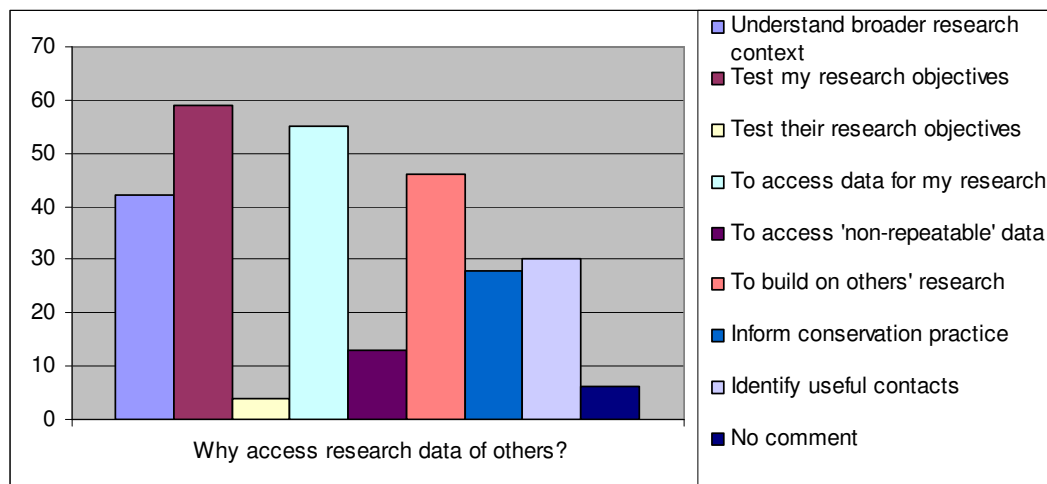
“In terms of Geographical datasets, a comprehensive Metadata Structure is suggested within specialized software such as ArcGIS”

“Metadata also designated by students responsible for current research project. No one person has an overall responsibility”

The students name their own files

F. DATA ACCESS AND SHARING

17. Why might you wish to access the research data generated by other researchers?



Respondents selected multiple responses to this question.

Almost all researchers chose multiple responses to this question and the selections were fairly evenly spread, with the exception of ‘To test the uniqueness and validity of **their** research objectives’, selected by only four researchers. ‘To test the uniqueness and validity of **my** research objectives’ was the most preferred option, chosen by 83% of respondents.

The next three most preferred responses were also selected by many researchers: ‘To access data that are useful or necessary to my research’ (77%); ‘To build on the work done by other researchers’ (65%) and ‘To understand the broader context and orientation of my research’ (59%). There were fewer free text comments associated with this question. These include:

“To test hypotheses I am interested in”, “To see what methods are useful.”

“But for my students rather than myself”

“to avoid repeating work already done!”

“To illustrate statistical methods, either to my students or in a research paper”

“I don't wish to access other researchers' data”

18. How would you normally access the research data of other researchers?

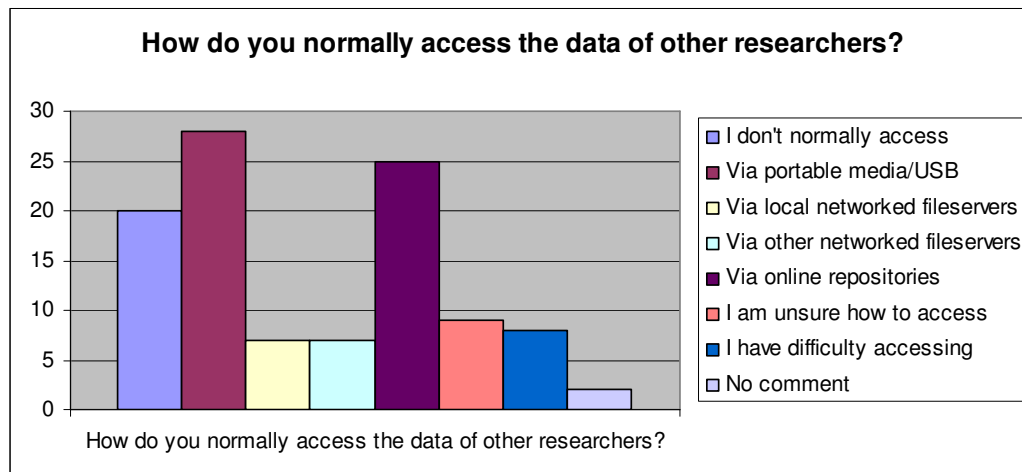
The two most strongly preferred options were ‘Through the exchange of data held in portable media’ (39%) and ‘Through online access to source/discipline repositories (35%). The third most preferred option was ‘I do not normally access others’ research data’ (28%).

Some researchers included free text comments which seem to contradict this third statement: (The use of the word ‘normally’ may be confusing and suggest varying interpretations of this question.)

“For specific projects that would benefit from another's research data, I would contact the researcher(s) involved.”

“I do not normally access others' research data...except postgraduate students under my supervision.”

“Occasionally through CRI monographs & reports and [underlined] through networks of colleagues”



Respondents selected multiple responses to this question.

The number of researchers who are unsure about how to access the data of others, or who experience difficulty accessing others’ data (or who make no comment), also add up to a significant 27% of all responses.

Free text comments focus primarily on other ways that the research data of others can be accessed, and direct ‘face-to-face’ communication is a recurring theme:

“Researchers' web sites.”

“By working collaboratively - thus sharing data”

“Request personally, by email. Research data sent (& received) as spreadsheets, pdf, etc. attachments (underlined) to email.”

“Contact them and ask for it. I feel that data are ultimately the property of the person who collected them, so permission must [underlined] be granted for data access.”

“I contact the person who ‘owns’ the data”

“Occasionally through CRI monographs & reports and through networks of colleagues”

“By personal contact”

“Through person-to-person contact primarily”

“Via hardcopy also eg maps, field notes, contact names”

“Docdm system is very good for this.”

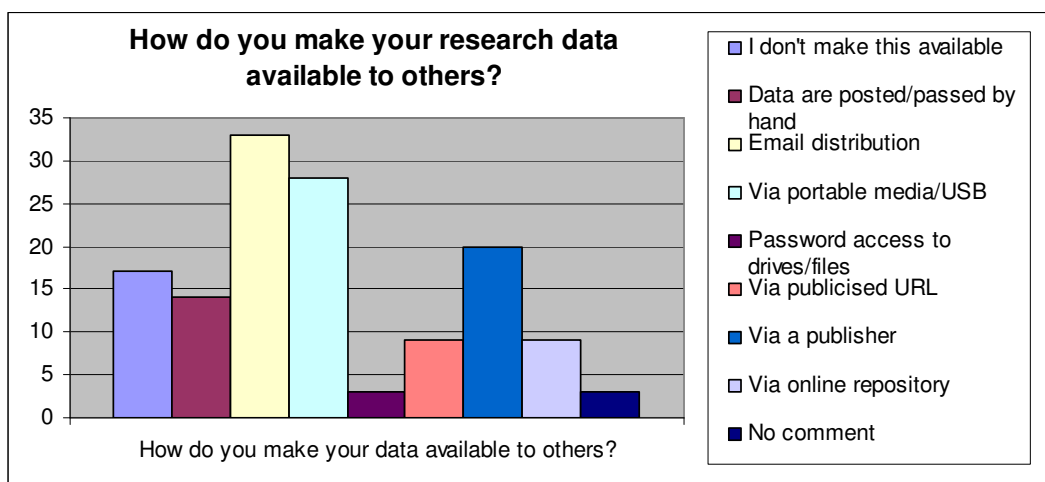
“They come to me for statistical advice”

“email attachments”

“library talking/meeting/emailing with researcher, community member, employees from other agencies (MFish, runanga, landcare, DoC, MfE, etc.)”

