Otago Biodiversity Data Management Project Part 2: A Qualitative Report

Gillian Elliot
Biodiversity Project Manager
University of Otago Library
New Zealand

December 2009
Acknowledgements

I wish to thank the University of Otago Library for providing me with the opportunity to share the findings of this qualitative research, as part of the wider ‘Otago Biodiversity Data Management Project’. I would especially like to thank Dr Donald Kerr, for his invaluable tips and also Barbara Taylor, who encouraged me to complete this report.

I would also like to thank Senior Lecturer, Dr Martin Tolich (Sociology) whose advice and encouragement has enabled me to tackle this research project with confidence.

Finally, I would like to thank the many University of Otago researchers, and other biodiversity and data management experts, who were so generous in sharing their time and insights; this report – and its findings – could not have been achieved without them.

Note: In other ‘Otago Biodiversity Data Management Project’ documentation, this report is referred to as the Otago Biodiversity Data Management Project Part 2: Survey Report
Contents

Acknowledgements ................................................................................................................ 2
Executive Summary ................................................................................................................... 5
  Recommendations .............................................................................................................. 5
Introduction ............................................................................................................................ 6
  Otago Biodiversity Data Management Project – activities and reports.............................. 6
  Timeline of Library Project activities and reports .............................................................. 8
Research question ................................................................................................................... 9
Ethics ...................................................................................................................................... 9
Methodology ........................................................................................................................ 10
  Data collection techniques ............................................................................................... 10
  Study population .............................................................................................................. 11
  Interview types and phases .............................................................................................. 11
  Expert consultation and document review .................................................................... 12
  Researcher/Interviewer role ............................................................................................ 13
Timeline of qualitative enquiry/Research process ............................................................... 14
Coding interviews and emerging themes .......................................................................... 15
Analysis ................................................................................................................................ 16
  Analysis of interview themes .......................................................................................... 17
Evaluation ............................................................................................................................ 34
Concluding comments ......................................................................................................... 36
Bibliography .......................................................................................................................... 37
Appendices ............................................................................................................................ 39
  Appendix A: Quantitative and qualitative research ......................................................... 40
  Appendix B: Selected glossary of qualitative research ...................................................... 42
  Appendix C: Permission request form ............................................................................. 51
  Appendix D: Study sample population – further information ........................................... 52
  Appendix E: Emerging themes by interview phase ........................................................... 54
Most research data ends up as what I called semi-structured data. Invariably these are tables in Word. If you’re lucky the scientist is more tech savvy and puts them into Excel. At that point they think they have ‘structured databases’ where in most instances research data requires more complex structures than could ever be expressed naturally in Excel tables. Living under this illusion when they want to analyze data many will spend days integrating the data, massaging it into a form where they can do a particular piece of analysis. Often they will change the data without recording how or what they changed. Then if they need to do it again they frequently have to start all over again. The Popperian notion of falsifiability in science requires that these analyses are reproducible by others, and yet most of the time we do not preserve the data to easily allow this.

Email from an external expert, June 2008

The role of the library in data-intensive research is important and a strategic repositioning of the library with respect to research support is now appropriate.

From the UK’s Joint Information Systems Committee website, retrieved 23rd Nov 2009

[Practitioner-researchers] are often not recognized as researchers. They certainly do not have the traditional image of the researcher, and they may not always be in a position to conduct their research in a most satisfactory way, nor do they necessarily meet the stringent demands of some members of the traditional research community. Nevertheless, this does not mean that they should not be viewed as practitioner-researchers, because that is what they are.

Jarvis, 1999, p.7
Executive Summary

When a life scientist pointed to a pile of unpublished, uncatalogued reports, stacked untidily on the floor of his office, he unwittingly started a series of conversations that were to bring the Library and many researchers at the University of Otago into a much wider global debate about how we define, value, store, use and share ‘research’ in the 21st century.

This initial ‘chance’ conversation rapidly led to other conversations, both within and beyond this department. During the next 15 months, more than 70 University of Otago researchers and external experts – both local and international - were interviewed. All of these conversations were based loosely around the same question: *If the Library could do one thing to better support your research, what would that be?*

This question generated wide ranging discussions around the challenges and benefits of managing (largely) unpublished research information – information ranging in format from traditional, text-based print to often highly unstructured research data (image, text and statistical). Other topics or ‘themes’ to emerge during these interviews included the roles and potentials of the university library, policy makers and funders, particularly around managing research data (both at Otago and more generally). Many researchers also shared comments, suggestions and insights about the research process itself.

The interviews were analysed within the context of a ‘mixed methodology’, combining a qualitative research approach (coding interview themes to extract meaning), with knowledge gained from the literature and also from data management and biodiversity experts. This *Qualitative Report* should also be read in conjunction with the earlier quantitative *Questionnaire Report*, as there is considerable overlap (questions and findings) between these reports. This robust research approach enhances the ability to validate any research findings.

Recommendations

The report’s SIX recommendations address some of the more complex information issues and challenges facing research organisations today. All of these recommendations are based on the findings of conversations with University of Otago and other researchers; all fall outside the scope of traditional librarianship and all could be actioned immediately:

1. Find out what collections currently exist at the University of Otago
2. Find out how researchers currently manage their own websites/bibliographies
3. Develop a ‘support package for emerging researchers’ (in collaboration with others)
4. Capture/manage data now, in conjunction with thesis deposit
5. Lobby for new information professionals/specialists (such as data scientists)
6. Identify collaborative projects, targeting those likely to attract external funding

---

1 As the audience for this report will be information professionals, rather than qualitative researchers, information about the research process, in general, and qualitative research, in particular (including a glossary), has been included.
Introduction

A chance email in April 2007 led to a conversation with a researcher in the Life Sciences Division of the University of Otago. The conversation was broadly on the topic of managing research information. It was an interesting discussion, for both myself and for the researcher, who was keen that I speak with others in the department.

Over the next few months I met with more than a dozen life scientists and other researchers with an interest in biodiversity. Each conversation was based loosely around the question: ‘If the Library could do one thing to better support your research, what would that be?’

All conversations were carefully documented and a picture of some of the issues that were of significance to this group of researchers began to emerge. Reading widely was also important during this time and a number of external experts were also consulted. This ‘framing of the problem’ was significant, enhancing understanding of the issues being shared during these informal interviews, and also enabling others in the Library to appreciate the need for further research into the issues of managing (largely unpublished) research information.

Otago Biodiversity Data Management Project – activities and reports

In January 2008, the University of Otago Library commenced a 12-month feasibility study, the ‘Otago Biodiversity Data Management Project’. This initiative recognised the value of further study into the effective management of unpublished research information and a number of activities, including a questionnaire and ongoing conversations with University of Otago researchers, were supported. Major Project activities and reports are:

Otago Biodiversity Data Management Project Part 4: Final Project Report (Feb 09)
This report includes all Project activities and findings.

Otago Biodiversity Data Management Project Part 1: Questionnaire Report (Oct ’08)
A quantitative report on the current data management practices of University of Otago researchers, with an interest in New Zealand biodiversity.

A fourth anticipated report; Part 3: Pilot Project Report was not completed during the Library Project, but has been taken up by a University Research Project, Long-term preservation and curation of primary research data:
http://eresearch.wiki.otago.ac.nz/Longterm_preservation_and_curation_of_primary_research_data

Current report

Otago Biodiversity Data Management Project Part 2: A Qualitative Report (Dec 09)
In addition to the findings of the quantitative questionnaire, substantial qualitative findings also emerged through a series of face-to-face interviews and phone/email communications

---

2 The Project was extended for a further six months, until the end of June 2009.
3 The findings of this questionnaire, including both quantitative, tabulated responses to multi-choice questions and some qualitative free text comments are available via the Library Projects website: http://library.otago.ac.nz/services/projects.html
4 This was an attempt to understand some of the technical and other challenges presented by managing a real-life research collection (of images), including format, storage, standards, linking and access.
with more than 70 life scientists and data managers. These conversations took place between May 2007, before the commencement of the Library Project, and September 2008, after the findings of the questionnaire had been collated. The findings of these conversations have furthered understanding of the information concerns and interests of Otago researchers, particularly in the areas of unpublished print outputs and research data.

The current report aims to document and report the activities and findings of this qualitative research. Furthermore, as the conversations with researchers informed both the quantitative questionnaire and the Library Project itself, this qualitative data will be analysed within the context of a ‘mixed methods’ research approach (understanding the connections between the qualitative/conversations and quantitative/questionnaire research, the knowledge gained from the literature and also from data management and biodiversity experts).

The conversations with researchers (and any findings), i.e. the qualitative aspect of the Library’s ‘Biodiversity Data Management Project’, are documented here.

**Qualitative research framework**

This report is presented within a qualitative research framework. As the main audience for this document will be information professionals, rather than qualitative researchers, the following information has also been included:

- Quantitative and qualitative research (*Appendix A*)
- Selected glossary of qualitative research (*Appendix B*)
- The place of the ‘research question’ within qualitative research (page 9)
- The function of ‘coding interviews’ in qualitative research (page 15)

---

5 In this report the word ‘conversations’ denotes any substantive communication, including F2F, phone or email.
**Timeline of Library Project activities and reports**

The timeline below places the conversations/interviews, i.e. the qualitative research activity (indicated by the blue buttons), within the wider context of the other events, activities and reports which took place leading up to and during the Library Project. It is clear that the conversations impacted on all aspects of the Project including the questionnaire and even ‘kick-starting’ the Project itself.

![Timeline Diagram](image-url)

**Fig. 1 Timeline of activities and reports**

Note: This diagram is a simplification of the actual processes that took place and is meant as a general guide only. The iterative nature of a ‘mixed methodology’ means that this is seldom a linear process.
**Research question**

A cornerstone of qualitative research is that the research question is not known at the start, but that it ‘emerges’ through the inductive/iterative nature of the research design process.

“The concept of an emergent design is based on the belief that the researcher ‘does not know what he or she doesn’t know’ (Lincoln and Guba, 1985 p.209) at the beginning of a study.” (Pickard, p.17)

While there is no hypothesis, however, there is a need to discover:

“…watching…listening…recording…We begin with a single, open question then explore that question in all its potential forms; the salient issues emerge as we begin to interpret what is going on.” (Pickard, p.158)

So while there was interest and some general ideas at the start of these informal conversations, a clearly defined ‘research question’ had not yet been defined. It was the process of interviewing, of gathering data, of listening to the researchers’ voices, which led to the discovery of patterns and it was these patterns or ‘themes’ which defined the focus of the research.

**Ethics**

This research activity did not begin as a formal research project, but rather as a series of largely unplanned conversations with researchers in the Life Sciences Division of the University of Otago. Furthermore, these conversations were the result of ‘workplace curiosity’ rather than any formal academic research enquiry. As a result, ethical approval, to use information gathered from these conversations, was sought after the research process was established. While this is a somewhat unconventional approach, all of those who had been interviewed were subsequently comfortable with the process and gave approval to use information gained during the earlier conversations (including phone and email communication).

When the ‘permission request form’ (*Appendix C*) was emailed to all interviewees, many requested to view (and subsequently confirmed) the relevant interview notes. This is an important part of the research validation process and is discussed further in *Evaluation*.

**Anonymity v confidentiality**

Unlike the data gathered within quantitative questionnaires, the interview data is confidential, rather than anonymous: “Anonymity means nobody knows who the participant is; confidentiality means nobody will be told the identity of the participants.” (Pickard, p.77)

Any information, gathered during from the 90+ conversations, which identifies the interviewee has been anonymised i.e. the speaker cannot be identified. Some information, particularly that providing detail about the interviewees’ particular research interests has not been included in this qualitative report.

---

6 Member checking (to validate what was said) is important, but not all interviewees felt the need to reflect on comments made/opinions shared, beyond the interview itself.

7 The Library Project questionnaire (informed, at least in part, by the interviews/conversations) gathered anonymous data i.e. the respondents were unknown.
Methodology

Information/data can be gathered in a variety of ways, for different purposes. The focus of this interview research was not to test or substantiate a pre-existing hypothesis, but rather to listen to and to document the interests and concerns of a particular research community (in this case, life scientists), within the broad context of ‘research information management’. Other research methods include: document review; case study; observation\(^8\) and action research and while all these methods are evident within the Library Project and the current research, their influence and significance (with the exception of document review) is not substantial; these methods are evident in this report but not discussed in any detail.

Data collection techniques

Two techniques were used to gather information throughout the 15 month interview period, from May 2007 to September 2008, including:

Conversations/interviews and note-taking

Otherwise known as interviews, these conversations took place in a variety of locations including the researcher’s office and the local coffee shop. No meetings were held at the interviewer’s workspace or elsewhere in the Library, although this was suggested as an option. The most unusual meeting place was in an office 65 feet below ground.

All conversations were documented immediately following the interview.\(^9\) Some notes were also taken during the interview (particularly direct quotes, contact names and other follow-up details) but in the interests of ‘naturalistic enquiry’ note-taking during the interview was kept to a minimum:

“In interviewing, for example, the researcher seeks to establish a rapport and sensitivity with those being interviewed so that they will feel free to expound on topics openly and personally and in the depth that formal interviewing techniques do not foster.” (Gorman, p.67)

Remote communication

Phone and email were routinely used to communicate with experts outside Dunedin. Local interviewees also made contact by phone and email following the interview, to clarify issues, make further suggestions and to generally stay in touch. One obvious benefits of email is that the information is already in print and is easily retained for re-use (with the sender’s permission). A potential drawback of this technology is that emails can be open to re-interpretation (no visual or audio clues associated with the comments). Phone conversations were generally documented as they took place (again with the speaker’s permission).

Other possible data gathering techniques include group discussion (such as focus groups), although this group method was not part of the current research approach.

Memo writing

In addition to keeping notes from the interviews, ‘memo-writing’ was an important aspect of this information gathering process:

---

\(^8\) One humorous incident involved watching a researcher’s slides land on the floor in a heap, supporting the researcher’s claims that these images could be stored more efficiently.

\(^9\) As it was important to record the salient issues from the conversations as accurately as possible, further notes were added immediately following the interview; much of the interview documentation was done in corridors, on street benches and even in departmental toilets!
“Memo-writing is the pivotal intermediate step between data collection and writing drafts of papers’ (Charmaz, 2006, p.72). Theory emerges as a process of constant interaction with the evidence and emerging categories; this means continuous commentary by the researcher on that data in the form of memos (Glaser, 1998)” (Pickard, p.161)

Memos were included with notes from each conversation. This process of constant reflection, interpretation and re-interpretation of the data ensured that the research remained thoroughly ‘grounded’ in the data, rather than focussed on some external reality or hypothesis.

**Study population**

As a result of an initial, informal conversation with a life scientist, further interviews were held with other life scientists and those with an interest in biodiversity. This network approach to a study population is known as interactive or ‘snowball sampling’ and is an appropriate population sampling method within the qualitative theoretical framework. This is markedly different from the quantitative research approach (where anonymous statistical sampling of large study populations is common).

![Fig. 2 Network of conversations](image)

[“Snowball sampling] takes an inductive approach to ‘growing’ the sample as the research progresses… more truly qualitative as it maintains the emergent nature of the research.” (Pickard, p.64)

After the initial interview ‘phase’ (see below), the population sample extended beyond the University of Otago, to include other experts based locally, nationally and internationally. The sample population totalled 78 individuals and the research documentation is based on 90+ interviews (some interviewees provided comments on more than one occasion) and other substantive communication (by phone and email).10

Conversations, informing this qualitative report, ceased when the data received i.e. interview comments, confirming themes and categories, no longer contained any new ideas; this is known as ‘saturation’.

Further information about the study sample population, including sub-groups, interview location and time spent interviewing, is included in *Appendix D*.

**Interview types and phases**

Spradley (1979)11 describes interviews and interviewing as ‘conversations with a purpose’. In addition to recognising the importance of this purposeful approach, it is also necessary to understand the difference between **types** of conversations. This particular research was

---

10 A total of 90+ conversations is given, as the use of data from follow-up emails and phone calls makes it difficult to give a precise number of conversations.
11 In Gorman and Clayton, p.194
carried out over a period of 15 months, commencing well before the start of the Library Project (which began in Jan ‘08, conversations were held from May ‘07) and extending well into the Project period (Project concluded in Dec ‘08). Interviews that were done during this protracted period of investigation changed in both content and structure, as the understanding – of both the interviewer and some of the interviewees\textsuperscript{12} - developed.

In his classic text, *Qualitative Research and Evaluation Methods*, Patton identifies three types of interviews\textsuperscript{13}: unstructured (informal conversations); semi-structured (following a general guide) and structured (standardized, open-ended). The interviews that were part of this qualitative research were unstructured, in the early phases of research, becoming increasingly semi-structured, as data was collated and patterns emerged. In addition to these unstructured and semi-structured approaches, other interview types emerged throughout different phases of this research, including:

1. Informal conversational – unstructured, open-ended enquiry (19 interviews)
2. Expert consultation – seeking validation from experts (12 interviews)
3. Modified conversational – semi-structured (20 interviews)
4. General guide – increasingly informed, associated with the questionnaire (35)
5. Outcome focussed – targeted enquiry, strategic (11 interviews)

The arrow indicates the changing nature of the conversations, moving from a highly inductive approach of discovery through informal conversations, to an increasingly deductive outcome-focussed, strategic approach. While this process moved in a general direction (from discovering patterns to testing hypotheses), the iterative/interpretivist nature of qualitative research means this cannot be described as a linear process. These interview phases are illustrated on Fig. 3 (p.14).

**Expert consultation and document review**

As part of the mixed methods approach, information gained from experts, both in person and in writing (document review) was significant throughout this research. A further distinction (a liberty perhaps?) has also been taken, by including ‘expert’ as both an interview type (expert consultation) and a research method (expert advice). For the purposes of this review, this distinction is meaningful and can be described as:

- **Expert consultation**: a particular interview type or phase representing 12 (wide-ranging) conversations with external experts in the areas of research information management and/or biodiversity. Interviewees were identified as part of the ‘snowball sample’ population.

- **Expert advice**: a research method, often part of a mixed methods approach to understanding and clarifying a research problem. The advice of external experts was sought throughout this research and communication was generally more targeted, specific to a particular question or issue, e.g. *where might I find more information on…?* or *what is meant by…?*.

\textsuperscript{12} A number of the interviewees were interviewed more than once and their interest and understanding of the issues changed.