19. How do you make your research data available to others?

Email is the most preferred method for making research data available to others. This option was selected by 46% of respondents. ‘Through the exchange of portable media’ was also popular (39%).



Respondents selected multiple responses to this question.

Free text comments associated with ‘I don’t make my research data available to others’ (selected by 24% of respondents) include:

“I “don’t” is not really the right term -- few have requested access to data generated in my projects. When requested, I’m willing to transfer by whatever means agreed.”

“I don’t...normally...make my research data available to others”

“I havent been asked by anyone to share data”

“Communication research data through Departmental Seminars”

“But with help this could change”

Other free text comments relate to:

(a) Direct (face-to-face) sharing of data:

“At my discretion, following a request, data are usually emailed”

“Data are distributed via e-mail...if someone requests data”

“Only when we have discussed publishing rights, etc”

“But really only been requested by collaborators”

“Either by direct correspondence with other researchers or by journal publications.”

“If requested”

“oral presentations at conferences, meetings, etc.”

“The URL is publicised only to a few colleagues, so this is not really published. Otherwise through publications in journals”

(b) Indirect access/sharing of data

“Through the Otago IRR.”

“DVDs distributed - Books published

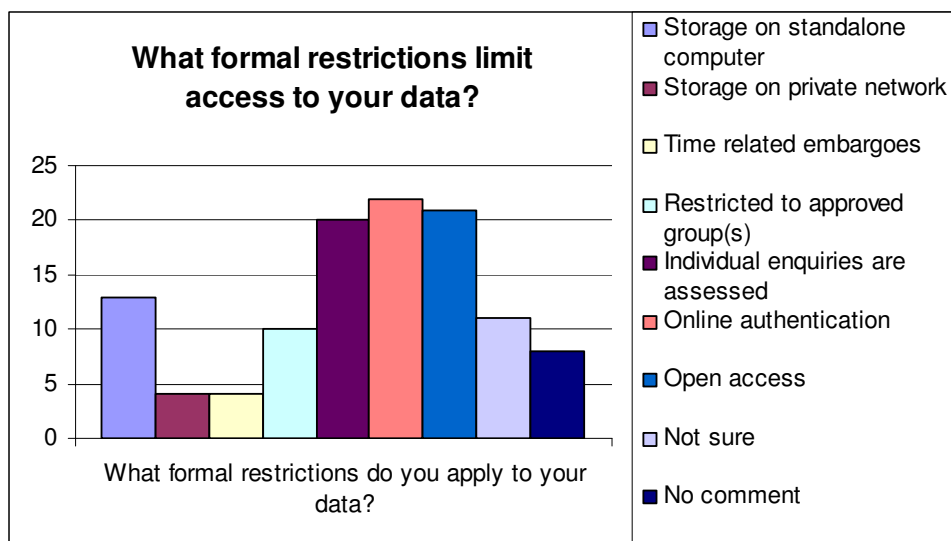
“Through networked file servers at my own institution”

“wiki”

“TreeBase; GenBank” [repositories suggested]

“Usually only final product not data is made available”

20. What formal restrictions do you apply to your research data to control access?



Respondents selected multiple responses to this question.

A significant number of researchers indicated some uncertainty about data restrictions. 27% of responses indicated either ‘Not sure what formal restrictions are applied’ or made no comment. Two of the respondents who selected ‘Not sure...’ also selected ‘No formal restrictions – there is open access.’

The most preferred response to this question was ‘Authentication of ID and password for online access’ (31%), closely followed by ‘No formal restrictions – there is open access’ (30%) and ‘Individual enquiries/requests for access are judged on their merits’ (28%).

Interestingly, two of these three most preferred responses emphasise security and permissions, while the third (open access) imposes no restrictions whatsoever. Of those who selected open access, 81% selected **only** open access, while the other 19% chose open access, in combination with other formal restrictions.

Free text comments support this combination of approaches:

“A mix - depending on the type of data and how it will be used by others.”

“Only restrictions on locations of species subject to predatory collecting.”

[Open access] *“to data / findings published / released on DVD”*

“I expect that my data is subject to the Official Information Act.”

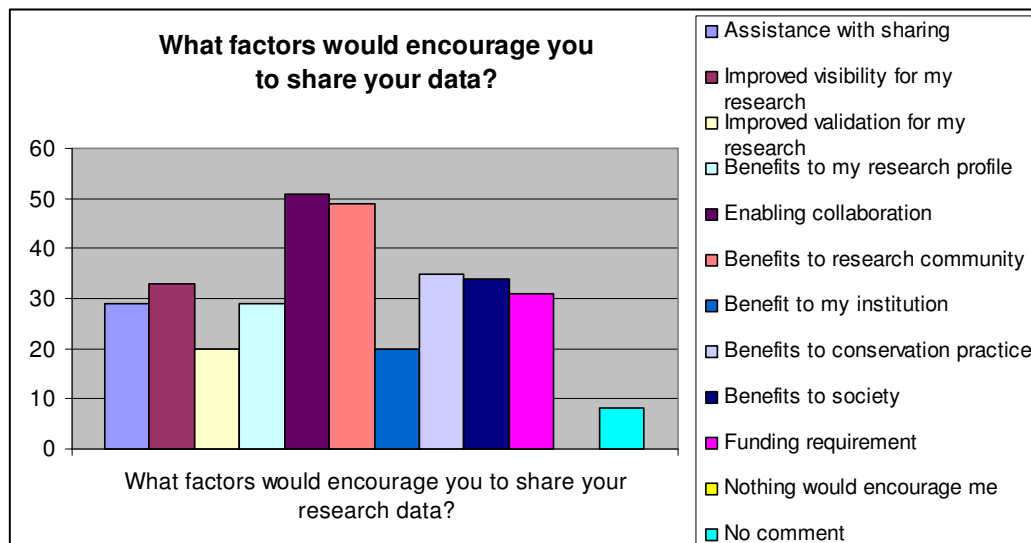
“Providing dataset to other researchers always involve a Memorandum of understanding stating their intention of use and possible counterparts (joint publications...)”

[Not sure...] *“But use has been confined to collaborators”*

“But a charge is made for any commercial requests”

“[X] research is based on open book policy, whatever research is going on can be enquired about and is usually passed on, but as the [X] server (which holds most of our data) is based with ITS and needs ID & password its not accessible for everyone”

21. What factors would encourage you to share your research data?



Respondents selected multiple responses to this question.