\textsuperscript{13} These are explained in more detail in the *Glossary* (Appendix B)
**Researcher/Interviewer role**

While it was important to establish a rapport with the interviewees “so they will feel free to expound on topics openly and personally, and in depth” (Gorman, p. 67), it was vital to recognise the role of the researcher/interviewer and the influence that may be brought to the conversations. Issues of subjectivity, partiality and bias are very real concerns and needed to be acknowledged and addressed\(^\text{14}\).

When interviewing this particular study group it is significant to note that both the researcher and the interviewees, particularly in the early phases of the research, knew little, if anything, about ‘research data management’ (which emerged as a key focus of interest and concern). While it could be argued that there was bias in the selection of the sample population (initially life scientists), the evidence for bias and partiality within the research findings is less apparent. It is also significant that there was no research agenda as the research process itself emerged gradually as conversations snowballed and understanding accumulated.

**Being human is a good thing**

The flip side of these negative aspects of being a ‘human instrument’ (subjective, biased, etc) includes the ability to respond to the data. This is a vital aspect of qualitative research, enabling the research focus to adapt to changing information, to continually respond to the data being collated. This iterative process, ‘to-ing and fro-ing’ as more data is gathered, literature consulted, notes reviewed and experts consulted, made the formulation of hypotheses more complex, but it also provided much richer understanding.

**The Aha! moment**

The Aha! moment or ‘tipping point’ occurs when a common theme recurs so frequently, that a particular problem (or problems) can be identified. During the initial phase of interviews, i.e. May - Aug 2007, issues and concerns around managing biodiversity research data recurred so frequently that it led to an Aha! moment.

The result of this Aha! was twofold. Firstly, it informed the direction of the Library feasibility study, the ‘Otago Biodiversity Data Management Project’ (Jan – Dec 2008). Secondly, it guided the focus and direction of subsequent interviews with Otago researchers and other external experts, i.e. the qualitative research activity which informs this report.

It is worth noting that while this Aha! was intuitively recognised by the interviewer it was articulated by one of the interviewees during the first phase. This is significant in the context of valid qualitative research as it adds weight to the impartiality of the interviewer, i.e. the problem emerged – and progressed - during the course of the informal conversational interviews and was identified and articulated by the interviewees themselves.

---

\(^{14}\) Quantitative research is routinely described as objective, partial and unbiased, despite the fact that real people formulate hypotheses and design questionnaires.
## Timeline of qualitative enquiry/Research process

<table>
<thead>
<tr>
<th>Informal conversational (19)</th>
<th>Expert consultation (12)</th>
<th>Modified conversational (20)</th>
<th>General guide (35)</th>
<th>Outcome Focussed (11)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-proposal</td>
<td>Pre-project</td>
<td>Project pre-questionnaire</td>
<td>Project with questionnaire</td>
<td>Project post-questionnaire</td>
</tr>
</tbody>
</table>

### Interview character
- **Interview character**  
  - Unstructured, open-ended enquiry
  - Seeking validation from experts
  - Semi-structured
  - Associated with questionnaire
  - Targeted enquiry, strategic

### Researcher bias
- **Uninformed, tacit knowledge**  
  - Emerging awareness, reflective
  - Increasing understanding, informed
  - Informed, increasingly directive

### Research question
- **Undefined**  
  - Emerging
  - Consolidation, defined
  - Defined (iterative)
  - Defined (ongoing iteration)

### Associated methods
- **Document research, expert advice**  
  - Document research
  - Local consultation
  - Quantitative research, contextualising

### Research approach/ methodology
- **Inductive analysis (discovering patterns)**
  - Deductive (testing hypothesis)

### The Aha! moment
- Information problem recognised (not yet clearly articulated)

### Fig. 3 Interview phases
Coding interviews and emerging themes

Coding interviews

“At the core of qualitative analysis lies a twofold task: to select a bit of data [relevant unit of meaning], and assign it to a category. This has become known as ‘coding’ data... [Qualitative analysis] requires the analyst to create or adapt concepts relevant to the data rather than to apply a set of pre-established rules.” (Dey, p.57-58)

The task of coding interview data is not straightforward and requires a considerable degree of testing and re-testing of categories, recognising patterns of similarity and difference. Certain themes recur and these suggest a logical approach to grouping comments.

All substantial interview comments, quotes and paraphrased statements were documented in Excel. This enabled text to be moved and grouped within appropriate categories. All data have been anonymised and permissions to use this information have been granted.

Time and interview phases

Time was another consideration with this qualitative research (regarding the categorising of responses) i.e. when a conversation took place. The five interview phases (between May ’07 and Sept ’08) must be taken into account when analysing the data.

Interview themes

The main themes which emerged from the 90+ conversations are:

1. Digital management of unpublished research
2. Manage/share research data – positive neutral aspects
3. Manage/share research data – negative aspects
4. Metadata
5. Role of the Library
6. Research Process
7. Funding and collaboration
8. Institutional Research Repositories (IRR)
9. Theses
10. Other

To view each theme by interview phase, together with information relating to each particular phase (population profile, time spent, research approach, quotes, comments and observations and also concurrent Project documentation) see Appendix E.

---

15 NVivo and other software products are available to assist in the categorisation of qualitative data
Analysis

Interview phases
As previously discussed, the interviews were done over an extended 15 month period (May '07 – Sept '08). As a result, there were five distinct interview phases:

1. **Informal conversational** (19 interviews): Exploratory conversations, highly unstructured and open-ended. 75% of the conversations were with Otago researchers.
2. **Expert consultation** (12 interviews): Consultation with data management and biodiversity experts from around the world (international conference and visits). 83% of interviews during this phase were with external experts.
3. **Modified conversational** (20 interviews): The Library Project now started, influencing conversations during this phase and the following two phases, Apr - Jul (with Questionnaire) and Aug - Sept (post-Questionnaire). 83% of interviews with Otago researchers.
4. **General guide** (35 interviews): 23% of interviews with academics and external experts who had already talked about the Library Project. 43% of interviewees also completed the *Otago Biodiversity Project Questionnaire* (shared during these interviews).
5. **Outcome focussed** (11 interviews): Half of these meetings were repeat conversations with interested academics.

Interview theme analysis
For the purposes of analysis, the comments relating to each of the 10 emerging themes have been grouped together (drawn from each of these five interview phases).

To ensure a meaningful analysis of each theme the interview phases have been retained and are indicated by the following coloured backgrounds:

<table>
<thead>
<tr>
<th>Phase 1 (blue)</th>
<th>Informal conversational</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase 2 (red)</td>
<td>Expert consultation</td>
</tr>
<tr>
<td>Phase 3 (green)</td>
<td>Modified conversational</td>
</tr>
<tr>
<td>Phase 4 (orange)</td>
<td>General guide</td>
</tr>
<tr>
<td>Phase 5 (purple)</td>
<td>Outcome focussed</td>
</tr>
</tbody>
</table>

To ensure that the ‘voice of the interviewee’ is retained, selected quotes and paraphrases are included and key comments have been bolded.

Analysis of interview themes

The 10 interview themes are presented on the following pages, together with interview summaries and extracts from each interview phase (indicated by different colours). Comments and observations are also included with each theme (and sub-theme).

| Phase 1 (blue) | Informal conversational |
| Phase 2 (red) | Expert consultation |
| Phase 3 (green) | Modified conversational |
| Phase 4 (orange) | General guide |
| Phase 5 (purple) | Outcome focussed |

1. DIGITAL MANAGEMENT OF UNPUBLISHED RESEARCH

All conversations included some discussion around this theme.

1a. Value/support for action

There was an overwhelming positive response to the possibility that the Library may be able to assist with enhancing access to (largely unpublished) information. One researcher was “amazed that the University is not already supporting researchers and their data’ – it’s hard to believe”. A small number of interviewees indicated that libraries should not step outside their ‘normal business’ or at least “start with traditional library interests i.e. PDFing research...”

The management of unpublished research was of interest to all interviewees, indeed for a number of those interviewed this was their core business. Email comments included “I must say that I am quite excited by your proposed Biodiversity Information programme. It is ambitious in its scope and outcomes, but should it eventuate it will be invaluable to a wide range of scientific and social interests.”

All general comments in response to the value of managing biodiversity research data were positive e.g. “...definitely interested in this project, thank you for being persistent...”, “this project is about more than biodiversity... it is extending the boundaries of research.” Most interviewees put forward specific recommendations for action, rather than simply making general supportive statements.

Three researchers indicated that their primary interest was in better management of key print/eprint resources, not currently managed by the Library. However, the majority of interviewees expressed support for the current Project focus, “The whole subject is of great interest to me, even though I am in [discipline] rather than biology” and “idea of sharing research data would be wonderful”.

‘Digital management of unpublished research continued to generate positive feedback and many of these conversations were followed up with emails including: ‘I'm so heartened to hear of your work and wish you all the best in this endeavour’. Another respondent made the following observation: I'd just like to highlight that not only is the work you're doing of value to the academic community, but may also be very useful to government agencies and other organisations.’

Comments/Observations

Interviewee comments, in all interview phases, indicated that there was a need for better support of unpublished research information and that the Library was well positioned to provide support in this area. Some reservations were expressed (libraries should not step outside their normal business) and the management of eprint resources continues to be of interest and concern for some. Most surprising, perhaps, was the level of interest in, and enthusiasm for, discussing the potential role of libraries (in non-traditional areas) in the increasingly digital environment.
1b. Observation/recommendation/s for action

Many interviewees suggested that I contact one or more of their colleagues (within that department and also outside the University, in CRIs, such as Landcare Research); this led to a ‘snowballing’ of conversations. I was also encouraged to associate any Library activity with the Life Sciences departments, the Ecology Teaching Programme and the ERG\(^\text{17}\). A number of researchers made comments about the difficulties they had in managing their own websites and online bibliographies. One researcher, gave specific recommendations about how to ‘frame’ a proposal “Any proposal should target this area as heritage, rather than biodiversity…. This is our country, our history, our species and if we don’t record in no-one else will. It’s vital to defining who we are…” [intrinsic value]

Most of those interviewed suggested practical ways for progressing the Library Project. Recommendations included: targeting individuals and institutions already active in this area (names supplied); collaboratively managing research collections (by species); linking any Otago project with established global initiatives and architecture (GBIF) and exploring the potentials of online tools, such as ATRIUM, for managing biodiversity information: “an excellent shareware platform for bringing together plot observational data with weather data and GIS information.”

Suggestions included: examine department [X’s] unique collections; promote the Library project at a University lunchtime seminar; establish an interested governance group (to monitor project activities); share project activities with research students and collaborate with other parts of the University (two respondents, details specified). Other comments included the importance of keeping the project regional and the need for an ‘academic champion’; two researchers offered to be the ‘points of contact’ [with others/within the department]. One respondent summed up their interest in an email “So the opportunity remains here for us at Otago to develop some ground-breaking ideas and practices (at least for Australasia) in this area of electronic data management. We begin with a pilot study involving Biodiversity in the Zoology Dept. and possibly [individual’s name] and proceed from there. I am excited about being part of this initiative…”

Many interviewees recommended further contacts, both within and beyond the Life Sciences. Some specific recommendations were also suggested for the Library Project including: the need for an ‘academic champion’; the importance of knowing what collections currently exist at Otago and the value of collaboration (particularly with the Research Office). One interviewee suggested: “I see the bulk of the issues are around document management/institutional repositories… research data repositories… [and] GIS data management…”

Substantial recommendations were made by most (now more informed) interviewees. These included: speak with the Research Office; focus on funding; take a national approach (‘biodiversity isn’t bounded by regions anyway’); pay attention to the ‘cultural dimension’; aim for ‘proof of concept phase’; look for collaborative opportunities (for example Sun Microsoft and the National Library). Further contacts were also suggested (particularly senior researchers) and practical suggestions were shared for dealing with both quantitative and qualitative data (Project reports). I was also advised not to ‘marginalise interest’ by focussing too much on the interests of one department, suggesting researchers in other parts of the organisation may also wish to be involved.

Comments/Observations

There were numerous observations and recommendations, both about managing research information and the Library Project itself (to ensure a successful outcome). Many

\(^{17}\) Ecology, Conservation and Biodiversity Research Group: www.erg.otago.ac.nz/
interviewees suggested further contacts, which led to the ‘snowballing’ of interviews. The need for collaboration (both within and beyond the University) was a strong recurring theme, as was the need for an ‘academic champion’ (to lead a collaborative initiative). Some comments were extremely targeted and specific, such as the importance of framing any activity around heritage (to guarantee interest and investment), while others were contradictory i.e. pilot a study in one department/don’t focus on one department as this will marginalise interest. The comments in this section are extremely rich and highly diverse and many could be usefully explored further, e.g. ‘what collections currently exist at Otago?’ and ‘how do researchers currently manage their own websites/online bibliographies?’

1c. Collection/s suggested
Researchers suggested numerous personal and other research collections which they would be interested in managing and/or accessing more effectively; a number of these are listed in the Library project proposal (Elliot, Aug 2007). These collections included a range of information types and formats, from published print (e.g. bibliographies) to raw data, such as images and spreadsheets. Some suggested ‘collections’ included information from a variety of institutions and individuals, beyond the University of Otago. Most conversations were with external experts who were not familiar with specific Otago collections. Comments related to ‘how to approach’ this challenge, rather than which collection to address, “it is also important to have something to show for the project after 12 months…”

Comments/Observations
Comments from the Pre-Proposal and Pre-Project phases only – this is a result of the timing of the interviews in relation to the Project. Once the Project commenced the ‘research information’ focus (which had initially been very broad) narrowed to ‘research data’ and this was reflected in these later conversations. The collections that were identified early on (Elliot, Aug 2007) included a wide range of information types and formats. There could be value in revisiting these collections as new technologies emerge, library interests change, and new collaborations develop.

2. MANAGE/SHARE DATA – POSITIVE/NEUTRAL ASPECTS

2a. Access/sharing of research data
More than half of the conversations included some discussion about managing research information (other than published or unpublished print information). Several researchers volunteered that they would be willing to share their data, but that they were uncertain of how they might do so. Another observed that it wasn’t until they joined the staff at Otago that they became aware of a database, which is significant in their research field. The issue of why research data might be shared was almost redundant in these conversations. Many in this group were already investing time, energy and resources into sharing digital biodiversity information (including research data). Almost half of the interviewees commented on the value of sharing publicly funded data (for the public good) with one adding that “perhaps researchers working in biodiversity are more
likely to wish to share their data”. Specific observations about data sharing included how a ‘lack of access to climate data (several years ago) had impacted on farmers suffering from extreme weather conditions’ – “now information is shared every three months.” Another interviewee observed that “private companies who gather data on wildlife for their own ...benefits/information... may be prepared to contribute (if a structure was in place)”. While the following statement was made about data sharing in the UK “With respect to the research councils' strategies, the statement [publicly funded research data are a public good] remains true, if not more so. Five of [the] seven Research Councils now have data management/sharing strategies, and the Higher Education Funding Council for England (HEFCE) has recently decided to sponsor a feasibility study into managing research data on a national shared services basis.” Another comment was less positive about the local situation: “...there are numerous [global] initiatives to open up access to publically funded research, yet in New Zealand we seem to be going the other way.” However, at least one positive local example was also shared “It was moved... that as a matter of policy we share biological data from Orokonui with other organisations with the proviso that our staff don’t have to do any reformatting or data input and also with the proviso that Orokonui is acknowledged within any documents that use the data”.

Over half the researchers in this group shared personal experiences about the value of managing and sharing research data. A recurring observation was the “intrinsic value of the data” i.e. “data has value for its own sake.” Several researchers were interested in combining different datasets, “very interested in revealing and layering data... adding [discipline data] layers to topographic maps.” Others remarked on why the pressure to share data in the life sciences may be more compelling than the reasons not to share, such as the threat of mass extinction.

Comments relate to the practicalities of sharing: ‘I sort of know where things are’, and (in an email) ‘While much information exists in theses and publications, the raw data would be of most use to us. Accessing this is not necessarily straightforward.’

Comments/Observations

Many of the interviewees, in all interview phases, commented on the value of sharing and having access to research data including: why it should be shared (public good, advancing knowledge); how it might be shared (work being done internationally) and the problems that can arise through not sharing (economic impact, lack of research information). The overwhelming sense from these conversations was that there was a surprising willingness and openness towards sharing (at least some) data; how to do so, however, was generally unclear or unknown.

2b. Loss/re-use of data/new research potentials

The assumption that ‘earlier research actually informs conservation practice’ was challenged by one researcher, who was not convinced that this is the case. “Practitioners don’t read old sightings of species; rather they are interested in land use, current practices.” One researcher expressed concern about what will happen to their research when they retire. Another made the following observation:

“Capture recent research data now... earlier research was used for a single outcome, then discarded and lost. A new ethos or principle should be to capture this data, digitise, store in a repository... look forward from here, before tackling a huge backlog...”

18 ‘How’ to manage research data more effectively is an important question for many research organisations around the globe and is one of the three research questions which are part of the current University of Otago Research Project, ‘Long term preservation and curation of primary research data’.
One researcher related a very specific example of how unshared data (discovered at a much later date) had skewed the results of a long-term field study. The potential research value of sharing legacy biodiversity print information and also research data, including herbaria, was also discussed.

A number of researchers raised concerns about losing data, and how this impacts (negatively) on new research potentials. The following illustration of the ‘stop/start’ nature of some research projects was shared by one academic “Biodiversity projects start, funding dries up, the student puts the research away and forgets about it - no longer their concern. The research is effectively lost and future researchers need to start from scratch.” Another put forward the following suggestion “Would it be too unreasonable to add to the existing University regulation requiring a student to deposit two hardcopies of a thesis in the library, to also deposit a CD with key data from the thesis?”

**Data loss** was an issue for a number of researchers, with one email respondent providing a detailed example of the problem “I lost to "updating" the entire set of digitised outline maps of [X]!” Several researchers discussed the enduring value of research data, “the value of [long term studies] are tremendous, so whether or not the data is shared, it should be managed... Hocken would have a keen interest in this purely from a New Zealand archival perspective.” Another email concludes, “From what I can gather, my data is seen to be quite valuable. I really was on the verge of throwing all of that stuff out.” Some interviewees also addressed the potential to create new research “…sharing research data would be fabulous for [my] students, who look at published research, but currently are unable to repeat the experiments as they are unable to access the original data. They could re-test or possibly build new areas of research.”

**Comments/Observations**

Again, personal experiences and observations richly illustrated some of the many issues and challenges around the retention and loss of research data. Issues raised included: what happens to data when the researcher retires; changing technologies (and subsequent data loss); the cost of losing data (possible impact on research projects, loss of research potentials); the (arguable) long-term value of data and again, the ‘intrinsic’ value of New Zealand data (it’s our heritage). Practical suggestions included: “Capture recent data now” (worry about the backlog later) and require that data be deposited together with a thesis.