The two most preferred options are possibly not surprising, ‘Enabling collaboration and contributions by others’ (72%) and ‘Potential benefits to the research community’ (69%). However, ‘Demonstrable benefit to my institution’ attracted only 28% of responses, suggesting a different relationship between researchers and their ‘research community’ and researchers and their ‘institution’.

‘Improved level of validation for my research findings’ also attracted just 28% of responses, while the remaining options were each selected by between 41 – 50% of all respondents.

None of the respondents selected ‘Nothing would encourage me to share my research data’ and the free text comments are generally positive about data sharing, although one of these comments also raises an important concern:

Better access to [X]. Having supplied data in the past, I'm loathe to do so again unless I know that I can access the system."

"I don't need encouragement -- I'm already willing."

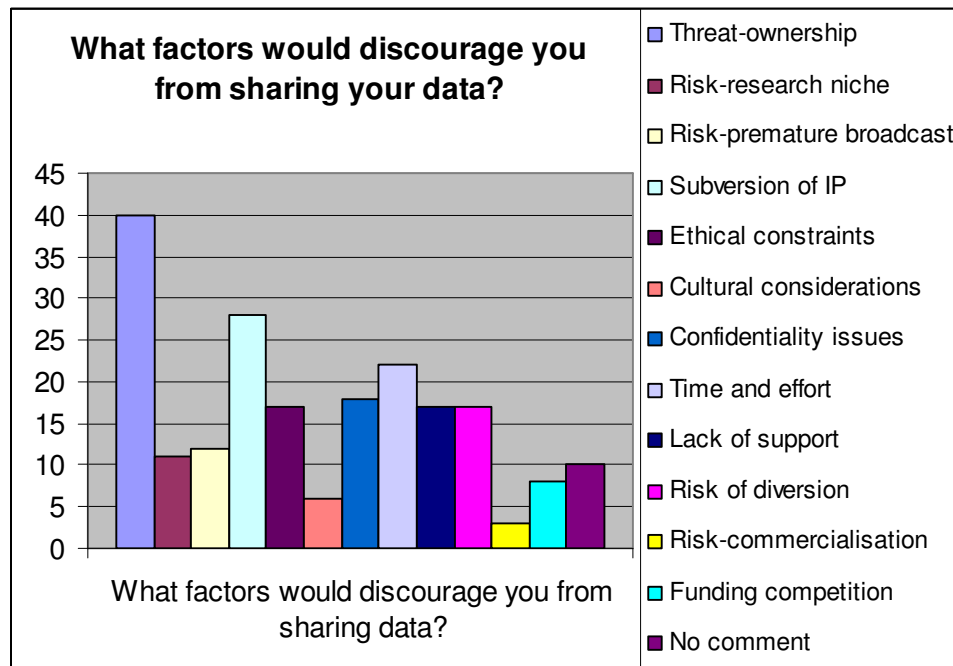
"I'd be willing to share - if asked - but I don't think I ever have been!"

"I publish my results, then all can read it."

"And interest from other people that are mutually helpful."

"because its part of the project our research will be available on our web-page anyway for my personal work (publications on past work) I just need some time to write them up"

22. What factors would discourage you from sharing your research data?



Respondents selected multiple responses to this question.

The top three factors which would discourage researchers from sharing their data were a combination of privacy issues and practicality: ‘The threat of loss of ownership’ (56%);

‘Subversion of intellectual property rights, including copyright’ (39%) and ‘The time and effort required to enable sharing’ (31%).

Of least concern were: ‘Cultural considerations relating to my research’ (8%); ‘Risk to commercialisation opportunities’ (4%) and ‘Increased competition for funding’ (11%).

A considerable number of researchers (14%) chose to give no response to this question. It is impossible to know whether these respondents didn’t know how to respond, didn’t care or didn’t have time to think about this question.

Free text comments associated with these responses included the following suggestions:

“potential for misinterpretation of data”

“Not necessarily at liberty to share data collected under contract.”

*“Subversion of intellectual property rights, including copyright...of others
Risk to commercialisation opportunities...of others”*

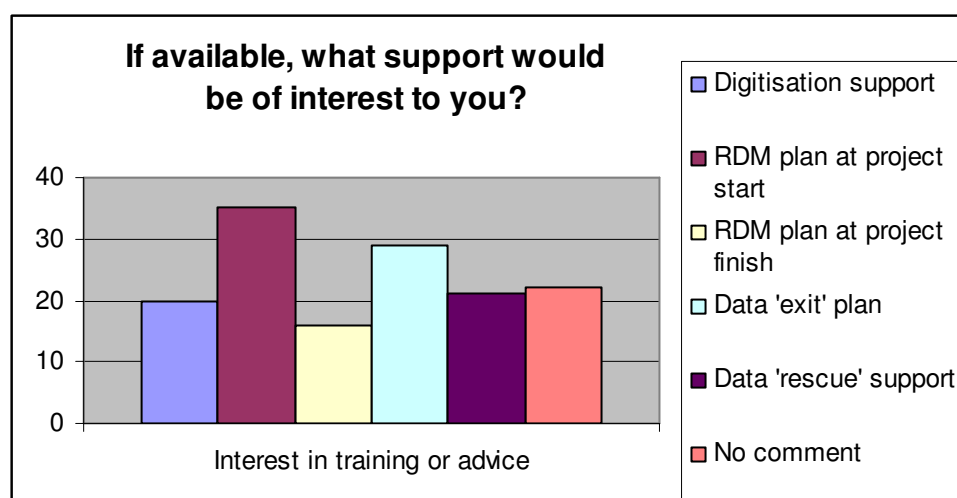
“Raw data and computer files have been made freely available to researchers in this field and the sharing of information has been encouraged throughout this programme”[example given]

“Feeling quite non-competitive as a PhD student at this time - if I had something worth sharing, I would”

*“Predatory or unpleasant behaviour of the other person.
Cynical use against conservation or social interests.”*

G. SUPPORT AND FURTHER INTEREST

23. If support were available, would you be interested in training or advice on any of the following?



Respondents selected multiple responses to this question.

The most preferred support option was ‘Creating a research data management plan at the beginning of a project’, selected by 49% of respondents. Also popular was ‘A data ‘exit’ plan (for retiring academics or departing academics or postgraduate students)’ (41%).

Fewer respondents were, unsurprisingly, interested in ‘Creating a research data management plan after the project has finished’ (just 23%), while interest in the remaining two options was slightly higher: ‘Digitisation advice, tools and services’ (28%) and ‘Data ‘rescue’ for older digital materials, such as data on older media or migration of data from legacy (now unreadable) systems’ (30%).