**2c. Current practice: data format/deposit (repositories)**

Several current information sharing practices were discussed, including personal website bibliographies and the use of well-established national and international data repositories. Concern was also raised about the ability of some organisations to maintain digital collections in the longer term, with one researcher pointing to the “temporary nature of some Australian government departments”. Several researchers noted the substantial activities of Landcare Research, who build a large proportion of New Zealand’s biodiversity databases (see Part 1: Questionnaire Report).

There were just three comments here; one researcher explained their department’s local solution for managing research outputs (EndNote Library), while two others described their experiences of accessing data from international sources.

Two researchers indicated that it was easier to get data from (trusted) international online sources, while others shared concerns with current, local, practices: “Lots of files on variable discs in offices... unreadable, need assistance now... data being lost”; “issues of who updates, maintains database, also issues of access (must be on campus to use this)” and “Central management of research data... would eliminate the proliferation of 'bad’ practices.”
Comments/Observations
Comments by a few of the interviewees, from three interview phases, emphasised the value of established ‘trusted’ international repositories for depositing/accessing research data. Apart from the activities of Landcare Research, no other (funded, sustainable) New Zealand depositing options were shared; rather concerns were raised about local ‘inadequacies’.

3. MANAGE/SHARE DATA – NEGATIVE ASPECTS

<table>
<thead>
<tr>
<th>Phase 1 (blue)</th>
<th>Phase 2 (red)</th>
<th>Phase 3 (green)</th>
<th>Phase 4 (orange)</th>
<th>Phase 5 (purple)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Informal conversational</td>
<td>Expert consultation</td>
<td>Modified conversational</td>
<td>General guide</td>
<td>Outcome focussed</td>
</tr>
</tbody>
</table>

3a. Ownership/IP/copyright/credit for data (PBRF19)
The negative aspects of managing and sharing research data were discussed in less than half of the conversations. Several researchers noted unsuccessful sharing practices i.e. they had submitted data to a repository but were experiencing difficulty in subsequently retrieving this data (loss of control). Another was a “supporter of PBRF” and suggested “this may impact negatively on collaboration.”

Three comments here, all relating to the absolute requirement that researchers be acknowledged for their data, “needs [to be] a policy change to recognise impact factors on use of this information... University rewards system should recognise use of data/datasets...” One interviewee went further, suggesting that we “publish – don’t share data”.

Several interviewees addressed the legal aspects of data sharing, “ownership, this is certainly a concern and something to be examined in detail...researchers are understandably sensitive about who uses their [data], where and why; acknowledgement is critical...” A couple of responses were couched quite strongly “[some departments] don’t share...they want credit for their data” and “[some researchers] only cite articles written by other academics [from their own particular institution] (to keep PBRF’s up).” Cultural considerations and security issues were also of concern to two researchers, “cultural sensitivity...again an issue with native/endemic species” and “[X] stressed the anxiety surrounding the ongoing safety of endangered species...”

Just under half of the interviewees expressed concerns about the negative implications of managing and sharing research data. About a third of this group raised concerns about ‘ownership’: “I have worked with [X]patient data which could never be shared”; “a major issue with this is copyright, as this resides with [Y] and no-one is quite sure who can okay ‘not-for-profit’ digitisation,” and “when staff leave, they may take a copy of their work, but what happens to the original, which is owned by the University?” One researcher was particularly concerned about research data being managed by a third party “There is no function in my vision for centralized repositories other than as an archive for stuff that nobody wants to actively curate.”

Comments/Observations
Less than half the interviewees raised concerns around data ownership. Concerns were: legal (ownership, copyright); cultural (responsibilities with native species data) and institutional (impact factors, PBRF). One interviewee suggested ‘publishing’ rather than sharing data. Concern was also raised about the reliability and value of centralised repositories; a topic

19 Performance Based Review Fund – (funded) measure of excellence in New Zealand tertiary research
which has also been raised with regard to IRRs (see *Library Trends* special issue on ‘Institutional Repositories: Current State and Future’, 2008)

3b. Cost/complexity/workload/lack of time

Several comments here, including: “*lack of funds can be a disincentive to manage e-stuff*” and “*be wary of digitisation... it’s a bottomless pit...*” One researcher was critical of the government’s approach to KAREN20, suggesting that “*high speed capacity has been developed without adequate infrastructure... the potential is there but it won’t cope... the government has created a race-track, but no cars to run on it.*”

Just a single comment about how *workload and lack of time* makes it difficult to effectively share data, “[X] felt this time of year was too busy, for either himself or his postgrads...”

Where ‘cost’ was felt to be an issue, concerns focussed mainly on technical issues, i.e. *technical obsolescence and lack of infrastructure* “*KAREN is a great big pipe which ends in a bit of string... infrastructural backbone doesn’t extend to within the institution itself... capabilities are partial only and problems remain for researchers, storing, managing and sharing their own data here at Otago.*” One researcher raised a further concern, relating to *career stage* “*The issues for established researchers are quite different. Time is a huge issue and they already have established patterns of research. They also have collections and will be more interested in deposit of (already formatted) data than needing to access the data of others.*”

Three separate concerns were raised by three interviewees i.e: lack of *time* (‘too busy at this time of year’); *ethical* considerations (especially within the area of health) and *cultural* responsibility (particularly with regard to Maori and endemic species).

Comments/Observations

There were fewer comments around the different costs of managing research data. Those who did comment mentioned: financial cost (lack of funds, “*bottom-less pit*”); workload (too busy already) and two comments, one each about ethics (particularly in the area of health) and culture (with regard to Maori and endemic species). Two interviewees mentioned the inadequacies of KAREN, “*race-track but no cars to run on it*” and “*great big pipe which ends in a bit of string*”. These comments point to the need for greater communication and collaboration between those working in the area of research and research infrastructure.

3c. Quality/reliability of data

Just one interviewee expressed concern about the quality and reliability of data, “*Combining different sets of data within the one Excel spreadsheet presents huge problems of interpretation and certainly brings into question the possibility of ‘re-use’ – who would want to, or even could reuse this data?*”

Data quality and reliability concerned a number of researchers and some compelling personal experiences were described: “*Quality control of data is an important issue and something that the University is (understandably) acutely aware of...its reputation is at stake... I have lots of ‘unpublished data’ in my office, which I wouldn’t necessarily wish to reveal or share with other researchers.*” “*The research data, associated with the published research is also included but only that research which has been through some validation/quality control process or process of peer review. If this isn’t done, the data is regarded as not worth sharing and is not revealed.*” “*Digital manipulation is of concern i.e. maintaining data integrity (difference between ‘cleaning up’ the image and actually changing the imaging to something different; also the importance of appropriate imaging*
when capturing data.’ One researcher also talked about the data gathered by their own students, “The data they collate is good, but not verified to the highest standard and therefore must be discarded.” Others felt that some researchers collate data ‘badly’ (at least some of the time), making it difficult to re-use. The following quote from one researcher is revealing “Most research data ends up as... semi-structured data. Invariably these are tables in Word. If you’re lucky the [researcher] is more tech savvy and puts them into Excel. At that point they think they have ‘structured databases’ where in most instances research data requires more complex structures than could ever be expressed naturally in Excel tables. Living under this illusion when they want to analyze data many will spend days integrating that data, massaging it into a form where they can do a particular piece of analysis. Often they will change the data without recording how or what they changed. Then if they need to do it again they frequently have to start all over again.”

Comments/observations
Interestingly, there were a number of comments on the quality/reliability of data, and almost all of these were raised during the fourth interview phase (with the questionnaire). The reason for this is unclear, but it may be that the questions in the written questionnaire inadvertently stimulated discussion around this topic. Concerns about data quality included: the importance of quality control (University’s reputation); inadequacies of some data (gathered by students for a single purpose); maintaining data integrity/digital manipulation (distinguishing between cleaning up and changing an image) and poor data collection techniques (use/misuse of flat files, such as Excel, rather than relational databases, overuse of proprietary software e.g. Word). An email from one interviewee explains the problem of ‘semi-structured data’ in some detail, outlining a series of inappropriate/inadequate processes in data handling which are commonplace. This is one aspect of research data management that could be addressed locally/institutionally by providing appropriate training to emerging researchers.

4. METADATA

<table>
<thead>
<tr>
<th>Phase 1 (blue)</th>
<th>Informal conversational</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase 2 (red)</td>
<td>Expert consultation</td>
</tr>
<tr>
<td>Phase 3 (green)</td>
<td>Modified conversational</td>
</tr>
<tr>
<td>Phase 4 (orange)</td>
<td>General guide</td>
</tr>
<tr>
<td>Phase 5 (purple)</td>
<td>Outcome focussed</td>
</tr>
</tbody>
</table>

Just one comment here, relating to MoRST and work on policy development “basically around metadata.”

Half of this group talked about the importance of metadata. Comments and suggestions focussed on appropriate metadata standards and also around individuals and groups already working in this area, including: Biodiversity Information Standards (formerly known as TDWG), National Biological Information Structure, INOTAXA Project (web workspace for taxonomic descriptions, catalogues, names specimen data, images and other resources) and the value of Darwin Core and ABCD (metadata standards for the life sciences).

Less than half of those interviewed discussed metadata. Comments included general statements about the value of standardisation and interoperability, the complexities of metadata, “it is the metadata that is the challenge” and more discipline-specific comments about ‘significant descriptive metadata in the life sciences’ i.e. name/taxonomy, location, time and date. One emailed comment was particularly illuminating “…but without a reasonable amount of knowledge of how [the data] was collected, stored, etc. it wouldn’t be usable by others -- and currently that meta-data is not associated with the dataset.”
Less than a third of this group provided comments about metadata. Several researchers addressed the need for ‘context’, “Data in isolation are largely meaningless.” Others mentioned the importance of adequate standards and the importance of naming/taxonomy in the life sciences, “…without sound taxonomy, there is no point in sharing research – becomes meaningless.” One researcher spoke of the increasing interest in (and potentials of) ontologies and the semantic web, “In the case of organism names that would be the ‘synonyms and classifications’ so for example the user could ask for all data-sets that contain information on the daisy family with all the names standardized to the currently accepted name.”

One researcher discussed the importance of standards for data migration, while another stressed the importance of key descriptive metadata within the life sciences, i.e: name/taxonomy; time/date and location/place (noting the value of the New Zealand Map Grid). A third commentator emailed “In addition, information regarding the methodology used is also important, especially if data sets from a variety of sources is to be combined.”

Comments/observations
Understanding about metadata varied substantially between different groups, with the ‘external expert’ group commenting less about the value of metadata (this was a given) and more about discipline standards, groups and projects (Darwin Core, TDWG21). Other comments which emerged from the conversations included: the importance of metadata standards and interoperability (and data migration); the need for discipline focussed descriptive metadata (taxonomy, time and location). One interviewee discussed ontologies and the potentials of the semantic web, while another pointed to MoRST22 and policy development around metadata. These and other issues and concerns around metadata need to be better understood by information professionals; this will ensure meaningful dialogue, not just between those working with information, but also between information professionals and those who generate the information in the first place.

5. ROLE OF THE LIBRARY

<table>
<thead>
<tr>
<th>Phase</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase 1 (blue)</td>
<td>Informal conversational</td>
</tr>
<tr>
<td>Phase 2 (red)</td>
<td>Expert consultation</td>
</tr>
<tr>
<td>Phase 3 (green)</td>
<td>Modified conversational</td>
</tr>
<tr>
<td>Phase 4 (orange)</td>
<td>General guide</td>
</tr>
<tr>
<td>Phase 5 (purple)</td>
<td>Outcome focussed</td>
</tr>
</tbody>
</table>

Just two comments on the changing role of libraries, but notably from opposing perspectives i.e. “be careful to direct thinking within traditional library boundaries” and “libraries need to wake up... blurring of publication outputs and databases... what will the library be doing... not housing print – there won’t be any!” Several researchers talked about how libraries are ‘trusted’ - “I would be keen to house this [data] in the Library and make more widely available” and an email “It would be great if the library could appropriately catalogue, care for and make available the wealth of information collected by [X]”

Several differing perspectives on the role of the library in the new digital research environment: the library is ‘trusted’ and should have a role in managing research data in future; data should be managed ‘at source’ i.e. the library should “work with the research scientist, curating their research data at source”. One interviewee ‘discussed the issues with library catalogues and how they are of little use to taxonomists (information structured very

21 Biodiversity Information Standards Working Group: www.tdwg.org/
22 Ministry of Science Research and Technology: www.morst.govt.nz/
differently)... need to extend granularity of information” and the following advice was also shared: “A huge culture change is coming to academia as a result of digitisation.”

Several researchers indicated that the Library should have a role in managing research data, “it is vital that any project is Library-based” citing the Library’s ‘neutrality, communication and ability to collaborate with all departments’. One interviewee suggested that the Library should “focus on sharing biodiversity ePrint resources”.

Around half the researchers shared information about how they perceived the role of the library and almost all of these comments related to the ‘changing responsibilities’ of libraries. Comments targeted three key aspects: What role? A surprising number of researchers had ‘forgotten’ about the Library, “I never go to the Library... It would be nice to visit.” and “I don’t normally think of libraries having much of a role.” Libraries have a role: Others expected libraries to embrace new challenges in the rapidly changing digital information environment. One researcher queried why the New Zealand Digital Content Strategy lacked a national direction or focus, with regard to the management of research data, while another researcher articulated why librarians should be interested in managing data ‘Librarians manage stuff and it shouldn’t matter what format. The process of change in institutions can be slow, but this shouldn’t stop libraries from moving forward, managing their own business, in the best way they know. They have a duty of care for stuff. It’s important to recognise the distinction between storage and access. Libraries should be dynamic, keep building, heading standards and then deal with the politics of access when it is appropriate. Just because the institution doesn’t support OA now, doesn’t mean it won’t do so in future. We can’t wait for these decisions or we’d never get anything done! Librarians undervalue their own skills – they are too modest. Yes, collaborate with ITS, departments, but librarians have huge competencies and they shouldn’t be afraid to use them.’ Librarians have no role: Several researchers were yet to be convinced that libraries and librarians had a role in managing research data, as articulated by one researcher “Most scientists see [libraries] as providing base-level support functions and are justifiably suspicious that they can do much more than that. Managing research data... goes beyond such base-level support. It requires cross-disciplinary domain expertise within that support function.”

Just two comments here, one restating the importance of ‘trust’ when managing data, the other questioning who should be leading activities in the area of data management (National Library?)

Comments/observations
There were a number of comments, particularly from those interviewed during the fourth phase (with the questionnaire), about the changing role of libraries (and librarians). Opinion seems to be divided as to what this future role might be, with some interviewees seeing no role for the library beyond “traditional library boundaries”, while others stressed the need for the library to “wake up” and embrace the challenges of the new research environment. Specific comments and suggestions ranged between these two positions, i.e. libraries have no role (and are unable to adapt) and libraries have a role (‘Librarians manage stuff and it shouldn’t matter what format’). A number of interviewees also stressed that libraries and librarians are widely trusted and that their neutral status could enable them to develop new potentials within the research community (such as working with researchers to manage their data at source). Other comments included: the need to distinguish between ‘storage and access’ (to progress useful dialogue in this area); the inadequacies of library catalogues (for

---

23 One researcher expressed a strong interest in discussing Library research support, rather than anything else.
taxonomists) and the need for strong national leadership in this area (questioning the role of the National Library).

6. RESEARCH PROCESS

| Phase 1 (blue) | Informal conversational |
| Phase 2 (red) | Expert consultation |
| Phase 3 (green) | Modified conversational |
| Phase 4 (orange) | General guide |
| Phase 5 (purple) | Outcome focussed |

Just three comments here, including observations about the place of new digital initiatives at Otago, which “would sit well” and more general statements about ‘rapid digitisation’ of literature and the need for changes to PBRF policy (to recognise data/dataset citation).

Just three comments here, about the value of Open Access (from international experience), the statement about the UK Research Council strategies (see 2a. Access/sharing of research data) and a general comment about “the role of data scientists and the need to reconsider how data is collated and managed, before researchers are in the field.”

A quarter of interviewees provided comments relating to the research process and three themes emerged: the value of managing research data at the earliest phase of the research life-cycle, “this is suggesting a need for a ‘seismic’ shift in thinking/attitudes and practice (think Kuhn)”; the publishing process “...to get something published there has to be some new methodology, justifying a publication. This may mean lots of data are never published (or retained)” and the need to collate data in re-usable format/train emerging researchers “The research advisors are critical to the research process and highly relevant to researchers. How research is done (collation, management of data) is fundamental to this Project and it is the emerging researchers who should be supported into this process.” “The Popperian notion of falsifiability in science requires that these analyses are reproducible by others, and yet most of the time we do not preserve the data to easily allow this. Science is not generally ‘auditable’ in this sense. The new landscape of... plug & play analysis and modeling tools, workflow tools etc, all require that this structured, auditable data access infrastructure is in place. In many areas we are probably waiting for a generation of old school scientists to retire before things will change. Those who do change and keep up will leap ahead quite dramatically.”

The ‘conservativism’ of most research (particularly in New Zealand) was raised as a concern by one interviewee, i.e. the requirement to specialise to a high degree, encouraging excellence in very narrow fields, ‘does not encourage cross-fertilisation and openness to change’. Further comments related to: time pressures (short time frame for many research projects); grant funding (the need to allocate a % for data management); use of appropriate file formats (for data use and reuse) and the value of a greenfield approach.

Comments/observations

These comments, from all but the external expert group, included a range of general observations: rapid digitization of literature; value of OA and concerns about the publishing process; conservativism of New Zealand research (narrow focus); time and (grant) funding pressures; the place of digital initiatives at Otago and UK Research council activities. One

---

24 Greenfield: A brand new installation of equipment without the requirement of integrating existing systems. Contrast with "brownfield," which is an upgrade to an existing system. Coined by the building industry to refer to clean, undeveloped land (greenfield) vs. contaminated land or land with existing structures (brownfield), the term may be used to refer to network installations. From PC Magazine Encyclopedia, www.pcmag.com/encyclopedia_term/0,2542,t=greenfield&id=43956,00.asp retrieved 27 July 2009
interviewee also discussed the value of a ‘greenfield approach’ to changing organisational practices. Specific recommendations included: the need for changes to PBRF (to recognise changing research outputs); the value of managing research data at the earliest possible phase (collaborative potentials); the need to train emerging researchers (data for reuse) and also the importance of recognising and training new types of research/information specialists (such as data scientists). The range of topics covered during these conversations was extensive and there would be considerable value in revisiting at least some of these. Most notable for libraries is the need to train new and different types of information professionals - a challenge, perhaps, for both the National Library and the educators of our future information professionals.

7. FUNDING AND COLLABORATION

| Phase 1 (blue) | Informal conversational |
| Phase 2 (red) | Expert consultation |
| Phase 3 (green) | Modified conversational |
| Phase 4 (orange) | General guide |
| Phase 5 (purple) | Outcome focused |

‘Funding and collaboration’ have been brought under a single theme as they were often discussed simultaneously. While they have been separated into two sub-themes, these groupings are not absolute, and some overlap still occurs.

7a. Funding

Funding and/or collaboration were mentioned in more than half of these initial conversations and some substantial opinions were shared. Several researchers shared personal experiences of seeking external funding for university-based projects (can be difficult). Canterbury University’s recent success in attracting funding to digitize entomology records was noted as a useful model and other researchers commented on the potential for funding to PDF existing biodiversity print documents. One researcher mentioned Australia’s commitment to ‘infrastructural IT’, putting “$8 million into biodiversity alone!” while another pointed to U.S. funded initiatives, such as the ‘Biodiversity Heritage Library’ and the ‘Encyclopaedia of Life’.

Two thirds of all interviewees shared examples and possibilities of collaboration and/or funding. Some of these conversations contained substantial local detail (not reproduced here). Several interviewees spoke of specific funding agencies and potentials for project funding. TFBIS25 funding was known to two researchers, funding success with FRST26 was shared by another researcher and another pointed to the possibility of TEC27 funding, “TEC – collaborative - initiative is looking to improve access to [other specified information] resources across universities, CRIs, government departments and local councils – would TEC funding also be available for this biodiversity initiative?”

Collaboration and/or funding were discussed by two thirds of those interviewed. Several researchers were disparaging about New Zealand funding, comparing this with funding support overseas “[This New Zealand Government initiative] has a chequered history,

---

26 Foundation for Research, Science and Technology: www.frst.govt.nz/
27 Tertiary Education Commission (TEC) ... [is] responsible for funding the government’s contribution to tertiary education and training offered by universities, institutes of technology and polytechnics...” Retrieved 20th June 2009 from: www.tec.govt.nz/templates/standard.aspx?id=448
initiated in 2006, then erratic funding and support thereafter... was to share data, but this hasn’t yet eventuated and last meeting was in December 2007.” “Cornell’s eBird... are going to assist with NZ material” and (slightly tongue-in-cheek), “If New Zealand cares so little about its own identity (unique biodiversity, science publications) then we would be better off as another state of Australia – better funding!” Two New Zealand funding sources were also suggested, i.e. TFBIS and MoRST contestable funding.

There was just one substantial comment which addressed the need for a national approach to funding, i.e. ‘regional approach will not attract sufficient/sustainable funding (and biodiversity is not bounded by regions anyway)’. Recommended an approach to national funders and interested parties (Landcare, DOC, MoRST, MfE, National Library) ‘sooner, rather than later – ‘not an easy task, but vital’.

Comments/observations
Funding was discussed by many of the interviewees, in all of the interview phases (just two suggestions from one interviewee in the post-questionnaire phase). Comments were both general (the need for/difficulties of securing external funding and the fact that better funding is available overseas, citing examples) and specific (Canterbury University’s success in digitizing entomology records is a useful model and “Cornell’s eBird... are going to assist with New Zealand material”). Many of those interviewed recommended national funding sources, including TFBIS, FRST, MoRST and TEC, while one interviewee stressed the importance of a national approach ‘sooner, rather than later.’ It is clear from these conversations that funding is a major issue for all academic institutions. It is also a fact, however, that external/national funding is available and that activities around collaboration and sharing of resources are increasingly attracting interest and gaining support.

7b. Collaboration
Funding and/or collaboration were mentioned in more than half of these initial conversations and some substantial opinions were shared. Most (but not all) comments focussed on how collaboration could be improved within and across institutions. Comments ranged from the local, “closer collaboration with the University Library and information specialists could be beneficial” to the global, “...better to hook up with GBIF28, than any local initiatives.” Between these two positions, comments touched on the need for universities to work more closely together, “New Zealand is a small country and collaboration is paramount to achieve goals”. Observations were shared about current collaborations (e.g. New Zealand regional councils’ Envirolink programme) and also future University collaboration potentials, working with government departments, such DOC and also the Ministry for the Environment, “this is an effective organisation and their capabilities are under-rated.” Other comments stressed the need for a multidisciplinary approach, when considering ‘biodiversity’. Conversely, the observation that New Zealand universities are “not typically involved with [biodiversity projects] at a national level... smaller initiatives, often at the species level” was repeated during a later conversation, “Landcare Research [rather than the University] is doing this now”.

Comments included the importance of collaboration and communication when seeking funding, “[X]stressed the need for ‘collaboration’ with any digitisation project and pointed out that universities were too competitive in their approach” and “communication is key”. The challenges of “long-term sustainability, funding sources, project based start-up funds vs ongoing maintenance costs” were also raised. Comments also tended to be both externally and internally-focussed, “look at the Canadian universities, who recently committed six

28 Global Biodiversity Information Facility: www.gbif.org/
Two thirds of all interviewees shared examples and possibilities of collaboration and/or funding. Some of these conversations contained substantial local detail (not reproduced here). Many researchers are already routinely collaborating with others, within their normal research activities. These collaborations are with a wide range of individuals and groups, including: other departments; other New Zealand Universities; Private Training Establishments (PTEs); Maori organisations (locally and nationally); regional councils; CRIs\(^\text{29}\): NGOs and government departments. Other observations about collaboration were: directive, i.e. the Library project ‘must be a collaborative effort’ and the project must retain a ‘regional focus’; supportive, for example, “[I] would be willing to collaborate on an Otago research data project” and critical. One researcher shared detailed examples to demonstrate how little collaboration currently exists within New Zealand, with regard to biodiversity initiatives, suggesting that things were “less than ideal” or that researchers were “talking past each other.”

A number of examples of current collaborations (with other University departments and New Zealand organisations) were shared. One researcher also discussed the “slight relaxation in the ‘separateness’ of CRIs and universities... better communication, some collaboration and sharing... NIWA\(^\text{30}\) shared their West Coast coastal photos... (wouldn’t have happened in the past).” Several local and external researchers also expressed interest in collaborating with the Library (on similar projects) “[I am] ringing to see if there is any way the Library could work with [X] on developing a web-based data system...”

Most interviewees supplied comments about collaboration. Examples of collaboration included: sharing an international ‘multidisciplinary approach to scientific investigation’ (citing an example of a more holistic approach); working more closely with the Research Office; working with industry (citing the successful example of Sun Microsoft and National Library collaboration) and the importance of genuine consultation with Maori “Ticking the boxes approach is not necessarily the only, or even the best, way to consult.” Two researchers were interested in a symposium to explore the implications of data sharing.

Comments/observations
There was considerable discussion around the need for ‘better collaboration’ within all interview phases. The contexts for these collaborations (both potential and realised) included: local, i.e. within our own institution (between departments, ITS, Library, Research Office); regional (targeted to interest groups); national (other universities, DOC, MfE\(^\text{31}\),) and global (better funding potentials, sustainable), “…better to hook up with GBIF, than any local initiatives.” One interviewee suggested working with industry, such as Microsoft, while another stressed the importance of genuine consultation/collaboration with Maori. Opinions were sometimes contradictory, recommending both a large scale multidisciplinary approach and also a local ‘regional approach’. Other observations included the changing nature of CRIs, such as Landcare and NIWA (becoming less ‘separate’) and also the illusion of collaboration, “talking past each other”, (although many also argued that they were routinely collaborating with colleagues within their normal research activities). It is clear, from the many comments, that collaboration is not only desirable, but essential if individuals and organisations are to benefit fully from the vast intellectual and financial resources that are

\(^{29}\) Crown Research Institutes: [http://sciencenewzealand.org/](http://sciencenewzealand.org/)

\(^{30}\) National Institute of Water and Atmospheric Research: [www.niwa.co.nz/](http://www.niwa.co.nz/)

increasingly becoming available within the highly connected research and information environments.

8. INSTITUTIONAL RESEARCH REPOSITORIES (IRR)

<table>
<thead>
<tr>
<th>Phase</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Informal conversational</td>
</tr>
<tr>
<td>2</td>
<td>Expert consultation</td>
</tr>
<tr>
<td>3</td>
<td>Modified conversational</td>
</tr>
<tr>
<td>4</td>
<td>General guide</td>
</tr>
<tr>
<td>5</td>
<td>Outcome focused</td>
</tr>
</tbody>
</table>

Just a few comments here, including: one researcher claimed to send all research outputs to the Science Library already; another would like to know the research outputs of other universities and requested that student outputs be made available (and therefore known). A follow-up email contained the following remark, “IRRs are basically archival, rather than research focussed.”

There was just one, very practically-focussed, conversation about institutional repositories touching on current practice and potentials for the future.

Only two researchers had any knowledge of Institutional Research Repositories and both were uncertain about Otago’s activities in this area.

Around a third of this group knew of, or had an interest in, knowing more about, Institutional Repositories. The majority of the comments were positive, “…very interested in the Library’s work in this area as [I have] many documents I would like to see included in this repository.” Several concerns and questions were also raised, around: standards, “who applies the standards – KRIS32, National Library or Institutional contributors [at] Otago?”

impact on researcher practice, “…generally cautious about ‘process of deposit’ – workload and copyright issues are the main concerns” and PBRF, “… keen to know how (or if) the Library will accommodate unconventional, peer-reviewed staff outputs, eligible for PBRF.”

There was also a further comment about the role of the National Library “I remain deeply unconvinced by the current national institutional repositories approach. I see the National Library ‘in charge’ of implementing the Digital Strategy and doing so with no concept of ‘repositories’ as anything other than ‘library documents + metadata’.”

Two comments here, one relating to the ‘National survey of the perceptions of academic staff on institutional repositories’33 (uncertainty about both the letter and IRRs). A second interviewee believed all their data was available through institutional and other repositories, but was uncertain which ones.

Comments/observations
Conversation about IRRs was limited as few of the interviewees (from all interview groups) either knew about IRRs or wished to make any comment. Those who had some experience or understanding of IRRs were generally neutral or positive, but a number of concerns were raised, including: the value of IRRs (archival rather than research focussed); the importance of standards (and who should apply these); impact on researcher practice (workload, copyright); the need to expose student research and accommodation of unconventional research formats (for PBRF); future potentials for IRRs and the role of the National Library (perceived to be inadequate). Some confusion was apparent around: accessing research outputs from other universities; who is currently depositing and where and what the School of Information Management’s ‘National survey of the perceptions of academic staff on

32 Kiwi Research Information Service (Gateway to NZ open-access research): http://nzresearch.org.nz/
33 The survey was run by Victoria University’s School of Information Management
institutional repositories’ is all about. While there were a number of useful issues raised during these conversations the overwhelming impression was of researchers struggling to understand the changes that are taking place, with regard to how information is managed, and what potentials exist in this area. The ‘what’s in it for me?’ does not seem to have been shared with our own research community; a major challenge, perhaps, for all those working with researchers and information.

9. THESES

<table>
<thead>
<tr>
<th>Phase 1 (blue)</th>
<th>Informal conversational</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase 2 (red)</td>
<td>Expert consultation</td>
</tr>
<tr>
<td>Phase 3 (green)</td>
<td>Modified conversational</td>
</tr>
<tr>
<td>Phase 4 (orange)</td>
<td>General guide</td>
</tr>
<tr>
<td>Phase 5 (purple)</td>
<td>Outcome focussed</td>
</tr>
</tbody>
</table>

Two comments here, one researcher expressed interest in making all theses (all levels) digitally available. Another was interested in knowing Otago’s policy, with regard to “full text scanning of theses.”

Several interviewees shared comments or had questions relating to how Otago theses are, or will be, made available in future. Comments included “[I am] unclear of the University’s position with regard to full text scanning of theses.” “One thought I have had... is how to get the information collected in a hard copy thesis, into a machine readable/data base format?” and “...in the near future all PhD theses (including Otago theses) should be visible”

Just two comments about theses, both around which theses are to be included in Otago’s IRR (PhD, Masters) and when this will happen.

Comments/observations
A few questions, rather than comments, were raised during three of the five interview phases. The questions were all very much on the same few topics i.e. which Otago theses will be made digitally available, how will this be achieved and when will this happen? One interviewee specifically asked when there will be an Otago policy change to accommodate the digitisation of theses. None of the comments indicated any reservations about making Otago theses digitally (and widely) available. It is difficult to know why there were so few comments about theses; whether there is little interest in this topic (theses in other organisations have been digitally available for some time), whether the conversation did not encourage comments in this area or whether researchers are unaware of Otago’s activities with regard to the digitisation of theses (you can’t ask questions about what you don’t know). There would be value in further conversations with our research community about theses, at the very least to raise awareness of our own IRR and the potentials this has for exposing Otago’s research outputs.

10. OTHER

<table>
<thead>
<tr>
<th>Phase 1 (blue)</th>
<th>Informal conversational</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase 2 (red)</td>
<td>Expert consultation</td>
</tr>
<tr>
<td>Phase 3 (green)</td>
<td>Modified conversational</td>
</tr>
<tr>
<td>Phase 4 (orange)</td>
<td>General guide</td>
</tr>
<tr>
<td>Phase 5 (purple)</td>
<td>Outcome focussed</td>
</tr>
</tbody>
</table>

A number of more specific comments, mostly: questions, relating to personal research interests; recommendations about databases e.g. ‘Catalogue of Life’ and ‘Species 2000’ and tools (such as CONNOTEA and PITION). Further useful contacts, were also made during these early conversations.
Further comments were made about: **specific software, tools, initiatives and practices** (including DOIs, Australian and U.S. activities); **engaging with New Zealand research** and researchers and also ‘**staying in touch**’ with future developments and initiatives.

More than half of the researchers in this ‘Pre-Questionnaire’ group are from disciplines outside the life sciences; they were recommended to me because they are working in the areas of biodiversity and/or digital management. Other comments and concerns raised by this group were **wide ranging**, including: the **global lack of taxonomists**; the **reduction in New Zealand science print publishing**; the **importance of biodiversity data modelling**; defining ‘what is biodiversity’ and the need for a digital “**biodiversity landscape diagram [showing] complexity of relationships, key players, databases, infrastructure.”**

Other comments covered a **wide range** of issues, including: the **need for enhanced visibility** (who is doing what in New Zealand with regard to repositories, data management); **University responsibility re sustainability and biodiversity** (and what Australia is doing) and the **merging of New Zealand science publications** into a single online equivalent (one researcher stressed the immeasurable value of these print publications, in terms of content, uniqueness, identity, research merit and nomenclature/taxonomic requirement).

One researcher pointed to an article which confirmed the **University of Otago as a significant producer of high quality research in the area of biological conservation.** “**Otago is the only New Zealand university listed with 13 other southern universities as being ranked equivalently with US and Canadian universities in the top 40: Otago is 8/14 and equivalent to the top 20 North American rankings.”** The other comment related to **trust** (and a story was shared). Trust can be difficult to achieve but once gained it is invaluable – particularly when trying to make things happen; to make a difference.

**Comments/observations**

A number of interviewees from each interview group raised issues that fell outside the recurring themes that were identified and which have defined this qualitative report. These ‘other’ issues all relate to information, however, and it is appropriate that they be included as a tenth theme. These other issues are wide ranging and include (in no particular order): the **reduction in New Zealand science print publishing (merging of publications to a single online equivalent)**; **anxiety over the global shortage of taxonomists (naming underpins everything);** the **importance of defining what is meant by ‘biodiversity’** (commonality of understanding); the **value of biodiversity modelling** (invaluable for planning/policy making); **University of Otago responsibility with regard to sustainability and biodiversity.** Several interviewees also stressed the need for enhanced visibility regarding who is doing what with a) New Zealand repositories and data, and b) biodiversity data (need for a schematic to reveal the complexity of infrastructure, players, tools and projects).

Some of the individual conversations included comments relating to the emergent themes (pointing to tools, initiatives, practices and further contacts) but also extended beyond these themes, focussing more on issues relating to the researcher’s own area of speciality or personal interest. Much of this content is too specific to share but the value of this personal connection cannot be overemphasised; this foundation of shared interest and trust is absolutely fundamental if genuine collaboration is to become a reality.

---

34 One researcher stressed the immeasurable value of these print publications, in terms of content, uniqueness, identity, research merit and nomenclature/taxonomic requirement.

35 This information would readily identify the individual researcher; this falls outside the intent of this report.
Evaluation

Some aspects of this study, e.g. the ethics approval process, could mean that this research does not qualify as ‘pure research’ i.e. it may not “meet the stringent demands of some members of the traditional research community” (Jarvis, p.7). It is important, however, to recognise that this qualitative research was carried out in a robust manner and that the findings are meaningful. A number of criteria can be applied as evidence of the validity of the research process and subsequent findings.

a. Researcher role

The role of the researcher in qualitative research is critical and, as already discussed in Researcher role, it is important to openly recognise any partiality or bias during the interview process and subsequent interpretation and analysis. As stated earlier:

“When interviewing this particular study group it is significant to note that both the researcher and the interviewees, particularly in the early phases of the research, knew little, if anything, about ‘research data management’ (which emerged as a key focus of interest and concern). While it could be argued that there was bias in the selection of the sample population (initially life scientists), the evidence for bias and partiality within the research findings is less apparent. It is also significant that there was no research agenda as the research process itself emerged gradually as conversations snowballed and understanding accumulated.” (p.13)

Furthermore, the explicitly subjective approach of this qualitative research, the valuing of the interviewer’s own intuition and tacit knowledge, enabled the research problem to emerge from the data and be recognised, not only by the interviewer, but by the interviewees themselves.

b. Research data and findings

Mixed methodology

Validity “indicates that your methods warrant conclusions… validity indicates that the conclusions you have drawn are trustworthy….” (O’Leary, p.61)

The research documented in this report is largely qualitative, framed within a mixed methods approach (see Methodology). This methodology adds substantially to the validity of the research findings presented within this report. External expert knowledge (from the literature and from a wide range of individuals)\(^{36}\) has confirmed and contextualised many of the issues raised during the interviews and enhanced understanding of these issues. In addition, many of the topics included within the Project questionnaire\(^{37}\) were also raised during the interviews and similar findings emerged from both research activities. Cross-over of questionnaire/interview themes included: the Library’s Data Management Project; metadata; and, most significantly, the advantages, disadvantages and challenges of sharing research data. Other themes which emerged during the interviews included: the role of the library; institutional repositories; theses; the research process; and collaboration and funding.