Significantly, 31% of respondents chose not to select any of the options, although several did provide comments, which suggested (for these researchers at least) that not choosing a support option was a result of uncertainty or lack of time, rather than lack of interest in support with managing their data:

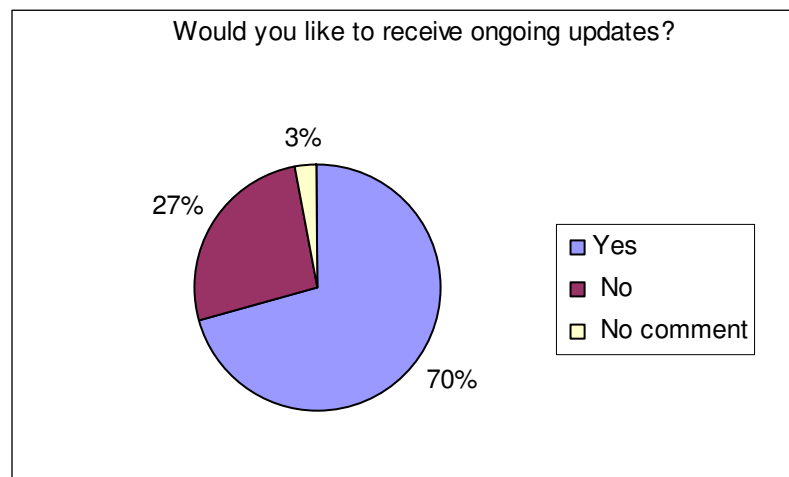
“Unlikely to personally have time for this”

“Don't know”

“Not sure”

“Probably not - lack of time, etc”

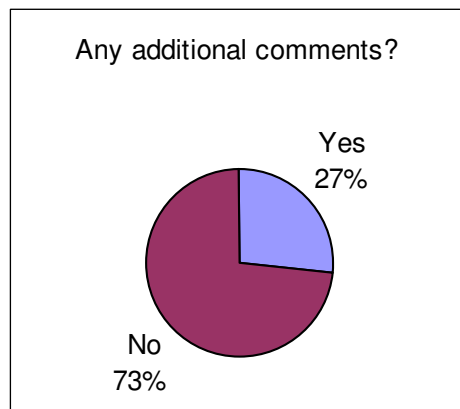
24. Would you like to receive ongoing email updates about the current Otago Biodiversity Project and the findings of this research data questionnaire?



The majority of researchers (70%) expressed ongoing interest in the current Otago Biodiversity Project and the findings of this research data questionnaire.

Each respondent selected a single response to this question.

25. Please feel free to add any other comments regarding data management, long term data storage and access, digitisation, training, etc



A summary of ‘other comments’ is included in *Appendix D*.

Many of these comments were of considerable length, giving researchers an opportunity to share, in greater detail, some of their concerns and suggestions about managing data more generally, as well as the issues surrounding management of their own research data.

Questionnaire responses – summary observations

This section summarises the findings from the questionnaire responses (presented in detail in the previous section) and adds further observations, in terms of formal data management.

A. Constituency [1 and 2]

While considerable attempts were made to gain a representative sample of University of Otago researchers, working in the area of biodiversity, the majority of respondents (74%) were established, rather than emerging, researchers. This does not undermine the validity of the responses, but it is worth noting, especially as the interests and concerns of established researchers are likely to differ from those who are embarking on an academic career, and have yet to establish significant amounts of data. As discussed, 7% of responses were also received from Otago DOC managers.

B. Project aims [3]

More than a quarter of respondents felt that formal management of biodiversity research data would be of use to them. 43% suggested this would be ‘useful,’ and a further 21% that this was interesting (although perhaps not particularly useful). Only 6% indicated that research data management was ‘of no interest to me’. The associated free text comments (and also the ‘other comments’, question 25) suggested that there may be even greater support for formal data management if the Project scope were different, or if the researcher had a better understanding of what data management might mean for their research.

C. Your research data

Non-digital data [4]

The survey group, many of whom are Life Scientists, still have quantities of non-digital data, particularly field notes, photographs, observational records and reports. The questionnaire did not quantify this non-digital data, but as the majority of respondents indicated that their data would be of value for more than 10 years, it seems likely that much of this non-digital material continues to have value and would be worth digitising. This raises a number of questions for potential data managers, not least of which is cost of digitising widely differing non-digital formats.

Digital data

There were a number of questions around digital data, generated by the researchers, including types of data, formats, whether or not this data is a combination of formats, data size and life span, back-up systems for this data, and current data management practices.

Data types, formats and size [5-8]

With regard to data types and formats, almost all respondents had multiple kinds of data, in a variety of formats. Especially popular are datasets, digital images and derived data, utilising spreadsheets, word processed files and image files. Numerous other data types and formats were also suggested, in the free text comments, and few respondents were either ‘unsure’ or gave no comment. The responses suggest that researchers are confident in using digital data, however the high uptake of proprietary software is something to be aware of in terms of long term management of this data (e.g. obsolescence, ongoing costs).

Generating data in a combination of different data formats is common – 82% of data was reported as being ‘often’, ‘sometimes’ or ‘potentially’ a combination of different formats. What data format combinations are most prevalent was not explored, but again, having data in more than one format has implications for formal data management.

'Data size' and 'length of time this data has value' generated responses consistent with what is reported in the literature on this topic, within the Life Sciences. Life scientists often work on small projects, with limited funding, generating small, localised collections of data, i.e. "the Long Tail of Science" (Palmer, 2007¹³) and much of this data retains its value over time. 52% of respondents to the Otago questionnaire indicated that they held less than 1GB¹⁴ of data and 82% held less than 1TB of data. Compared with the 'Big Sciences',¹⁵ these are relatively small quantities of data.

Data life, storage [9-10]

The quantities of data may not (typically) be large in the Life Sciences, but the data is held to be of value over time (illustrations of long term studies are given in the free text comments). 66% of respondents indicated that their data had value for up to, or more than, 10 years. A further 25% didn't know how long their data would be of value or provided no comment.

If data retains its value over time, it is important that this data is stored securely. Initial findings of this questionnaire suggest that researchers appreciate the value of secure storage and the importance of back up systems - 87% of respondents indicated their data was stored in more than one location. However, around only half the respondents stored data either on their Local Area Network (LAN), 'Offsite storage' (just 13%) or used a 'Third Party' (17%). For many, 'back up systems' meant the use of an external hard drive, CD, DVD or Flash drive.

Data management [11-12]

Researchers manage their own data – 96% checked 'I manage' and 64% checked *only* 'I manage'. A third of researchers indicated that they manage their data with others, often their postgraduate students. Just 3% indicated their data was managed by a third party.

Researchers clearly work closely with their data, throughout the research cycle and also the entire data management and curation life cycle (see DCC Curation Lifecycle Model).¹⁶ How well this data is managed, however, was not explored, but it seems reasonable to suggest that, with such a high % of researchers managing their own data, there will be considerable variety in how well this data is managed, particularly over the longer term. It also raises the question of how much interest there might be in a more standardised approach to data management and third party support, when so much is currently being done by the individual.

A more structured approach to managing researchers' data, in the form of a 'Research Data Management Plan' (DMP) was not something many researchers had yet considered. Almost half of the respondents indicated they 'haven't yet considered this'. Just two respondents currently had a DMP in place¹⁷. While many respondents indicated they lacked knowledge or time to consider a DMP, almost 30% of respondents indicated that they would consider a DMP. Only seven respondents indicated they did not 'see the need for this'.

Free text comments and also the high number of 'don't know', 'unsure' and 'no comment' responses throughout this questionnaire, suggest that there is a need to provide more information about what 'data management' means in practice (and may mean in future) to individual researchers. The issues around data management are not widely understood and remain unclear to many.