\(^{36}\) Already shared, through the outputs (activities and documentation) of the ‘Otago Biodiversity Data Management Project’

Significantly, a number of interviewees also participated in the questionnaire, adding a complimentary, quantifiable dimension to the understanding gained through the interviews, i.e. adding ‘breadth to depth’. This research population overlap is shown in Fig. 4.

![Fig. 4 Overlap of survey sample populations (interview/questionnaire)](image)

**Member checking**

Interviewees were encouraged to review any interview notes, both during the interview itself (specific comments were reflected back to the interviewee) and also when the permission request form was sent out. This member checking helped to validate what was being shared and documented, and also to highlight any issues that were of particular importance to the interviewee.

c. **Research process**

To ensure a valid research process it is important to provide evidence of thoroughness in all activities by maintaining adequate documentation. This clear ‘audit trail’ makes it possible for other researchers to review the process of research and ensure thoroughness and accuracy throughout. This research thoroughly documented all conversations/interviews, memos, expert advice and literature consulted (including notes) in Word and EndNote. All qualitative analysis was documented and managed within Excel.

**Transferability of research findings**

Generalizability or external validity is generally associated with quantitative research, whereby the “findings of a sample are directly applicable to a larger population.” (O’Leary, p.62). In other words, the findings of the few\(^\text{38}\) can be generally representative of the many.

The very nature of qualitative research (with its small population sample) seldom enables this wider ‘representativeness’. However, what can be achieved is a richness of understanding, of issues and concerns, and a ‘transferability’ of this understanding within a larger population. Put another way, the ‘rich pictures’ shared by the few can shed light on understanding for the many.

---

\(^\text{38}\) While quantitative population samples are generally ‘large’ (possibly in the thousands), the sample population is always still much smaller than the actual population.
Concluding comments

The interview themes that have emerged from this qualitative research reflect some of the ‘research information management’ and wider interests and concerns of more than 70 University of Otago researchers and other external experts with an interest in biodiversity. Furthermore these themes highlight a range of issues falling both within and beyond the traditional scope of information professionals working within the academic research environment. These issues – and many challenges - are of significance and many are compelling and worthy of further investigation, not just for their own sake, but also because they have been recognised as having value by the researchers themselves.

In addition to the six Recommendations proposed in the Executive Summary of this report, Key Findings that can be drawn from this qualitative research include:

Discovering the value of…
- genuine outreach and the possibilities that can emerge through active listening
- narrative in revealing the richness of personal experience, leading to understanding
- a mixed methods approach in recognising and understanding new and complex issues

Resonating themes...
- ‘research data is our heritage’ – if we choose not to care, no-one will
- ‘libraries need to wake up’ – the world is changing rapidly and so too must libraries
- ‘what’s in it for me (the researcher)’ – we need to know what researchers value
- ‘greenfield approach to thinking’ – step outside existing frameworks to enable new and different possibilities

Recurring themes...
- Researchers are ‘time-poor’ and this impacts significantly on decision-making
- Collaboration is the key (to resources, funding) – it is the way of the future
- Libraries are trusted and neutral (this can be said of few institutions today)
- The Library could have a role beyond traditional areas of information management
- Different academic disciplines have different requirements (must not be overlooked)
- The value of institutional repositories is unknown (libraries should care about this)
- Digital theses are a good idea, but knowledge/understanding in this area is limited
- PBRF matters and many have concerns about its current scope, approved formats
- KAREN is a national good, but the infrastructure is not yet there at the local level
- Much confusion about metadata – what it is, what it does and why it matters
- There is a perceived lack of a clear national direction/leadership in ‘digital NZ’

The recommendations and findings of this qualitative research are significant for University of Otago researchers (particularly those with an interest in biodiversity) and for information professionals at the University of Otago Library. These research findings are also likely to be of interest, and arguably transferable to, other researchers and tertiary information specialists, within and beyond the University of Otago.
Bibliography


Appendices

Appendix A  Quantitative and qualitative research (2 pages)

Appendix B  Selected glossary of qualitative research (9 pages)

Appendix C  Permission request form (1 page)

Appendix D  Study sample population – further information pie charts (2 pages)

Appendix E  Emerging themes by interview phase, together with information relating to each particular phase: population profile; time spent; research approach; quotes; comments and observations. Concurrent Project documentation is also included (23 pages)
Appendix A: Quantitative and qualitative research

To be genuinely robust, any research must be undertaken within a theoretical framework. When this framework includes a variety of research methods, this is known as a ‘mixed methods’ approach.

**Mixed methods**
This research approach may include one or more of the following methods: quantitative (largely statistical); qualitative (interpretative); document research and expert consultation. The advantage of using more than one of these methodologies is the ability to ‘triangulate’ or cross-check data and findings. Some researchers prefer the idea of ‘crystallisation’, but the point is that data obtained from a variety of sources strengthens the validity of any findings (see *Evaluation*).

How and why a researcher may wish to consult the literature or an expert on any given topic is fairly self-explanatory. However, it is worth providing a brief explanation about what is implied by quantitative and qualitative research.

![Fig. 5 Mixed methodology](image)

**Quantitative and qualitative research**
Quantitative research often begins with a literature review, from which a hypothesis is derived and variables identified. The hypothesis is then tested, using statistical processes and procedures. It typically involves a large population sample and is highly measurable.

In qualitative research a hypothesis emerges from the data, which is collected and interpreted; human understanding is central to this methodology. Study populations are generally small and research is difficult to analyse and measure. The research process is highly iterative.

“**Quantitative** measures are succinct, parsimonious, and easily aggregated for analysis; quantitative data are systematic, standardized, and easily presented in a short space. By contrast, the **qualitative** findings are longer, more detailed, and variable in content; analysis is difficult because responses are neither systematic nor standardized. Yet, the open-ended responses permit one to understand the world as seen by the respondents. The purpose of gathering responses to open-ended questions is to enable the researcher to understand and capture the points of view of other people without predetermining those points of view through prior selection of questionnaire categories...“To capture participants’ in their own terms’ one must learn *their* categories for rendering explicable and coherent the flux of raw reality. That, indeed, is the first principle of qualitative analysis.” (Lofland, 1971, p.7 in Patton, pp.20-2)
Characteristics of quantitative and qualitative research

Certain characteristics are associated with a particular research framework. Fig. 6 summarises the key features of qualitative and quantitative methodologies (which framed the two key activities of this Project ie. the conversations/interviews and the questionnaire).

The terminology is explained in more detail in the Glossary of this report (Appendix B).

<table>
<thead>
<tr>
<th>WHAT</th>
<th>Qualitative method</th>
<th>Quantitative method</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Paradigm</td>
<td>Interpretivist</td>
<td>Positivist</td>
</tr>
<tr>
<td>2. Methodology</td>
<td>Inductive; Intuitive; Iterative; Naturalistic Emergent design; Empathetic</td>
<td>Deductive; Theory-based; Objective</td>
</tr>
<tr>
<td></td>
<td>Mixed methodology – triangulation</td>
<td></td>
</tr>
<tr>
<td>3. Methods</td>
<td>Survey; elements of Action Research and Case Study; Document review;</td>
<td>Survey</td>
</tr>
<tr>
<td>4. Techniques</td>
<td>Interviews (informal, structured, semi-structured); Document research; Expert consultation</td>
<td>Questionnaire</td>
</tr>
<tr>
<td>5. Instrument</td>
<td>The Researcher Reliability/bias</td>
<td>Questionnaire Neutrality</td>
</tr>
<tr>
<td>6. Analysis</td>
<td>Thematic exploration; Coding</td>
<td>Statistical analysis</td>
</tr>
<tr>
<td>7. Evaluation</td>
<td>Transferability Data credibility Member checking</td>
<td>Generalizability Data validity</td>
</tr>
</tbody>
</table>

Fig. 6 Characteristics of qualitative and quantitative research (associated with this Project)

Library Project

While the quantitative method was not part of this qualitative research (and is therefore not the focus of this document), the following should be restated:

- the conversations (qualitative) informed the questionnaire (quantitative)
- the questionnaire pushed awareness of the issues out to a wider audience
- the questionnaire also enabled the issues (from the conversations) to be quantified
- the conversations gave much fuller understanding to individual researcher problems, providing context and meaning around many of the questionnaire responses

The quantitative and qualitative research activities carried out during this Project are inherently linked and this connection impacted on both activities and outcomes.

---

The quantitative questionnaire resulted from the qualitative interviews and is included here as an element of this mixed methods approach. As the findings of this questionnaire have already been reported (Part I: Questionnaire Report), this quantitative component of the research is not discussed in detail here.
Appendix B: Selected glossary of qualitative research

**Action research method**
This method is characterized by the researcher examining current processes, taking action to improve those processes, then analysing the results of the action. It is most common in educational research but can be found in many professions where improvement of practice is being sought. The action research model takes an approach that is described as either cyclical or spiral, and includes the following stages: identifying problems; action planning; implementation; evaluation and reflection. (Pickard, 2007, p.134)

A research strategy that pursues action and knowledge in an integrated fashion through a cyclical and participatory process. In action research, process, outcome, and application are inextricably linked. (O’Leary, 2004, p.139)

The goal of action research is to work with stakeholders to generate knowledge in order to action change. (O’Leary, 2004, p.98)

**Aha! moment** (aka Eureka!)
The Eureka Effect is the sudden realization of the solution to a problem. Such inspiration usually follows considerable perspiration. Also called the aha phenomenon. ([http://siliclone.tripod.com/books/history/H111.html](http://siliclone.tripod.com/books/history/H111.html) retrieved 14 May 2009)

**Anonymity** (see also Confidentiality)
Anonymity means nobody knows who the participant is; confidentiality means nobody will be told the identity of the participants. (Pickard, 2007, p.77)

**Case study method**
A case study is an empirical inquiry that investigates a contemporary phenomenon in depth and within its real-life context, especially when the boundaries between phenomenon and context are not clearly evident. In other words, you would use the case study method because you wanted to understand a real life phenomenon in depth, but such understanding encompassed important contextual conditions – because they were highly pertinent to your phenomenon of study. (Yin, 2009, p.18)

A method of studying elements of the social through comprehensive description and analysis of a single situation or case, for example, a detailed study of an individual or group, episode, event, or any other unit of social life organization. Emphasis is often placed on understanding the unity and wholeness of the particular case. (O’Leary, 2004, p.115)

**Confidentiality** (see also Anonymity)
Anonymity means nobody knows who the participant is; confidentiality means nobody will be told the identity of the participants. (Pickard, 2007, p.77)

**Constant comparative method**
This strategy involves taking one piece of data and comparing it with all others that may be similar or different in order to develop conceptualizations of the possible relations between various pieces of data…the creation of categories is driven by the raw data and not established a priori, although it is inevitable that prior research will have identified some of the salient issues: ‘The original version of grounded theory stressed the idea that theory emerged from, and was grounded in, data. Careful analysis of data items using the constant comparative method would lead to the emergence of conceptual categories that would describe and explain the phenomenon under study.’ (Melia, 1997, p.31 in Pickard, 2007, p.241)
Credibility (see also Reliability; Validity)
The quality, capability, or power to elicit belief. (O’Leary, 1999, p.56)

Crystal/Crystallisation (see also Triangulation)
[some writers] dispute the usefulness of the concept of triangulation, asserting that the central image for qualitative inquiry should be the crystal, not the triangle. Mixed-genre texts in the postexperimental moment have more than three sides. Like crystals...the mixed-genre text “combines symmetry and substances, transmutations... Crystals grow, change, alter... Crystals are prisms that reflect externalities and refract within themselves, creating different colors, patterns, arrays, casting off in different directions.” (Richardson, 2000, p.934 in Denzin & Lincoln, 2005, pp. 5-6)

Data analysis
... the process of qualitative data analysis involves the use of inductive (discovering) and/or deductive (uncovering) reasoning to generate and interpret relevant themes in order to achieve meaningful understanding. (O’Leary, 2004, p.195)

The analysis of qualitative data is usually seen as arduous. The reason why it is found to be difficult is that it is not fundamentally a mechanical or technical exercise. It is a dynamic, intuitive and creative process of inductive reasoning, thinking and theorizing. (Basit, 2003, p.143)

Data – bit
The first question we encounter involves deciding what constitutes a ‘bit’ of data... We may want to categorize words, lines, sentences or paragraphs... Since ideas can be expressed succinctly or expansively, the number of words is less important than the meaning they convey. The underlying consideration should be the relevant ‘unit of meaning’ which is conveyed by content rather than form. (Dey, 1993, p.115)

Data - coding
At the core of qualitative analysis lies a twofold task: to select a bit of data, and assign it to a category. This has become known as ‘coding’ data... [Qualitative analysis] requires the analyst to create or adapt concepts relevant to the data rather than to apply a set of pre-established rules. (Dey, 1993, p.57-58)

In qualitative research, coding is the process whereby data are broken down into component parts, which are given names. (Bryman, 2008, pp.692-3)

Data – qualitative
Observations that yield detailed, thick description; inquiry in depth; interviews that capture direct quotations about people’s personal perspectives and experiences; case studies; careful document review [my emphasis]. (Patton, 2002, p.40)

Pure description and quotations are the raw data of qualitative inquiry. The description is meant to take the reader into the setting. The data do not include judgements about whether what occurred was good or bad, appropriate or inappropriate, or any other interpretative judgements. The data simply describe what occurred. (Patton, 2002, pp.26-27)

Data - recontextualization
The first thing we may do after creating and assigning categories to the data is consider ways of refining or focusing our analysis. To do this, we can shift attention from the ‘original’ data itself to the data as reconceptualised through the results of our labours... we can now organize and analyse our data in terms of the categories which we have developed. This shift in focus has been described as a ‘recontextualization’ of the data (Tesch 1990), as it can now be
viewed in the context of our own categories rather than in its original context. (Dey, 1993, p.129)

**Data reduction**
During data reduction the researcher-scientist condenses volumes of data into quantifiable analytic units; data are manipulated and reconfigured in an attempt to discover patterns and connections not previously apparent. (Gorman & Clayton, 2005, p.205)

**Deduction (see also Induction)**
A deductive approach to research is one where a hypothesis is derived from existing theory and the empirical world is explored, and data are collected, in order to test the truth or falsity of the hypothesis. (O’Reilly, 2005, p.26)

*Inductive analysis* involves discovering patterns, themes, and categories in one’s data. Findings emerge out of the data, through the analyst’s interactions with the data, in contrast to *deductive analysis* where the data are analyzed according to an existing framework. (Patton, 2002, p.453)

**Document analysis**
Document analysis refers to the collection, review, interrogation, and analysis of various forms of text as a primary source of research data [not a literature review]. Documents suitable for analysis might be authoritative, agenda-based, personal, multimedia-based, or historical.

Document analysis sees researchers working with pre-produced, rather than generated, texts. This requires researchers to consider two potential sources of bias; both the original author’s and their own. (O’Leary, 2004, p.183)

**Emergent design (see also Grounded theory)**
The concept of an emergent design is based on the belief that the researcher ‘does not know what he or she doesn’t know’ (Lincoln and Guba, 1985, 209) at the beginning of a study. Therefore it would be impossible to establish the means by which the unknown could manifest itself to the researcher during the course of the study. Because of this, qualitative research allows the design to emerge as the study progresses. A research model can and should be developed that allows for the iterative nature of the study. (Pickard, 2007, p.17)

...adapting inquiry as understanding deepens. (Patton, 2002, p.39)

**Empirical/empiricism**
An approach to the study of reality that suggests that only knowledge gained through experience and the senses is acceptable. (Bryman, 2008, p.693)

**Ethics in research**
Generally we think of research ethics in terms of our impact on subjects, respondents, or participants. (O’Leary, 2004, p.151)

Research ethics is a field which is constantly changing and its boundaries are at times quite fuzzy. What might be acceptable research behaviour one year, may be unacceptable the next. (Bow, 1999, 254) (Pickard, p.71)

**Ethnographic research method**
[Ethnographic research] focuses on interpretations of behaviour or specific events in the everyday lives of individuals. The primary data collection technique is participant observation... Ethnography uses descriptive narrative to analyse and present the findings from the investigation; the purpose is to build theory by describing situations as they are and presenting patterns within a context. (Pickard, 2007, p.111)
**Generalizability** (see also Transferability; Validity)

The degree to which it is justifiable to apply to a wider population explanations and descriptions that research has found apply in a particular sample or example. (Gibbs, 2007, p.149)

A concern with the external validity of research findings. (Bryman, 2008, p.694)

Generalizability (also referred to as external validity) indicates that the findings of a sample are directly applicable to a larger population. This is most likely to be possible for studies whose large sample sizes often limit data to that of a quantitative nature. (O’Leary, 1999, p.62)

**Grounded theory** (see also Emergent design)

Grounded theory researchers collect data and analyze it simultaneously from the initial phases of research. Researchers cannot know exactly what the most significant social processes are in particular settings, so they start with areas of interest to them and form preliminary interviewing questions to open up those areas. They explore and examine research participants’ concerns and then further develop questions around those concerns, subsequently seeking participants whose experiences speak to those questions. Hence grounded theory methods keep researchers close to their gathered data rather than to what they may have previously assumed or wished was the case. (Charmaz, 2002, p.675)

Emergent methodological design is often central to grounded theory methodology. In grounded theory, researchers work inductively to generate theories strictly from the data. (O’Leary, 1999, p.96)

**Human instrument**

When human experience and situations are the subject of research, then the human as instrument is ‘the only instrument which is flexible enough to capture the complexity, subtlety, and constantly changing situation which is the human experience’ (Maykut and Morehouse, 1994, 26). This researcher as instrument is also in a position to apply appropriate tacit knowledge to each situation and event as it occurs. (Pickard, 2007, pp.16-17)

**Hypothesis**

The research process begins with a topic of interest to the research and, more often than not, requires that problem to be solved. While defining this problem, the researcher generates a series of questions and also formulates hypotheses, or “educated guesses,” as to what the research findings will be. Hypotheses both provide a logical means to locate the data and serve as a way of resolving the research problem. (Beck & Manuel, 2008, p.15)

Logical conjecture (hunch or educated guess) about the nature of relationships between two or more variables expressed in the form of a testable statement. (O’Leary, 2004, p.36)