¹³ Palmer, C.L., Cragin, M.H., Heidorn, P.B. & Smith, L.C. (2007) *Data curation for the Long Tail of Science: the case of Environmental Sciences*, Urbana-Champaign: University of Illinois.

¹⁴ KB = 1 thousand bytes; MB = 1 million bytes; GB = 1 billion bytes; TB = 1 trillion bytes

¹⁵ "Big Science usually implies one or more of these specific characteristics: big budgets; big staffs; big machines; big laboratories..." [predominately physics and astronomy] Retrieved 1 Sept 2008 from www.wikipedia.org/wiki/Big_Science

¹⁶ DCC Curation Lifecycle Model: www.dcc.ac.uk/FAQs/dcc-curation-lifecycle-model/

¹⁷ For more on Research Data Management Planning see ANU's DMP site: <http://ilp.anu.edu.au/dm/>

D. Source and discipline repositories [13]

The majority of respondents (61%) do not currently submit their research data to any repository¹⁸, either New Zealand (e.g. those managed by DOC, NIWA or Landcare) or international (e.g. OBIS, GLORIA¹⁹). Some of the free text comments indicate further uncertainty about data deposit (data format, value of depositing data). Of those who do currently deposit data in a repository, only 7% do so frequently, while 15% give no indication of how frequently they deposit.

These results suggest that University of Otago researchers do not, routinely, deposit data into repositories. *Why* they do not deposit data is, again, not explored in this questionnaire, although free text comments provide selected insights.

E. Metadata [14-16]

Most metadata options suggested in question 14 were selected by most respondents and there seems to be little uncertainty associated with questions about metadata ‘types’. No respondent selected ‘Don’t know’ and many researchers included further metadata suggestions in the free text. However, when asked when metadata was assigned to the researchers’ data, the responses indicated widely differing practices and 44% of researchers indicated (together with other responses) that either no metadata was assigned, didn’t know when metadata was assigned or gave no comment. These differing results may suggest a disconnection between knowledge, of what ‘should be done’, and practice or what is ‘actually done’.

When questioned about *who* assigns the metadata to the researchers’ data, again the results suggest that this ‘data management’ activity is largely done by the individual researchers. Two thirds of the researchers decide which metadata terms to assign, and 54% of researchers are the only ones to do so. A significant 28% of respondents indicated that no metadata are assigned, don’t know or make no comment. Again, with decisions about metadata being made primarily by the individual researchers, there are implications (as previously discussed) with implementing a more standardised approach to managing the research data.

F. Data access and sharing

Accessing others’ research data [17-18]

The majority of respondents would wish to access the research data of others to test the uniqueness and validity of *their own* research (83%), to access useful data (77%) and to build upon the research of others (65%). Other reasons were also selected, but identifying real benefits to the individual’s own research objectives are, understandably, the most compelling.

The three most selected responses to ‘how respondents normally access the research data of others’ suggests a variety of approaches and raises a number of issues. ‘Exchange of data held in portable media’ was selected by 40% of respondents. While this may be a quick and easy way to share limited amounts of data, this method also carries the risk of data corruption and file damage. ‘Through online access to source/discipline repositories’ was selected by more than a third of respondents. This figure seems to suggest that while researchers do not generally *deposit* data in repositories (qu.13), they are comfortable *accessing* data from these sources. The third most preferred choice, ‘I do not normally access others’ research data’ (28%) implies that a significant number of researchers do not access others data. However, the accompanying free text comments suggest that, for some at least, the situation is otherwise. The use of the word ‘normally’ in this question may have caused some confusion, affecting some of the responses.

¹⁸ “A repository is a place where data or specimens are stored and maintained for future retrieval”

Retrieved 1 Sept 2008 from: <http://en.wikipedia.org/wiki/Repository>

¹⁹ OBIS: Ocean Biographic Information System: www.iobis.org/; GLORIA: Global Observation Research Initiative in Alpine Environments: www.gloria.ac.at/

Sharing your research data [19-20]

As in the previous question, 'Through the exchange of portable media' was a popular method for making research data available, in this case, *to others*. This option was selected by almost 40% of respondents, just slightly less than the most preferred method of sharing, which was 'via email' (46%). Again the issues of portable media sharing are of concern, while sharing via email may present issues of limits to size and speed of data transfer.

Both this, and the previous question (concerning how research data is shared), generated a large number of free text comments. Reflecting on these, certain 'themes' emerge, including: the continued importance of personal contact and direct 'face-to-face' sharing; the ongoing prevalence/popularity of email and the continuing importance of print (non-digital) formats.

It is not always possible to share all research data at all times, and researchers apply a variety of restrictions to accessing their data. Interestingly, two of the three most preferred responses to the question of 'formal data restriction', 'Authentication/password for online access' and 'Individual enquiries are judged on their merits' (each of which was selected by a third of respondents), emphasise security and permissions, while the third, 'open access' imposes no restrictions whatsoever. This latter option was selected by 28% of respondents and 81% of this group selected *only* open access. Over a quarter of respondents were either unsure what formal restrictions applied or gave no comment. The free text comments reflect this mixed approach and slight degree of uncertainty.

Factors encouraging/discouraging you to share [21-22]

Most notable, perhaps, is that no respondent selected 'Nothing would encourage me to share my research data', in response to 'What factors would *encourage* you to share your research data?' Half of the respondents selected most of the suggested options and two factors were particularly popular, 'Enabling collaboration' (72%) and 'Potential benefits to the research community' (69%). 'Demonstrable benefit to my institution' was selected by just 28% of respondents, suggesting a differing relationship between researchers and their 'research community' and researchers and their 'research institution'.

Responses to the question of why you (the researcher) might be discouraged from sharing your research data, suggests that the 'threat of loss of ownership' (over 50%) and 'subversion of IP rights' (almost 40%) are the most concerning factors. Interestingly 'lack of time and effort required' and 'lack of information and support' also rank fairly highly (31% and 24% respectively). This interest in 'practical' barriers to data sharing suggests that *not* sharing data may not simply be a matter of ownership and privacy (as is often suggested).

G. Support and further interest [23-25]

A range of information and support options were selected by two thirds of respondents. Almost 50% of researchers indicated interest in 'creating a research data management plan at the beginning of a project' and 41% were interested in 'a data 'exit' plan'. (This may be in response to the researchers' varying situations, with regard to their research careers). 31% of respondents choose to make no comment, but whether this indicates that they want no support, or that they are unsure about what support they would find helpful, is not clear.

Again the supporting free text comments and 'other comment's' suggest a combination of lack of time and understanding of the issues of 'data management' may also be a barrier to researchers seeking support with managing their data.

The majority of respondents (70%) expressed ongoing interest in the current Otago Biodiversity Project and the findings of this research data questionnaire.

An edited list of 'other' researcher comments is included in *Appendix D*.

Appendix A: Otago Biodiversity Project Questionnaire

BACKGROUND

In 2007 a series of informal interviews, concerning research data management and curation²⁰, were held with life scientists at the University of Otago. These conversations highlighted the following:

- Internationally, New Zealand organisations and individuals are playing significant roles in the creation of databases and infrastructures, such as GBIF²¹, to improve management of, and access to, biodiversity information.
- Nationally, organisations such as Landcare Research and DOC are driving numerous initiatives – managing biodiversity databases and repositories and ensuring access to, and interoperability between, these resources.
- Locally, at the University of Otago, much biodiversity information is non-digital, locally-stored and difficult to access.

The Otago Biodiversity Project aims to investigate the feasibility of managing, and making available, University of Otago and regional biodiversity research data to empower research, further enable collaboration and inform conservation practice.

DEFINITIONS (within the context of this document)

- Database: information stored on a computer, which can be searched and updated
- Interoperable: ability of different computer systems to exchange information
- Open Access (OA): electronic resources are visible and available to all
- Research Data: facts, records and observations (numerical, textual or visual) which inform research intended for publication
- Digital/Online Repositories
 - Discipline Repository: electronic storehouse for subject related information e.g. Bioweb and herpetofauna databases
 - Institutional Research Repository (IRR): electronic storehouse for institutional research outputs, typically ‘high-end’ unpublished e-prints (Otago’s IRR is viewable through KRIS²²)
 - Source Repository: electronic storehouse for research data

Acknowledgement

This questionnaire is based on the *Project StORe survey: cross-discipline report* (2006), Edinburgh Research Archive: www.era.lib.ed.ac.uk/handle/1842/1419
Kind thanks to Graham Pryor for his advice and access to this report.

Permission was also granted to adapt elements from two further surveys: *Data Needs Assessment Survey* (2008) – thanks to Melissa Cragin, University of Illinois and also the University of Queensland survey (2007), available via the APSR website: www.apsr.edu.au/currentprojects/data_management_at_uq_blank_survey.pdf.

²⁰ *Data curation is the active and on-going management of data through its lifecycle of interest and usefulness to scholarship, science, and education. Data curation activities enable data discovery and retrieval, maintain its quality, and provide for re-use.* From Data Needs Assessment Survey, p.2, 2008, University of Illinois

²¹ Global Biodiversity Information Facility <http://www.gbif.org/>

²² Kiwi Research Information Service <http://nzresearch.org.nz/>

A. CONSTITUENCY

1. Please identify your own role by selecting from the following list:	
	University academic staff
	University research assistant
	Postgraduate student
	Contracting researcher
	Independent researcher
	Visiting researcher
Other (<i>please specify</i>)	

2. Please indicate your area of research:
--

B. PROJECT AIMS

This project aims to enhance scholarly research and collaboration by identifying Otago biodiversity research data and planning for its storage, management and access.

3. How would formal management of biodiversity research data be useful to you?	
	Of significant advantage to my work
	Useful but not of major significance
	Interesting but not particularly useful
	Of no interest to me
	Not sure what this means
	Don't know
Other (<i>please comment as fully as possible</i>)	

C. YOUR RESEARCH DATA

4. What kinds of non-digital data do you generate or collect for your research (or have you generated or collected in the past)? Please check all that apply:

	Aerial photographs or maps
	Audio or video tapes
	Catalogued specimens
	Field notes
	Index cards
	Laboratory notes
	Observational records
	Photographs
	Print data
	Reports
	Slides or film strips
	Specimens or field samples
	None
Other (<i>please specify</i>)	

5. What kinds of digital data do you generate or collect for your research? Please check all that apply:

	Audio or video
	Bibliography
	Databases
	Datasets
	Derived data (eg table or graph)
	Digital images
	Digital objects
	Drawings, plots
	Geospatial data
	Plans, maps
	Raw, experimental data
	Remote sensing data
	Statistical data
	Text-based files (eg unpublished reports)
	None
Other (<i>please specify</i>)	

6. In what formats are these digital information sources held? Please check all that apply:

	Computer-aided design (CAD)
	Database files (eg Filemaker, MS Access, MySQL, SQL Server)
	Extensible mark-up language (XML)
	Geographical information systems (GIS)
	Hypertext mark-up language (HTML)
	Image files (eg .jpg, .tif, .bmp, .gif)
	Plain text (.txt)
	Portable document format (.pdf)
	Reference management software (eg EndNote, Zotero)
	Rich text files (.rtf)
	Spreadsheets (eg Excel/.xls)
	Statistical software (eg SPSS)
	Word processed files (eg Word/.doc)
	Unsure of the format
Other (<i>please specify, giving as much detail as possible</i>)	

7. Are the research data you generate sometimes a combination of different data formats?²³

	Often
	Sometimes
	Rarely
	Never
	Potentially
Other (<i>please comment</i>)	

8. How large (in total) are your digital research data? Please estimate:

	Less than 100MB
	100MB – 1GB
	1GB – 1TB
	Greater than 1TB
Don't know (<i>please comment</i>)	

²³ “Combination data set: a set of data generated from a specific piece of research that is produced and presented in a mixture of formats. For example, a report that is written in a format that has a .doc extension but also includes charts in .xls format and images saved as .jpg files, or such an entity as a Geographical Information System (GIS)” (*Project StORe*, p.20)

9. How long do you think your research data will have value?	
	Up to 5 years
	Up to 10 years
	More than 10 years
	Don't know
Other (<i>please comment as fully as possible</i>)	

10. What data storage and back up system(s) do you currently have in place? Please check all that apply:	
	Own computer hard drive
	External hard drive
	CD or DVD
	Flash drive/USB
	Local area network
	Offsite storage
	Third party (including commercial data storage)
	Don't know
Other (<i>please specify</i>)	

11. Who currently manages your data? Please check all that apply:	
	I manage my own data
	Research project manager
	Designated person on project
	External project partners
	ITS
	IT staff within my department, centre or research institute
	Research Assistant
	Don't know
Other (<i>please specify</i>)	

12. Do you currently have a formal Research Data Management Plan in place and, if not, please indicate the reasons. Please check all that apply:

<input type="checkbox"/>	Yes, I have a Data Management Plan currently in place
<input type="checkbox"/>	No, I haven't yet considered this
<input type="checkbox"/>	No, I don't see the need for this
<input type="checkbox"/>	No, I don't have time for this
<input type="checkbox"/>	No, but I would consider developing a formal plan in future (given time and support)
<input type="checkbox"/>	I don't know enough about this topic to respond
<input type="checkbox"/>	No comment
Other (<i>please comment as fully as possible</i>)	

D. SOURCE AND DISCIPLINE REPOSITORIES

13. To which repositories – and how often - do you submit your research data? Please check all that apply:

13a. Landcare Research e.g. but to	frequently	on several occasions	once	never	never, intend
Animal Survey Reports					
Ant Distribution Data					
Ecological Traits of NZ Flora					
Geospatial Data Integration Portal					
LENZ: Land Environments of NZ					
New Zealand's Virtual Mycota					
Nga Tipu o Aotearoa: NZ Plants					
Nga Tipu Whakaoranga: People Plants					
NVS: National Vegetation Survey					
NZAC: NZ Arthropod Collection					
NZBRN: NZ Biological Recording Network					
NZFungi Database					
NZOR: New Zealand Organisms Register					
None of the above <input type="checkbox"/>					
Other (<i>please specify</i>)					