**Induction** (see also Deduction)

A simplistically inductive approach to research is one where the researcher beings with as open a mind and as few preconceptions as possible, allowing theory to emerge from the data. (O’Reilly, 2005, p.26)

*Inductive analysis* involves discovering patterns, themes, and categories in one’s data. Findings emerge out of the data, through the analyst’s interactions with the data, in contrast to *deductive analysis* where the data are analyzed according to an existing framework. (Patton, 2002, p.453)
Interpretivism (see also Research paradigm)
Interpretivism acknowledges and explores the cultural and historical interpretations of the social world. (O’Leary, 2005, p.10)

Interview/Interviewing
...‘a conversation with a purpose’ (Spradley, 1979, p.58 in Gorman & Clayton, 2005, p.194)

Interviewing may be done in a variety of ways, from highly formal and structured (appointments, set questions) to casual and non-directive (in the staff room, anecdotal), and each approach has its advantages. (Gorman & Clayton, 2005, p.41)

Interview – semi-structured (general guide)
An interview guide lists the questions or issues that are to be explored in the course of the interview. An interview guide is prepared to ensure that the same basic lines of inquiry are pursued with each person interviewed. The interview guide provides topics or subject areas within which the interviewer is free to explore, probe, and ask questions that will elucidate and illuminate that particular subject. (Patton, 2002, p.343)

Interview – structured (standardized open-ended)
[The] standardized open-ended interview consists of a set of questions carefully worded and arranged with the intention of taking each respondent through the same sequence and asking each respondent the same questions with essentially the same words. (Patton, 2002, p.344)

Interview – unstructured (informal conversation)
The informal conversational interview is the most open-ended approach to interviewing. It is also called “unstructured interviewing” (Fontana and Frey 2000:652). [It] offers maximum flexibility to pursue information in whatever direction appears to be appropriate...The informal conversational interview relies entirely on the spontaneous generation of questions in the natural flow of an interaction... (Patton, 2002, p.342)

Iterative process
The main feature of [the iterative theory building process] is that research questions may be changed over time based on material collected and that research strategies, data collection and analysis methods and tactics should fit the (changing) research questions and process phases. (Kerssens-van Drongelen, 2001, p.503)

Iterative implies both a spiral and a straight line, a loop and a tail... (O’Reilly, 2007, p.27)

Member checking (negotiated outcomes)
One additional – and important – way in which to establish validity [of research findings] is by the use of what Guba has termed ‘member checks’: the circulation of a draft report to members of the organization or group for review at the end of the case. (Gorman & Clayton, 2005, p.60)

...negotiated outcomes or ‘member checking’ is a vital component of a study, not just in terms of adding to the credibility of the study, but also in improving the quality of the final case report. [Stake] stresses that ‘all [his] reports have been improved by member checking’. (Stake, 1995, p.116 in Pickard, 2007, p.17)

Memo writing
Memo-writing is the pivotal intermediate step between data collection and writing drafts of papers’ (Charmaz, 2006, 72). Theory emerges as a process of constant interaction with the evidence and emerging categories; this means continuous commentary by the researcher on that data in the form of memos (Glaser, 1998)... Essentially it is a written commentary on the data, an interpretation of data as it is gathered. Because you are writing about real data this
process helps to keep you grounded in the evidence and not force data into extant concepts that may be far removed from the actual evidence in front of you. Memo-writing is a very useful way of identifying gaps in your data, showing up areas which appear to have been overlooked or need further investigation...Memos link the whole process of data collection, analysis and research reporting. (Pickard, 2007, p.161)

**Mixed methodology**
A term that is increasingly employed to describe research that combines the use of both quantitative and qualitative research. (Bryman, 2008, p.695)

**Naturalistic enquiry**
Qualitative designs are naturalistic to the extent that the research takes place in real-world settings and the researcher does not attempt to manipulate the phenomenon of interest...[which] unfolds naturally in that it has no predetermined course established by and for the researcher...people are interviewed with open-ended questions in places and under conditions that are comfortable for and familiar to them. (Patton, 2002, p.39)

**Objectivity** (see also Subjectivity)
Objectivity has been considered the strength of the scientific method. The primary methods for achieving objectivity in science have been conducting blind experiments and quantification. “Objective tests” gather data through instruments that, in principle, are not dependent on human skill, perception, or even presence. (Patton, 2002, p.50)

There are no objective observations, only observations socially situated in the worlds of – and between – the observer and the observed. (Denzin & Lincoln, 2005, p.21)

Objectivity is measured by the extent to which the findings from an investigation would remain constant regardless of the character of the researcher. Findings are a result of the research investigation, not a result of the researcher’s interpretation of those findings. The goal is to demonstrate that the investigation is value-free, free from any personal constructs of the researcher. (Pickard, 2007, p.22)

**Positivism** (see also Research paradigm)
Put simply, positivism assumes that social phenomenon can be approached with scientific method and makes a number of assumptions about the world and the nature of research. (O’Leary, 2004, p.5)

**Qualitative analysis**
Qualitative analysis is applied in any study that focuses on emerging theory, using the inductive analysis process to arrive at an understanding of the phenomenon under investigation. When applying qualitative analysis the purpose is to generate a hypothesis based on the data gathered and interpretation of that data. (Pickard, 2007, p.239)

**Qualitative research**
Qualitative research is a situated activity that locates the observer in the world. It consists of a set of interpretative, material practices that make the world visible. These practices transform the world. They turn the world into a series of representations, including field notes, interviews, conversations, photographs, recordings, and memos to the self. At this level, qualitative research involves an interpretative, naturalistic approach to the world. This means that qualitative researchers study things in their natural settings, attempting to make sense of, or interpret, phenomena in terms of the meanings people bring to them. (Denzin and Lincoln, 2005, p.3)

Qualitative research is a process of enquiry that draws data from the context in which events occur, in an attempt to describe these occurrences, as a means of determining the process in
which events are embedded and the perspectives of those participating in the events, using induction to derive possible explanations based on observed phenomena. (Gorman & Clayton, 2005, p.3)

**Quantitative research**
Quantitative research begins with a theoretical framework established from the literature review; from this framework a hypothesis will emerge and the variables within that hypothesis can be identified. The notion of hypothesis can also be translated into research aims and objectives; it is only compulsory to have a hypothesis when true experimental research is chosen as the method. (Pickard, 2007, p.18)

**Reliability**
Reliability is premised on the notion that there is some sense of uniformity or standardization in what is being measured, and that methods need to consistently capture what is being explored. Reliability is thus the extent to which a measure, procedure, or instrument provides the same result on repeated trials [e.g. bathroom scales]. (O’Leary, 1999, p.59)

**Research method**
This is the bounded system created by the researcher to engage in empirical investigation, the overall approach, often referred to as a ‘strategy’. (Pickard, 2007, p.xvi)

Qualitative research methods include: case study; survey; ethnography and action research.

**Research methodology**
This is the theoretical perspective of the research, that is the overall nature of the research activity... I believe there are only two fundamental methodologies: qualitative or quantitative. (Pickard, 2007, p.xvi)

**Research techniques**
These are the individual data collection techniques [interview; questionnaire and focus groups] applied within the method. (Pickard, p.xvii)

**Research paradigm**
In its established usage, a paradigm is an accepted model or pattern (Kuhn, 1970, p.23)

A research paradigm...is often an individual’s view of the world that dictates the nature of the research they engage with... Positivist thinking is associated with quantitative research, interpretivist thinking with qualitative research and postpositivist thinking with a dualism that attempts to include both methodologies. (Pickard, 2007, p.xvi)

**Sample population** (aka Research population)
Your research population is the entire set of individuals about which inference will be made. (Pickard, 2007, p.60)

**Sampling – a priori**
[A priori sampling] establishes a sample framework before sampling begins... (Pickard, p.64)

**Sampling – purposive** (aka non-random)
The logic and power of purposeful sampling lies in selecting information-rich cases for study in depth. Information-rich cases are those from which one can learn a great deal about issues of central importance to the purpose of the inquiry, thus the term purposeful sampling. Studying information-rich cases yields insights and in-depth understanding rather than empirical generalizations. (Patton, 2002, p.230)
Sampling – snowball (aka interactive)
Interviewees may well recommend other people you should talk to, which leads to a snowballing of your sample. (O’Reilly, 2005, p.40)

[Snowball sampling] takes an inductive approach to ‘growing’ the sample as the research progresses… more truly qualitative as it maintains the emergent nature of the research. (Pickard, 2007, p.64)

With this approach to sampling, the researcher makes initial contact with a small group of people who are relevant to the research topic and then uses these to establish contacts with others... There is a much better ‘fit’ between snowball sampling and the theoretical strategy of qualitative research than with the statistical sampling approach of quantitative research. (Bryman, 2008, pp.184-185)

Saturation (aka Theoretical saturation)
In grounded theory, the situation where predictions and expectations based on existing data and categories are repeatedly confirmed by data from additional categories or cases. The additional categories or cases seem to contain no new ideas and they are then said to be saturated. The search for further appropriate instances seems futile and data collection ceases. Also referred to as data saturation. (Gibbs, 2007, p.151)

Subjectivity (see also Objectivity)
Subjectivism emphasises the subjective elements in experience and accepts that personal experiences are the foundation for factual knowledge. (O’Leary, 2005, p.10)

Survey research
The purpose of survey research is to gather and analyse information by questioning individuals who are either representative of the research population or are the entire research population… Survey research can include qualitative and quantitative research… (Pickard, p.95)

Tacit knowledge
Tacit knowledge…is learned from experience, either pre-consciously – that is, without having entered the conscious mind – or consciously, and has been forgotten or even repressed... Picardi (1988, p.91) defines tacit knowledge… as the “type of knowledge which, even though it can somehow be manifested, need not, and in many cases cannot be articulated linguistically”… Not all knowledge is narrative or discourse. (Jarvis, 1999, p.48)

Theme
A recurring issue or an idea or concept either derived from prior theory or from respondents’ lived experience that emerges during the analysis of qualitative data. It can be used to establish a code with which text can be coded.

Theory
Outhwaite (1983, p.395) suggested that theory is a “body of law-like generalizations, logically linked to one another, which can be used to EXPLAIN empirical phenomena… In the social sciences the term ‘theory’ is used very loosely, and may mean no more than a set of assumptions or concepts, or a relatively abstract inquiry distinguished from empirical research or practical recommendations.” (Jarvis, 1999, p.143)

Transferability (see also Generalizability)
There are a range of studies that do not have the sample size necessary to ensure generalizability. These may be case studies, action research projects, or studies focused on the collection of qualitative data. Nevertheless, illustrating the significance of findings to larger
populations or within other contexts is still often a goal. For research of this nature, transferability can be a useful indicator of applicability. (O’Leary, 1999, p.62)

**Triangulation** (see also Crystallization)

The use of more than one method or source of data in the study of a social phenomenon so that findings may be cross-checked. (Bryman, 2008, p.700)

[Triangulation] assumes a fixed point, a single truth... (Woods, 2006, p.5)

**Validity** (see also Generalizability; Transferability)

The extent to which an account accurately represents the social phenomenon to which it refers. In realist research it refers to the degree to which the research provides a true picture of the situation and/or people being studied and is often referred to as internal validity. External validity refers to the extent to which the data collected from the group or situation studied can be generalized to a wider population. Postmodernists, who contest that research can ever provide a single true picture of the world, contest the very possibility of validity. (Gibbs, 2007, p.152)

Validity is premised on the assumption that what is being studied can be measured or captured and seeks to confirm the truth and accuracy of this measured and captured ‘data’, as well as the truth and accuracy of any findings or conclusions drawn from the data. It indicates that the conclusions you have drawn are trustworthy... validity indicates that your methods warrant conclusions. (O’Leary, 1999, p.61)
Appendix C: Permission request form

Subject: Permission request for Library report

Good morning

I met with you (some time ago now) and discussed how the Library might better support your research. Our informal conversation, together with many other similar conversations around the University, and also elsewhere, led to the Library-funded Otago Biodiversity Data Management Project: http://library.otago.ac.nz/services/projects.html

This Project has been hugely successful, generating substantial interest locally (MoRST, DOC, Landcare and others) and internationally (through presentations in Singapore and Melbourne and also meetings with staff at the Smithsonian, in Washington and the Digital Curation Centre at Edinburgh University). The Project has also provided a foundation for the current University of Otago research project, championed by Dr Nigel Stanger (Information Science): Long-term Preservation and Curation of Primary Research Data http://eresearch.wiki.otago.ac.nz/Longterm_preservation_and_curation_of_primary_research_data

As a result of all this interest, I have been asked to document the issues that were raised during the informal conversations which I had with you, and others, before and during the Project period. I am hoping to write up the findings of these conversations, phone calls and emails in the next few weeks. All data appearing in this, or any subsequent reports, will be confidential and any quotations that are used will not be identifiable; confidentiality will be ensured.

Can you please reply to this email, putting AGREE in the subject line if you will allow me to include information that I learned during our conversation.

If I haven’t heard back from you by [date supplied], I hope you won’t mind if I give you a call.

Thank you once again for your time and support

Kind regards,
Appendix D: Study sample population – further information

Total interviewees (%)
May 07 - Sept 08
- Academic: 62%
- Expert: 24%
- General: 6%
- Postgrad: 8%

Total interviewees (no.)
May 07 - Sept 08
- Academic: 48
- Expert: 19
- General: 5
- Postgrad: 6

Total timespent (mins)
May 07 - Sept 08
- Academic: 2665
- Expert: 1550
- General: 225
- Postgrad: 255
Appendix E: Emerging themes by interview phase

1. Pre-Proposal: May – August 2007

<table>
<thead>
<tr>
<th>Research group profile</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No. conversations</strong></td>
</tr>
<tr>
<td><strong>First contact</strong></td>
</tr>
<tr>
<td><strong>Repeat contact</strong></td>
</tr>
<tr>
<td><strong>Total contact time</strong></td>
</tr>
<tr>
<td><strong>Group composition</strong></td>
</tr>
<tr>
<td><strong>Contact mode</strong></td>
</tr>
<tr>
<td><strong>Location</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Interviewees May - Aug 2007 Pre-Proposal</th>
</tr>
</thead>
<tbody>
<tr>
<td>academic 69%</td>
</tr>
<tr>
<td>expert 26%</td>
</tr>
<tr>
<td>postgrad 5%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time spent May - Aug 2007: Pre-Proposal</th>
</tr>
</thead>
<tbody>
<tr>
<td>interviewees (19)</td>
</tr>
</tbody>
</table>

Research approach
- context: life sciences
- focus: management of information (unspecified)
- researcher bias: uninformed, tacit knowledge applied
- interview type: unstructured, informal conversational, open-ended enquiry
- research question: undefined, emerging
- associated methods: preliminary document research and expert advice

General comments/observations:
- establishing key information concerns of academic researchers
- registering interest in researcher-suggested project(s)
- documenting project suggestions (relating to personal collections – digital & print)
- researcher response led to ‘Eureka moment’ and the Library project proposal, presented to Library management on 14th Aug 2007:


---

40 Combined: F2F, phone and email conversations
Conversation themes, sub-themes and selected quotes (May – Aug 2007)

1. Digital management of unpublished research
All conversations included some discussion around this theme.

a. Value/support for action
There was an overwhelming positive response to the possibility that the Library may be able to assist with enhancing access to (largely unpublished) information. One researcher was “amazed that the University is not already supporting researchers and their data – it’s hard to believe”. A small number of interviewees indicated that libraries should not step outside their ‘normal business’ or at least “start with traditional library interests i.e. PDFing research…”

b. Observation/recommendation/s for action
Many of those I spoke with suggested that I contact one or more of their colleagues (within that department and also outside the University, in CRIs, such as Landcare Research); this led to a ‘snowballing’ of conversations. It was during these initial conversations also, that I was encouraged to associate any Library activity with the Life Sciences departments, the Ecology Teaching Programme and the ERG. A number of researchers made comments about the difficulties they had in managing their own websites and online bibliographies, which were ’difficult to maintain and update’. One researcher, keen to see activity in this area, gave specific recommendations about how to ‘frame’ a proposal:

“All researchers should target this area as heritage, rather than biodiversity... you may find it difficult to convince there is practical value in this but there is, without doubt, intrinsic value. This is our country, our history, our species and if we don’t record in no-one else will. It’s vital to defining who we are…”

c. Collection/s suggested
Researchers suggested numerous personal and other research collections which they would be interested in managing and/or accessing more effectively; a number of these are listed in the Library project proposal (Elliot, Aug 2007). These collections included a range of information types and formats, from published print (e.g. bibliographies) to raw data, such as images and spreadsheets. Some suggested ‘collections’ included information from a variety of institutions and individuals, beyond the University of Otago.

2. Manage/share research data – positive/neutral aspects
More than half of the conversations included some discussion about managing research information other than published or unpublished print information. This section conveys some of the positive or neutral comments that were made about managing and sharing research data.

a. Access/sharing of research data
Several researchers volunteered that they would be willing to share their data, but that they were uncertain of how they might do so. Another observed that it wasn’t until they joined the staff at Otago that they became aware of a database, which is significant in their research field.

b. Loss/re-use of data/new research potentials
The assumption that ‘earlier research actually informs conservation practice’ was challenged by one researcher, who was not convinced that this is the case. “Practitioners don’t read old sightings of species; rather they are interested in land use, current practices.”

One researcher expressed concern about what will happen to their research when they retire. Another made the following observation:
“Capture recent research data now... earlier research was used for a single outcome, then discarded and lost. A new ethos or principle should be to capture this data, digitise, store in a repository... look forward from here, before tackling a huge backlog...”

c. Current practice: data format/deposit (repositories)
Several current information sharing practices were discussed, including personal website bibliographies and the use of well-established national and international data repositories. Concern was also raised about the ability of some organisations to maintain digital collections in the longer term, with one researcher pointing to the “temporary nature of some Australian government departments”. Several researchers noted the substantial activities of Landcare Research, who build a large proportion of New Zealand’s biodiversity databases (see Part 1: Questionnaire Report).

3. Manage/share research data – negative aspects
The negative aspects of managing and sharing research data was discussed in less than half of the conversations.

a. Ownership/IP/copyright/credit for data (PBRF)
Several researchers noted unsuccessful sharing practices i.e. they had submitted data to a repository but were experiencing difficulty in subsequently retrieving this data (loss of control). Another was a “supporter of PBRF” and suggested “this may impact negatively on collaboration.”

b. Cost/complexity/workload/lack of time
Several comments here, including: “lack of funds can be a disincentive to manage e-stuff” and “be wary of digitisation... it’s a bottomless pit...” One researcher was critical of the government’s approach to KAREN, suggesting that “high speed capacity has been developed without adequate infrastructure... the potential is there but it won’t cope... the government has created a race-track, but no cars to run on it.”