13b. NIWA e.g.	frequently	on several occasions	once	never	never, but intend to
Aquatic Plants Database					
FBIS: Freshwater Biodata Information System					
NZFFD: NZ Freshwater Fish Database					
None of the above <input type="checkbox"/>					
Other (<i>please specify</i>)					

13c. DOC e.g.	frequently	on several occasions	once	never	never, but intend to
ARDS: Amphibian & Reptile Distribution Scheme					
Bioweb Herpetofauna (now includes ARDS)					
Bioweb Bird Banding					
Bioweb Threatened Plants					
Bioweb Weeds					
NHMS: Natural Heritage Management System					
None of the above <input type="checkbox"/>					
Other (<i>please specify</i>)					

13d. Other New Zealand e.g.	frequently	on several occasions	once	never	never, but intend to
BUGZ: Bibliography of NZ Terrestrial Invertebrate					
Guide to NZ Soil Invertebrates					
NZ Fossil Record File (Geol Soc of NZ & GNS Science)					
NZ National Forestry Herbarium Database (Ensis)					
NZ Threatened Plants & Fungi (NZ Plant Conservation Network)					
NZRBN: NZ Recognised Bird Names (Ornithological Soc of NZ)					
None of the above <input type="checkbox"/>					
Other (<i>please specify</i>)					

13e. International e.g.	frequently	on several occasions	once	never	never, but intend to
AmphibiaWeb					
ARKive					
Avano: Marine & Aquatic Sciences					
BHL: Biodiversity Heritage Library					
CABI Bioscience databases					
Catalogue of Life					
Global Invasive Species Database					
GLORIA: GLObal Observation Research Initiative in Alpine Environments					
OBIS: Ocean Biographic Information System					
None of the above <input type="checkbox"/>					
Other (<i>please specify</i>)					

E. METADATA

The term metadata refers to the information or labels that you use to identify and describe your data. The principal purpose of metadata is to make it easy to recognise, access and retrieve data. A familiar example of this would be a library catalogue, which is a collection of metadata records.

14. Can you please indicate what types of metadata you consider important to assign to your data. Please check all that apply:	
<input type="checkbox"/>	Author/date creator name(s)
<input type="checkbox"/>	Date (eg of data creation)
<input type="checkbox"/>	Dates of project
<input type="checkbox"/>	Format (eg PDF or HTML)
<input type="checkbox"/>	Funding source
<input type="checkbox"/>	Geospatial location
<input type="checkbox"/>	Project title
<input type="checkbox"/>	Project description
<input type="checkbox"/>	Project reference number/identifiers
<input type="checkbox"/>	Subject keywords
<input type="checkbox"/>	Specimen taxonomic name (eg <i>Calloria inconspicua</i>)
<input type="checkbox"/>	Title of data set
<input type="checkbox"/>	None
<input type="checkbox"/>	Don't know
Other (<i>please specify</i>)	

15. At what stage are metadata assigned to your research data? Please check all that apply:

	Prior to data creation
	As part of the indexing process for research data files
	During file saving
	When submitting data to the repository
	After submission of my data to the repository
	No metadata are assigned
	Don't know
Other (<i>please comment as fully as possible</i>)	

16. Who assigns metadata to your research data? Please check all that apply:

	I decide which terms to use and I assign them
	Research colleague(s) assign metadata on the team's behalf
	Research support staff assign metadata on the team's behalf
	Metadata are assigned by library/information services staff
	Metadata are assigned by the repository administrators
	Metadata are generated automatically
	No metadata are assigned
	Don't know
Other (<i>please specify</i>)	

F. DATA ACCESS AND SHARING

17. Why might you wish to access the research data generated by other researchers? Please check all that apply:

	To understand the broader context and orientation of my research
	To test the uniqueness and validity of my research objectives
	To test the uniqueness and validity of their research objectives
	To access data that are useful or necessary to my research
	To access data gathered from non-repeatable 'experiments' (eg volcanic eruption)
	To build on the work done by other researchers (eg long term research on a particular species or ecosystem)
	The potential to inform conservation practice
	To identify useful contacts
Other (<i>please comment as fully as possible</i>)	

18. How would you normally access the research data of other researchers? Please check all that apply:

	I do not normally access others' research data
	Through the exchange of data held in portable media (disks, CD-roms, USB drives, etc.)
	By access to networked file servers at my own institution
	By access to networked file servers at other institutions
	Through online access to source/discipline repositories
	I am unsure how I might access the research data of other researchers
	I have difficulty accessing the research data of other researchers (due to subscription or other barriers)
Other (<i>please comment as fully as possible</i>)	

19. How do you make your research data available to others? Please check all that apply:

	I don't make my research data available to others
	Data are posted or passed by hand in printed format
	Data are distributed via e-mail
	Through the exchange of portable media (disks, CD-roms, USB drives, etc.)
	By the allocation of passwords to network drives or data files
	By the provision of a publicised URL
	Via a publisher
	Through a source/discipline repository (eg Bioweb)
Other (<i>please comment as fully as possible</i>)	

20. What formal restrictions do you apply to your research data to control access? Please check all that apply:

	Storage of data on standalone computers
	Storage of data on a private network/intranet
	Time related embargoes (eg after the end of the project)
	Restricted to immediate research team/programme members/approved list
	Individual enquiries/requests for access are judged on their merits
	Authentication of ID and password for online access
	No formal restrictions – there is open access
	Not sure what formal restrictions are applied
Other (<i>please comment as fully as possible</i>)	

21. What factors would encourage you to share your research data? Please check all that apply:

	Assistance with sharing my research data ie collaboration with library and/or technical staff
	Improved visibility for my research
	Improved level of validation for my research findings
	Demonstrable benefit to my research profile (improved status, future funding or new research prospects)
	Enabling collaboration and contributions by others
	Potential benefits to the research community
	Demonstrable benefit to my institution (research presence, income, etc)
	Potential to inform conservation practice
	Potential benefits to society
	Requirement of funding body/condition of funding
	Nothing would encourage me to share my research data
Other (<i>please comment as fully as possible</i>)	

22. What factors would discourage you from sharing your research data? Please check all that apply:

	The threat of loss of ownership
	Risks to an established research niche
	Risk of premature broadcast of research findings
	Subversion of intellectual property rights, including copyright
	Ethical constraints relating to my research
	Cultural considerations relating to my research
	Consideration of data protection and other confidentiality issues
	The time and effort required to enable sharing
	The lack of information and support to enable sharing
	Risk of diversion from principal objectives through the generation of additional work
	Risk to commercialisation opportunities
	Increased competition for funding
Other (<i>please comment as fully as possible</i>)	

G. SUPPORT AND FURTHER INTEREST

23. If support were available, would you be interested in training or advice on any of the following? Please check all that apply:

<input type="checkbox"/>	Digitisation advice, tools and services
<input type="checkbox"/>	Creating a research data management plan at the beginning of a project
<input type="checkbox"/>	Creating a research data management plan after the project has finished
<input type="checkbox"/>	A data 'exit' plan (for retiring academics or departing academics or postgraduate students)
<input type="checkbox"/>	Data 'rescue' for older digital materials, such as data on older media or migration of data from legacy (now unreadable) systems
Other (<i>please comment as fully as possible</i>)	

24. Would you like to receive ongoing email updates about the current Otago Biodiversity Project and the findings of this research data questionnaire?