4. Metadata
Just one comment here, relating to MoRST and work on policy development “basically around metadata.”

5. Role of the Library
Just two comments on the changing role of libraries, but notably from opposing perspectives i.e. “be careful to direct thinking within traditional library boundaries” and “libraries need to wake up... blurring of publication outputs and databases... what will the library be doing... not housing print – there won’t be any!”

Several researchers talked about how libraries are ‘trusted’ - “I would be keen to house this [data] in the Library and make more widely available” and an email “It would be great if the library could appropriately catalogue, care for and make available the wealth of information collected by [X]”

6. Research process
Just three comments here, including observations about the place of new digital initiatives at Otago, which “would sit well” and more general statements about ‘rapid digitisation of literature’ and the need for changes to PBRF policy (to recognise data/dataset citation).

7. Funding and collaboration
Funding and/or collaboration were mentioned in more than half of these initial conversations and some substantial opinions were shared.

a. Funding
Several researchers shared personal experiences of seeking external funding for university-based projects (can be difficult). Canterbury University’s recent success in attracting funding to digitize entomology records was noted as a ‘useful model’ and other researchers commented on the potential
for funding to PDF existing biodiversity print documents. One researcher mentioned Australia’s commitment to ‘infrastructural IT’, putting “$8 million into biodiversity alone!” while another pointed to U.S. funded initiatives, such as the ‘Biodiversity Heritage Library’ and the ‘Encyclopaedia of Life’.

b. Collaboration
Most (but not all) comments focussed on how collaboration could be improved within and across institutions. Comments ranged from the local, “closer collaboration with the University Library and information specialists could be beneficial” to the global, “...better to hook up with GBIF, than any local initiatives.” Between these two positions, comments touched on the need for universities to work more closely together, “New Zealand is a small country and collaboration is paramount to achieve goals”. Observations were shared about current collaborations (e.g. New Zealand regional councils’ Envirolink programme) and also future University collaboration potentials, working with government departments, such DOC and also the Ministry for the Environment, “this is an effective organisation and their capabilities are under-rated.”

Other comments stressed the need for a multidisciplinary approach, when considering ‘biodiversity’.

Conversely, the observation that New Zealand universities are “not typically involved with [biodiversity projects] at a national level... smaller initiatives, often at the species level” was repeated during a later conversation, “Landcare Research [rather than the University] is doing this now”.

8. IRR
Again, just a few comments here, including: one researcher claimed to send all research outputs to the Science Library already; another would like to know the research outputs of other universities and requested that student outputs be made available (and therefore known). A follow-up email contained the following remark, “IRRs are basically archival, rather than research focussed.”

9. Theses
Two comments here, one researcher expressed interest in making all theses (all levels) digitally available. Another was interested in knowing Otago’s policy, with regard to “fulltext scanning of theses.”

10. Other
A number of more specific comments, mostly: questions, relating to personal research interests; recommendations about databases e.g. ‘Catalogue of Life’ and ‘Species 2000’ and tools (such as CONNOTEA and PICTION). Further useful contacts, were also made during these early conversations.
2. Pre-Project: September – December 2007

<table>
<thead>
<tr>
<th>Research group profile</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No. conversations</strong></td>
</tr>
<tr>
<td><strong>First contact</strong></td>
</tr>
<tr>
<td><strong>Repeat contact</strong></td>
</tr>
<tr>
<td><strong>Total contact time</strong></td>
</tr>
<tr>
<td><strong>Composition</strong></td>
</tr>
<tr>
<td><strong>Contact mode</strong></td>
</tr>
<tr>
<td><strong>Location</strong></td>
</tr>
</tbody>
</table>

Research approach
- context: life sciences and research data management
- focus: management of biodiversity information
- researcher bias: emerging awareness, reflective
- interview type: expert consultation
- research question: defined
- associated methods: document research

General comments/observations:
- contextualising information concerns of academic researchers
- validation of initial (local) findings
- external focus - information gathering
- consultation with national & international experts
- reflection, ongoing iterative approach to research
- formal project proposal presented to the Library’s EMG\(^{41}\) on 30\(^{th}\) August 2007:

Clark, J. (Project Sponsor) and Andrew, R. (Project Manager) (Aug, 2007) *Biodiversity Repository Service Pilot Project: A Proposal for a joint initiative between the University of Otago Library and the Life Sciences Departments, in association with the Ecology Teaching Programme, and the Ecology, Conservation and Biodiversity Research Group*

---

\(^{41}\) Executive Management Group
Conversation themes, sub-themes and selected quotes (Sept – Dec 2007)

Attendance at an international conference, on the management of digital scientific information\(^{42}\), brought me into direct contact with data management and biodiversity experts from around the world and I had conversations with experts from the Coalition of Networked Information, UKOLN, the Digital Curation Centre, GBIF and the Smithsonian National Museum of Natural History. During this period I also communicated with several biodiversity experts within New Zealand. Topics raised during these conversations reflect the interest and focus of this (mainly) external expert group.

1. Digital management of unpublished research

a. Value/support for action
The management of unpublished research was of interest to all interviewees, indeed for a number of those interviewed this was their core business. Email comments included:

“I must say that I am quite excited by your proposed Biodiversity Information programme. It is ambitious in its scope and outcomes, but should it eventuate it will be invaluable to a wide range of scientific and social interests.”

“The biodiversity work you describe is of similar interest to me, so I wish you well with it. If you [plan to visit] please do let us know and we would be very happy to host you presenting on the project to staff here.”

b. Observation/recommendation/s for action
Most of those interviewed suggested practical ways for progressing the (now confirmed) Otago Biodiversity Project. Recommendations included: targeting individuals and institutions already active in this area (names supplied); collaboratively managing research collections (by species); linking any Otago project with established global initiatives and architecture (such as uBio, GBIF) and exploring the potentials of online tools, such as ATRIUM, for managing biodiversity information:

“[This] is an excellent shareware platform for bringing together plot observational data with weather data and GIS information. This is the only tool that readily brings together biodiversity and ecology data, for a bigger picture and more informative overview. It has ‘canned tools’ for plot methodology modelling, which is especially useful for New Zealand’s focus on climate change and species invasion – plots can be viewed over time.”

c. Collection/s suggested
Most conversations were with external experts who were not familiar with specific Otago collections. Comments related to ‘how to approach’ this challenge, rather than which collection to address, “it is also important to have something to show for the project after 12 months…”

2. Manage/share research data – positive/neutral aspects

a. Access/sharing of research data
The issue of why research data might be shared was almost redundant in these conversations. Many in this group were already investing time, energy and resources into sharing digital biodiversity information (including research data).

b. Loss/re-use of data/new research potentials
One researcher related a very specific example of how unshared data (discovered at a much later date) had skewed the results of a long-term field study. The potential research value of sharing legacy biodiversity print information and also research data, including herbaria, was also discussed.

3. Manage/share research data – negative aspects

a. Ownership/IP/copyright/credit for data (PBRF)
Three comments here, all relating to the absolute requirement that researchers be acknowledged for their data, “needs [to be] a policy change to recognise impact factors on use of this information... University rewards system should recognise use of data/datasets...” One interviewee went further, suggesting that we “publish – don’t share data”.

4. Metadata
Half of this group talked about the importance of metadata. Comments and suggestions focussed on appropriate metadata standards and also around individuals and groups already working in this area, including: Biodiversity Information Standards (formerly known as TDWG), National Biological Information Structure, INOTAXA Project (web workspace for taxonomic descriptions, catalogues, names specimen data, images and other resources) and the value of Darwin Core and ABCD (metadata standards for the life sciences).

5. Role of the Library
Several, quite different, perspectives on the role of the library in the new digital research environment: the library is ‘trusted’ and should have a role in managing research data in future; data should be managed ‘at source’ i.e. the library should “work with the research scientist, curating their research data at source”. One interviewee ‘discussed the issues with library catalogues and how they are of little use to taxonomists (information structured very differently)... need to extend granularity of information” and the following advice was also shared: “A huge culture change is coming to academia as a result of digitisation.” [my emphasis]

7. Funding and collaboration
Comments included the importance of collaboration and communication when seeking funding, “[X]stressed the need for 'collaboration’ with any digitisation project and pointed out that universities were too competitive in their approach” and “communication is key”. The challenges of “long-term sustainability, funding sources, project based start-up funds v ongoing maintenance costs” were also raised.

Comments also tended to be both externally and internally-focussed, “look at the Canadian universities, who recently committed six billion to biodiversity initiatives’ and “[this is] an excellent opportunity for departmental collaboration, learning in action and groundbreaking for New Zealand – including the research community...”

8. IRR
There was just one, very practically-focussed, conversation about institutional repositories touching on current practice and potentials for the future.

9. Theses
No comments

10. Other
Further comments were made about: specific software, tools, initiatives and practices (including DOIs, Australian and U.S. activities); engaging with New Zealand research and researchers and also ‘staying in touch’ with future developments and initiatives.
3. Pre-Questionnaire: January – March 2008

**Research group profile**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>No. conversations</td>
<td>20</td>
</tr>
<tr>
<td>First contact</td>
<td>17</td>
</tr>
<tr>
<td>Repeat contact</td>
<td>3</td>
</tr>
<tr>
<td>Total contact time</td>
<td>17.25 hours</td>
</tr>
<tr>
<td>Composition</td>
<td>80% academic</td>
</tr>
<tr>
<td>Contact mode</td>
<td>F2F, email</td>
</tr>
<tr>
<td>Location</td>
<td>office, coffee, remote</td>
</tr>
</tbody>
</table>

**Research approach**

- context: biodiversity and research data management at Otago
- focus: management of biodiversity research data
- researcher bias: increasing level of understanding, informed
- interview type: semi-structured, general guide
- research question: defined
- associated methods: consultation with local interest groups, ongoing document research

**General comments/observations:**

- establishment of Library/Faculty advisory/governance group\(^{43}\)
- establishment of Library reference group\(^{44}\)
- further communication with Otago researchers with an interest in biodiversity
- move from inductive to deductive analysis
- development of a local questionnaire, based on three international data management surveys:

Elliot, G. (April, 2007) *Otago Biodiversity Project Questionnaire*

\(^{43}\) Biodiversity Data Management Advisory Group (BDMAG)
\(^{44}\) Biodiversity Library Reference Group (BLRG)
Conversation themes, sub-themes and selected quotes (Jan - Mar 2008)
The Otago Biodiversity Project officially commenced at the start of 2008. The Project was now clearly defined in the Pilot Proposal (Clark, 2007), influencing conversations during this period and the following two periods, April - July 2008 (with Questionnaire) and August - September 2008 (post-Questionnaire).

1. Digital management of unpublished research

a. Value/support for action
All general comments in response to the value of managing biodiversity research data were positive e.g. ‘...definitely interested in this project, thank you for being persistent...”, “this project is about more than biodiversity... it is extending the boundaries of research”, and in a follow-up email, “let’s hope the momentum persists.” Most interviewees put forward specific recommendations for action, rather than simply making general supportive statements (see below).

b. Observation/recommendation/s for action
Recommendations and suggestions included that I should: examine department [X’s] unique collections; promote the Library project at a University lunchtime seminar; establish an interested governance group (to monitor project activities); share project activities with research students and collaborate with other parts of the University (two respondents, details specified). Other comments included the importance of keeping the project regional and the need for an ‘academic champion’; two researchers offered to be the ‘points of contact’ [with others/within the department]. One respondent summed up their interest in an email:

“So the opportunity remains here for us at Otago to develop some ground-breaking ideas and practices (at least for Australasia) in this area of electronic data management. We begin with a pilot study involving Biodiversity in the Zoology Dept. and possibly [individual’s name] and proceed from there. I am excited about being part of this initiative.”

2. Manage/share research data – positive/neutral aspects

a. Access/sharing of research data
Almost half of the interviewees commented on the value of sharing publicly funded data (for the public good) with one adding that “perhaps researchers working in biodiversity are more likely to wish to share their data”. Specific observations about data sharing included how a ‘lack of access to climate data (several years ago) had impacted on farmers suffering from extreme weather conditions’ – “now information is shared every three months.” Another interviewee observed that “private companies who gather data on wildlife for their own ...benefits/information... may be prepared to contribute (if a structure was in place)”. While the following statement was made about data sharing in the UK:

“With respect to the research councils' strategies, the statement [publicly funded research data are a public good] remains true, if not more so. Five of [the] seven Research Councils now have data management/sharing strategies, and the Higher Education Funding Council for England (HEFCE) has recently decided to sponsor a feasibility study into managing research data on a national shared services basis.”

another comment was less positive about the local situation:

“...there are numerous [global] initiatives to open up access to publically funded research, yet in New Zealand we seem to be going the other way.”

However, at least one positive local example was also shared:
"It was moved... that as a matter of policy we share biological data from Orokonui with other organisations with the proviso that our staff don’t have to do any reformatting or data input and also with the proviso that Orokonui is acknowledged within any documents that use the data”.

b. Loss/re-use of data/new research potentials
A number of researchers raised concerns about losing data, and how this impacts (negatively) on new research potentials. The following illustration of the ‘stop/start’ nature of some research projects was shared by one academic:

“Biodiversity projects start, funding dries up, the student puts the research away and forgets about it - no longer their concern. The research is effectively lost and future researchers need to start from scratch.”

Another put forward the following suggestion:

“Would it be too unreasonable to add to the existing university regulation requiring a student to deposit two hardcopies of a thesis in the library, to also deposit a CD with key data from the thesis?”

c. Current practice: data format/deposit (repositories)
There were just three comments here; one researcher explained their department’s local solution for managing research outputs (EndNote Library), while two others described their experiences of accessing data from international sources.

3. Manage/share research data – negative aspects

a. Ownership/IP/copyright/credit for data (PBRF)
Several interviewees addressed the legal aspects of data sharing, “ownership, this is certainly a concern and something to be examined in detail...researchers are understandably sensitive about who uses their [data], where and why; acknowledgement is critical...” A couple of responses were couched quite strongly “[some departments] don’t share...they want credit for their data” and “[some researchers] only cite articles written by other academics [from their own particular institution] (to keep PBRF’s up)”

Cultural considerations and security issues were also of concern to two researchers, “cultural sensitivity...again an issue with native/endemic species” and “[X] stressed the anxiety surrounding the ongoing safety of endangered species...”

b. Cost/complexity/workload/lack of time
Just a single comment about how workload and lack of time makes it difficult to effectively share data, “[X] felt this time of year was too busy, for either himself or his postgrads...”

c. Quality/reliability of data
Again, just one interviewee expressed concern about the quality and reliability of data captured,

“Combining different sets of data within the one Excel spreadsheet presents huge problems of interpretation and certainly brings into question the possibility of ‘re-use’ – who would want to, or even could reuse this data?”

4. Metadata
Less than half of those interviewed discussed metadata. Comments included general statements about the value of standardisation and interoperability, the complexities of metadata, “it is the metadata that is the challenge” and more discipline-specific comments about ‘significant descriptive metadata in the life sciences’ i.e. name/taxonomy, location, time and date. One emailed comment was particularly illuminating:
“...but without a reasonable amount of knowledge of how [the data] was collected, stored, etc. it wouldn't be usable by others -- and currently that meta-data is not associated with the dataset.”

5. Role of the Library
Several researchers indicated that the Library should have a role in managing research data, “[it is] vital that any project is Library-based” citing the Library’s ‘neutrality, communication and ability to collaborate with all departments’. One interviewee suggested that the Library should “focus on sharing biodiversity ePrint resources”.

6. Research process
Just three comments here, about the value of Open Access (from international experience), the statement about the UK Research Council strategies (see 2a. Access/sharing of research data) and a general comment about “the role of data scientists and the need to reconsider how data is collated and managed, before researchers are in the field.” [my emphasis]

7. Funding and collaboration
Two thirds of all interviewees shared examples and possibilities of collaboration and/or funding. Some of these conversations contained substantial local detail (not reproduced here).

a. Funding
Several interviewees spoke of specific funding agencies and potentials for project funding; TFBIS\(^45\) funding was known to two researchers, funding success with FRST\(^46\) was shared by another researcher and another pointed to the possibility of TEC\(^47\) funding,

“TEC – collaborative - initiative is looking to improve access to [other specified information] resources across universities, CRIs, government departments and local councils – would TEC funding also be available for this biodiversity initiative?”

b. Collaboration
Many researchers are already routinely collaborating with others, within their normal research activities. These collaborations are with a wide range of individuals and groups, including: other departments; other New Zealand Universities; Private Training Establishments (PTEs); Maori organisations (locally and nationally); regional councils; CRIs: NGOs and government departments.

Other observations about collaboration were: directive, i.e. the Library project ‘must be a collaborative effort’ and the project must retain a ‘regional focus’; supportive, for example, “[I] would be willing to collaborate on an Otago research data project” and critical. One researcher shared detailed examples to demonstrate how little collaboration currently exists within New Zealand, with regard to biodiversity initiatives, suggesting that things were “less than ideal” or that researchers were “talking past each other.”

8. IRR
Only two researchers had any knowledge of Institutional Research Repositories and both were uncertain about Otago’s activities in this area.

46 Foundation for Research, Science and Technology (FRST), “We invest close to $500 million a year in science and technology research on behalf of the New Zealand Government, seeking benefits to New Zealand’s economy, environment and society.” Retrieved 20th June 2009 from: www.frst.govt.nz/
9. Theses
Several interviewees shared comments or had questions relating to how Otago theses are, or will be, made available in future. Comments included:

“[I am] unclear of the University’s position with regard to full text scanning of theses.”

“One thought I have had... is how to get the information collected in a hard copy thesis, into a machine readable/data base format?”

“...in the near future all PhD theses (including Otago theses) should be visible”

10. Other
More than half of the researchers in this ‘Pre-Questionnaire’ group are from disciplines outside the life sciences; they were recommended to me because they are working in the areas of biodiversity and/or digital management.

‘Other’ comments and concerns raised by this group were wide ranging, including: the global lack of taxonomists; the reduction in New Zealand science print publishing; the importance of biodiversity data modelling; defining ‘what is biodiversity’ and the need for a digital “biodiversity landscape diagram [showing] complexity of relationships, key players, databases, infrastructure.”