<input type="checkbox"/>	Yes
<input type="checkbox"/>	No

Optional, but required if you wish to receive ongoing updates

24a. Your name

24b. Your email address

25. Please feel free to add any other comments regarding data management, long term data storage and access, digitisation, training, etc

Please return this questionnaire by [Date extended] 2008 to:

Gillian Elliot
Biodiversity Project Co-ordinator
Reference Department
Central Library
University of Otago
Dunedin

Further information

Please contact me if you wish to discuss this questionnaire or any other issues, relating to the management and curation of research data.

I am also available to give further presentations, to your colleagues or your department, about this questionnaire or about the wider Otago Biodiversity Project.

Feel free to contact me anytime:

Gillian Elliot
gillian.elliott@otago.ac.nz
ext. 8936

Thank you for your assistance.

Appendix B: List of questionnaire questions

1. Please identify your own role
2. Please indicate your area of research
3. How would formal management of biodiversity research data be useful to you?
4. What kinds of non-digital data do you generate or collect for your research (or have you generated or collected in the past)?
5. What kinds of digital data do you generate or collect for your research?
6. In what formats are these digital information sources held?
7. Are the research data you generate sometimes a combination of different data formats?
8. How large (in total) are your digital research data? Please estimate.
9. How long do you think your research data will have value?
10. What data storage and back up system(s) do you currently have in place?
11. Who currently manages your data?
12. Do you currently have a formal Research Data Management Plan in place and, if not, please indicate the reasons.
13. To which repositories – and how often - do you submit your research data?
14. Can you please indicate what types of metadata you consider important to assign to your data.
15. At what stage are metadata assigned to your research data?
16. Who assigns metadata to your research data?
17. Why might you wish to access the research data generated by other researchers?
18. How would you normally access the research data of other researchers?
19. How do you make your research data available to others?
20. What formal restrictions do you apply to your research data to control access?
21. What factors would encourage you to share your research data?
22. What factors would discourage you from sharing your research data?
23. If support were available, would you be interested in training or advice on any of the following?
24. Would you like to receive ongoing email updates about the current Otago Biodiversity Project and the findings of this research data questionnaire?
25. Please feel free to add any other comments regarding data management, long term data storage and access, digitisation, training, etc

Appendix C: List of respondents' research areas

1. Alpine processes
2. Behavioural ecology
3. Biology
4. Bird conservation
5. Bird ecology
6. Climatology
7. Coastal ecology
8. Communication of sustainability
9. Computational fluid dynamics
10. Conservation biology
11. Conservation management
12. Database systems
13. Ecological interactions
14. Ecological statistics
15. Ecology
16. Environmental chemistry
17. Environmental history
18. Environmental law
19. Environmental physiology
20. Environmental planning
21. Environmental statistics
22. Evolutionary biology
23. Fisheries management
24. Functional biology
25. Freshwater ecology
26. Geographical Information Systems
27. GPS, Earth deformation
28. Heritage tourism
29. Herpetology
30. Higher Education
31. Island management
32. Lichens
33. Marine biogeochemistry
34. Marine biology
35. Marine botany
36. Marine ecology
37. Marine molecular systematics
38. Marine sedimentology
39. Nature-based tourism
40. New Zealand glaciology
41. Organisations and sustainability
42. Parasitology
43. Philosophy of biology
44. Physiology of fish
45. Plant biotechnology
46. Plant conservation
47. Plant ecology
48. Plant evolution
49. Population biology
50. Population ecology
51. Predator control
52. Reproduction/conservation

53. Resource Management Act processes
54. Restoration ecology
55. Science communication
56. Species biology
57. Species management
58. Statistical ecology
59. Sustainable agriculture
60. Textile science
61. Traditional (Maori) management of resources
62. Water quality
63. Wildlife ecology/management

Appendix D: Summary of 'other comments'

"This questionnaire is very well put together & could serve as the catalyst for a very useful service. There is no central repository for spatial data on this campus - at least not one that is well publicized - and that should be."

[Identified University of Otago] *"collections are of national and international significance, and regional, and would be much more valuable if they were databased. [Details supplied of collection and currently available funding]... a drop in the bucket..."*

"We already have a database set up for a specific project, which is small scale and probably not of wide interest. Would consider a bibliography database though."

"It takes a lot of time to manage data for the benefit of others. Difficult to see how individuals will be able to do more of this given pressure to spend time in other ways (but, of course, potential for long-term benefit to society if they do). What incentives would work?"

"This is a wonderful idea - I hope this attempt is successful one."

"My experience may be different to many of the people you hope to deal with... It was never remotely envisaged that [the project] would turn into what it has [details supplied of supervision and published/unpublished outputs]. The irony of this work is that much of it has repeated and extended the work of [earlier researcher discussed]. Sadly all his data sets were destroyed for lack of somewhere to file them and we have only been able to deal with summaries, many of them quite inadequate for conservation purposes. [Illustration supplied of importance of having accessing to both published documents and unpublished data, to understand the full significance of the research.]"

"Although I am researching in [X] & much of this form/questionnaire does not apply, as a teaching tool and a potential for generating further research this project has a lot of merit. For instance... [specific example given]."

"Seems that there is no substitute to paper for archival storage. The need for an archival storage of electronic data is still unfulfilled. We live in an electronic tower of Babel plagued by non standardised hardware and software, rapid obsolescence, poor durability with humans that are as greedy as Microsoft, often paranoid and also those malicious hacker type people."

"Great project... and apologies for delay in feedback - this got buried somehow..."

"I consider myself to be a typical, if somewhat less than average as far as computer skills goes, [role identified]. I am all in favour of data (information) sharing but much of the data that we collect is not classic research data but more opportunistic and is stored in reports etc. Making this data available to a wide audience is good for conservation but i have neither the skills or to be honest the inclination to get it out there!" [Regional, successful example given.]

*"training/more info would be great
I personally like the idea of sharing as much info/data as possible as it saves time and money and gets people to network. But I do lack time and knowledge and would like to be acknowledged for my work. With community based research I have learnt that there is info out there that people do not want too see published or shared and in those cases I would respect that wish."*

Project scope outside current researcher interest:

"Not sure if this will help your survey, as my involvement with things biological is quite minimal."

"I am [interested in student] research following up this questionnaire..."

"Sorry I couldn't answer most of your questions - this doesn't particularly apply to me at the moment (but might in the future). Good luck."

"I think it is a great idea and well worth pursuing. It just doesn't have anything to do with my work - sorry! I wish you well with this valuable project."

"sorry but most of this survey is irrelevant to my type of research - hence I left out heaps! I understand why you only have a few responses from [X]."

Other comments:

"The chocolate frog worked - sorry for the delay!"

"why is this questionnaire not anonymous?" [no name supplied]