**Research group profile**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>No. conversations</td>
<td>35</td>
</tr>
<tr>
<td>First contact</td>
<td>27</td>
</tr>
<tr>
<td>Repeat contact</td>
<td>8</td>
</tr>
<tr>
<td>Total contact time</td>
<td>30.5 hours</td>
</tr>
<tr>
<td>Composition</td>
<td>66% academic</td>
</tr>
<tr>
<td>Contact mode</td>
<td>F2F, email</td>
</tr>
<tr>
<td>Location</td>
<td>office, coffee, remote</td>
</tr>
</tbody>
</table>

**Interviewees Apr - Jul 2008 with Questionnaire**

- academic: 66%
- expert: 17%
- general: 6%
- postgrad: 11%

**Apr - Jul 2008 with Questionnaire**

**Research approach**
- context: biodiversity and research data management at Otago
- focus: management of biodiversity research data (current practice)
- researcher bias: informed, increasingly directive
- interview type: semi-structured, associated with the *Otago Biodiversity Project Questionnaire*
- research question: defined
- associated methods: consultation with Project groups, ongoing document research

**General comments/observations:**
- first meetings of the Library/Faculty governance group\(^{48}\) and Library reference group\(^{49}\)
- 43% of this group completed the Questionnaire (following the interview)
- co-presentation to all Library staff on Library support for e-Research at Otago:


---

\(^{48}\) Biodiversity Data Management Advisory Group (BDMAG)

\(^{49}\) Biodiversity Library Reference Group (BLRG)
Conversation themes, sub-themes and selected quotes (April - July 2008)
23% of these interviews were with academics and external experts who had already talked with me about the current Library Project. 43% of those interviewed also completed the Otago Biodiversity Project Questionnaire (this was shared during the interview)

1. Digital management of unpublished research

a. Value/support for action
Three researchers indicated that their primary interest was in better management of key print/eprint resources, not currently managed by the Library. However, the majority of interviewees expressed support for the current Project focus, “The whole subject is of great interest to me, even though I am in [discipline] rather than biology” and “idea of sharing research data would be wonderful”.

b. Observation/recommendation/s for action
Many interviewees recommended further contacts, both within and beyond the Life Sciences. Some specific recommendations were also suggested for the Library Project including: the need for an ‘academic champion’; the importance of knowing what collections currently exist at Otago and the value of collaboration (particularly with the Research Office). One interviewee suggested:

“I see the bulk of the issues are around... document management/institutional repositories... research data repositories... [and] GIS data management...”

2. Manage/share research data – positive/neutral aspects
Over half the researchers in this group shared personal experiences about the value of managing and sharing research data. A recurring observation was the “intrinsic value of the data” i.e. “data has value for its own sake.”

a. Access/sharing of research data
Several researchers were interested in combining different datasets, “very interested in revealing and layering data... adding [discipline data] layers to topographic maps.” Others remarked on why the pressure to share data in the life sciences may be more compelling than the reasons not to share, such as the threat of mass extinction.

b. Loss/re-use of data/new research potentials
Data loss was an issue for a number of researchers, with one email respondent providing a detailed example of the problem:

“I lost to "updating" the entire set of digitised outline maps of [X]! The originals were on 8 inch discs for the PDP computer of the 1980s, and this machine, along with the disc reading programme was replaced. A tape copy of the data... was made as a contract for [Y]... but in the process the 'mappers' changed quite a lot of the original.”

Several researchers discussed the enduring value of research data, “the value of [long term studies] are tremendous, so whether or not the data is shared, it should be managed... Hocken would have a keen interest in this purely from a New Zealand archival perspective.” Another email concludes, “From what I can gather, my data is seen to be quite valuable. I really was on the verge of throwing all of that stuff out.”

Some interviewees also addressed the potential to create new research:
“...sharing research data would be fabulous for [my] students, who look at published research, but currently are unable to repeat the experiments as they are unable to access the original data. They could re-test or possibly build new areas of research.”
c. Current practice: data format/deposit (repositories)
Two researchers indicated that it was easier to get data from (trusted) international online sources, while others shared concerns with current, local, practices: “Lots of files on variable discs in offices... unreadable, need assistance now... data being lost”; “issues of who updates, maintains database, also issues of access (must be on campus to use this)” and “Central management of research data... would eliminate the proliferation of ‘bad’ practices.”

3. Manage/share research data – negative aspects
Just under half of the interviewees expressed concerns about the negative implications of managing and sharing research data.

a. Ownership/IP/copyright/loss of control/uncertainty
About a third of this group raised concerns about ‘ownership’: “I have worked with [X]patient data which could never be shared”; “a major issue with this is copyright, as this resides with [Y] and no-one is quite sure who can okay ‘not-for-profit’ digitisation,” and “when staff leave, they may take a copy of their work, but what happens to the original, which is owned by the University?” One researcher was particularly concerned about research data being managed by a third party:

“There is no function in my vision for centralized repositories other than as an archive for stuff that nobody wants to actively curate.”

b. Cost/complexity/workload/lack of time
Where ‘cost’ was felt to be an issue, concerns focussed mainly on technical issues, i.e. technical obsolescence and lack of infrastructure:

“KAREN is a great big pipe which ends in a bit of string... infrastructural backbone doesn’t extend to within the institution itself.... capabilities are partial only and problems remain for researchers, storing, managing and sharing their own data here at Otago.”

One researcher raised a further concern, relating to career stage:

“The issues for established researchers are quite different. Time is a huge issue and they already have established patterns of research. They also have collections and will be more interested in deposit of (already formatted) data than needing to access the data of others.”

c. Quality/reliability/credit for data (PBRF)
Data quality and reliability concerned a number of researchers and some compelling personal experiences were described:

“Quality control of data is an important issue and something that the University is (understandably) acutely aware of... its reputation is at stake... I have lots of ‘unpublished data’ in my office, which I wouldn’t necessarily wish to reveal or share with other researchers.”

“The research data, associated with the published research is also included but only that research which has been through some validation/quality control process or process of peer review. If this isn’t done, the data is regarded as not worth sharing and is not revealed.”

‘Digital manipulation is of concern i.e. maintaining data integrity (difference between ‘cleaning up’ the image and actually changing the imaging to something different; also the importance of appropriate imaging when capturing data.’

One researcher also talked about the data gathered by their own students, “The data they collate is good, but not verified to the highest standard and therefore must be discarded.” Others felt that some researchers collate data ‘badly’ (at least some of the time), making it difficult to re-use. The following quote from one researcher is revealing:
“Most research data ends up as... semi-structured data. Invariably these are tables in Word. If you’re lucky the [researcher] is more tech savvy and puts them into Excel. At that point they think they have ‘structured databases’ where in most instances research data requires more complex structures than could ever be expressed naturally in Excel tables. Living under this illusion when they want to analyze data many will spend days integrating that data, massaging it into a form where they can do a particular piece of analysis. Often they will change the data without recording how or what they changed. Then if they need to do it again they frequently have to start all over again.”

4. Metadata
Less than a third of this group provided comments about metadata. Several researchers addressed the need for ‘context’, “Data in isolation are largely meaningless.” Others mentioned the importance of adequate standards and the importance of naming/taxonomy in the life sciences, “...without sound taxonomy, there is no point in sharing research – becomes meaningless.”

One researcher spoke of the increasing interest in (and potentials of) ontologies and the semantic web,

“In the case of organism names that would be the ‘synonyms and classifications’ so for example the user could ask for all data-sets that contain information on the daisy family with all the names standardized to the currently accepted name.”

5. Role of the Library
Around half the researchers shared information about how they perceived the role of the library and almost all of these comments related to the ‘changing responsibilities’ of libraries. Comments targeted three key aspects:

What role?
A surprising number of researchers had ‘forgotten’ about the Library, “I never go to the Library... It would be nice to visit.” and “I don’t normally think of libraries having much of a role.”

Libraries have a role
Others expected libraries to embrace new challenges in the rapidly changing digital information environment. One researcher queried why the New Zealand Digital Content Strategy lacked a national direction or focus, with regard to the management of research data, while another researcher articulated why librarians should be interested in managing data:

“Librarians manage ‘stuff’ and it shouldn’t matter what format. The process of change in institutions can be slow, but this shouldn’t stop libraries from moving forward, managing their own business, in the best way they know. They have a duty of care for ‘stuff’. It’s important to recognise the distinction between STORAGE and ACCESS. Libraries should be dynamic, keep building, heading standards and then deal with the politics of ACCESS when it is appropriate. Just because the institution doesn’t support OA now, doesn’t mean it won’t do so in future. We can’t wait for these decisions or we’d never get anything done! Librarians undervalue their own skills – they are too modest. Yes, collaborate with ITS, departments, but librarians have huge competencies and they shouldn’t be afraid to use them.” [my emphasis]

Libraries have no role
Several researchers were yet to be convinced that libraries and librarians had a role in managing research data, as articulated by one researcher:

“Most scientists see [libraries] as providing base-level support functions and are justifiably suspicious that they can do much more than that. Managing research data... goes beyond such base-level support. It requires cross-disciplinary domain expertise within that support function.”

50 One researcher expressed a strong interest in discussing Library research support, rather than anything else.
6. Research process
A quarter of interviewees provided comments relating to the research process and three themes emerged:

i. The value of managing research data at the earliest phase of the research life-cycle, “this is suggesting a need for a ‘seismic’ shift in thinking/attitudes and practice (think Kuhn).”

ii. The publishing process “...to get something published there has to be some new methodology, justifying a publication. This may mean lots of data are never published (or retained).”

iii. The need to collate data in re-usable format/train emerging researchers

“The research advisors are critical to the research process and highly relevant to researchers. How research is done (collation, management of data) is fundamental to this Project and it is the emerging researchers who should be supported into this process.”

“The Popperian notion of falsifiability in science requires that these analyses are reproducible by others, and yet most of the time we do not preserve the data to easily allow this. Science is not generally ‘auditable’ in this sense. The new landscape of... plug & play analysis and modeling tools, workflow tools etc, all require that this structured, auditable data access infrastructure is in place. In many areas we are probably waiting for a generation of old school scientists to retire before things will change. Those who do change and keep up will leap ahead quite dramatically.”

7. Funding and collaboration
Collaboration and/or funding were discussed by two thirds of those interviewed.

a. Funding
Several researchers were disparaging about New Zealand funding, comparing this with funding support overseas:

“[This New Zealand Government initiative] has a chequered history, initiated in 2006, then erratic funding and support thereafter... was to share data, but this hasn’t yet eventuated and last meeting was in December 2007.”

“Cornell’s eBird... are going to assist with NZ material” and (slightly tongue-in-cheek), “If New Zealand cares so little about its own identity (unique biodiversity, science publications) then we would be better off as another state of Australia – better funding!”

Two New Zealand funding sources were also suggested, i.e. TFBIS51 and MoRST contestable funding.

b. Collaboration
A number of examples of current collaborations (with other University departments and New Zealand organisations) were shared. One researcher also discussed the “slight relaxation in the 'separateness' of CRIs52 and universities... better communication, some collaboration and sharing... NIWA53 shared their West Coast coastal photos... (wouldn’t have happened in the past)”

Several researchers, from both within and beyond the University, also expressed interest in collaborating with the Library (on similar projects):

“[I am] ringing to see if there is any way the Library could work with [X] on developing a web-based data system...”

51 Terrestrial and Freshwater Biodiversity Information Systems (TFBIS) Programme; Ministry of Research, Science and Technology
52 Crown Research Institutes
53 National Institute of Water and Atmospheric Research
8. IRR
Around a third of this group knew of, or had an interest in, knowing more about, Institutional Repositories. The majority of the comments were positive, “...very interested in the Library’s work in this area as [I have] many documents I would like to see included in this repository.”

Several concerns and questions were also raised, around: standards, “who applies the standards – KRS, National Library or Institutional contributors [at] Otago?”; impact on researcher practice, “...generally cautious about ‘process of deposit’ – workload and copyright issues are the main concerns” and PBRF, “...keen to know how (or if) the Library will accommodate unconventional, peer-reviewed staff outputs, eligible for PBRF.”

There was also a further comment about the role of the National Library:

“I remain deeply unconvinced by the current national institutional repositories approach. I see the National Library ‘in charge’ of implementing the Digital Strategy and doing so with no concept of ‘repositories’ as anything other than ‘library documents + metadata’.”

9. Theses
Just two comments about theses, both around which theses are to be included in Otago’s IRR (PhD, Masters) and when this will happen.

10. Other
Other comments covered a wide range of issues, including: managing digital manipulation (difference between ‘cleaning up’ a digital image and actually changing the image to something different); the need for enhanced visibility (who is doing what in New Zealand with regard to repositories, data management); University responsibility re sustainability and biodiversity (and what Australia is doing) and the merging of New Zealand science publications into a single online equivalent (one researcher stressed the immeasurable value of these print publications, in terms of content, uniqueness, identity, research merit and nomenclature/taxonomic requirement).
5. Post-Questionnaire: August - September 2008

Research group profile

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>No. conversations</td>
<td>11</td>
</tr>
<tr>
<td>First contact</td>
<td>5</td>
</tr>
<tr>
<td>Repeat contact</td>
<td>6</td>
</tr>
<tr>
<td>Total contact time</td>
<td>9 hours</td>
</tr>
<tr>
<td>Composition</td>
<td>64% academic</td>
</tr>
<tr>
<td>Contact mode</td>
<td>F2F, email</td>
</tr>
<tr>
<td>Location</td>
<td>office, coffee, remote</td>
</tr>
</tbody>
</table>

Research approach

- context: research data management at the University of Otago
- focus: management of University of Otago research data (particularly biodiversity)
- researcher bias: directive
- interview type: strategic, re ‘way forward’, valid research
- research question: defined
- associated methods: consultation with Project groups

General comments/observations:

- most of these communications were repeat contacts, with interested experts
- comments and feedback were more strategic, in line with the Project goals
- working towards final Project report
- developing Project recommendations
- sharing of findings of questionnaire report (locally, nationally, internationally):

Conversation themes, sub-themes and selected quotes (Aug - Sept 2008)
Half of these meetings were repeat conversations with interested academics.

1. Digital management of unpublished research

a. Value/support for action
‘Digital management of unpublished research’ continued to generate positive feedback and many of these conversations were followed up with emails: ‘I am most interested to be involved in your plans...Happy to support your initiative’; ‘i am very keen to help any way we can with your project’ and ‘I'm so heartened to hear of your work and wish you all the best in this endeavour’. Another respondent made the following observation:

I'd just like to highlight that not only is the work you're doing of value to the academic community, but may also be very useful to government agencies and other organisations.’

b. Observation/recommendation/s for action
Substantial recommendations were made by most (now more informed) interviewees. These included:

- speak with the Research Office;
- focus on funding;
- take a national approach (‘biodiversity isn’t bounded by regions anyway’);
- pay attention to the ‘cultural dimension’;
- aim for ‘proof of concept phase’;
- look for collaborative opportunities (for example Sun Microsoft and the National Library).

Further contacts were also suggested (particularly senior researchers) and practical suggestions were shared for dealing with both quantitative and qualitative data (Project reports). I was also advised not to ‘marginalise interest’ by focussing too much on the interests of one department, suggesting researchers in other parts of the organisation may also wish to be involved.

2. Manage/share research data – positive/neutral aspects

a. Access/sharing of research data
Comments here relate to the practicalities of sharing: ‘I sort of know where things are’, and (in an email) ‘While much information exists in theses and publications, the raw data would be of most use to us. Accessing this is not necessarily straight forward.’

3. Manage/share research data – negative aspects

b. Cost/complexity/workload/lack of time
Three separate concerns were raised by three interviewees i.e: lack of time (‘too busy at this time of year’); ethical considerations (especially within the area of health) and cultural responsibility (particularly with regard to Maori and endemic species).

4. Metadata
One researcher discussed the importance of standards for data migration, while another stressed the importance of key descriptive metadata within the life sciences, i.e: name/taxonomy; time/date and location/place (noting the value of the New Zealand Map Grid). A third commentator emailed:

In addition, information regarding the methodology used is also important, especially if data sets from a variety of sources is to be combined.

5. Role of the Library
Just two comments here, one restating the importance of ‘trust’ when managing data, the other questioning who should be leading activities in the area of data management (National Library?)
6. Research process
The ‘conservativism’ of most research (particularly in New Zealand) was raised as a concern by one interviewee, i.e. the requirement to specialise to a high degree, encouraging excellence in very narrow fields, ‘does not encourage cross-fertilisation and openness to change’. Further comments related to: time pressures (short time frame for many research projects); grant funding (the need to allocate a % for data management); use of appropriate file formats (for data use and reuse) and the value of a greenfield\textsuperscript{54} approach.

7. Collaboration and funding
About half of the interviewees supplied comments about collaboration and just one substantial comment regarding funding.

a. Funding
There was just one substantial comment which addressed the need for a national approach to funding, i.e. ‘a regional approach will not attract sufficient/sustainable funding (and biodiversity is not bounded by regions anyway)’. Recommended an approach to national funders and interested parties (Landcare, DOC, MoRST, MiE, National Library) sooner, rather than later – ‘not an easy task, but vital’.

b. Collaboration
Examples of collaboration included: sharing an international ‘multidisciplinary approach to scientific investigation’ (citing an example of a more holistic approach); working more closely with the Research Office; working with industry (citing the successful example of Sun Microsoft and National Library collaboration) and the importance of genuine consultation with Maori:

\textit{Ticking the boxes approach is not necessarily the only, or even the best, way to consult.}

Two researchers expressed interest in a symposium to further explore the implications of data sharing.

8. IRR
Two comments here, one relating to the School of Information Management’s ‘National survey of the perceptions of academic staff on institutional repositories’ (uncertainty about both the letter and IRRs). A second interviewee believed all their data was available through institutional and other repositories, but was uncertain which ones.

9. Theses
No comments

10. Other
One researcher pointed to an article which confirmed the University of Otago as a significant producer of high quality research in the area of biological conservation.

\textit{Otago is the only New Zealand university listed with 13 other southern universities as being ranked equivalently with US and Canadian universities in the top 40: Otago is 8/14 and equivalent to the top 20 North American rankings.}

The other comment related to trust (and a story was shared). Trust can be difficult to achieve but once gained it is invaluable – particularly when trying to make things happen; to make a difference.

\textsuperscript{54} Greenfield: A brand new installation of equipment without the requirement of integrating existing systems. Contrast with “brownfield,” which is an upgrade to an existing system. Coined by the building industry to refer to clean, undeveloped land (greenfield) vs. contaminated land or land with existing structures (brownfield), the term may be used to refer to network installations. From PC Magazine Encyclopedia, www.pcmag.com/encyclopedia_term/0,2542,t=greenfield&i=43956,00.asp retrieved 27 July 2009