Terrestrial and Freshwater Biodiversity Information System (TFBIS) Programme

What's in a name?

New Zealand Organisms Register Scope

February 2007

Prepared for the TFBIS Programme by Julian Carver, Jerry Cooper, Marianne Vignaux and Aaron Wilton

Table of Contents

Summary

In one of Shakespeare's most famous plays, the love struck Juliet says "What's in a name? That which we call a rose by any other word would smell as sweet". Names are not the things they describe. They are however the primary way we have of referring to things in the real world. The relationship between names and organisms is complicated. Our concept of what constitutes a particular species can change over time, and the names we use to refer to them change with it.

Over 250 year ago Carolus Linnaeus established conventions for the naming of living organisms and their placing in the 'tree of life' (or Linnaean taxonomy). These standards became universally accepted in the scientific world and have allowed biologists to name organisms in a consistent way, share and communicate those names, and debate and manage ongoing changes in species classification.

With technology now available in the information age organism names and taxonomic data are able to be stored in databases and shared between different organisations and countries. A number of countries have established national species databases, and there are several international initiatives creating frameworks and standards for the aggregation of these data globally. How is New Zealand placed in this setting?

In a draft Import Risk Analysis of corn seed from Australia, the fungus *Botryodiplodia theobromae* (pod rot of cocoa) was considered to be present in Australia but not in Samoa. As a result it was classified by Samoa as a regulated quarantine pest. Using the NZFungi system and web site however it was shown that the current name for this species was *Lasiodiplodia theobromae*. The specimen records showed that there were many recorded instances of this in Samoa. Further checking in Gerlach (1988) showed this species had indeed been recorded for Samoa under *Lasiodiplodia theobromae*, and it could also be found in CABI Crop Protection Compendium (2003) under both names.

The New Zealand Fungal Herbarium (PDD) website indicates that the N.Z. MAF status for *Lasiodiplodia theobromae* (on the BORIC website) is non-regulated, as the species occurs in New Zealand. The 3 PDD herbarium records for New Zealand for the species are only for one occasion in Avondale, Auckland. Checking the description available on the website for those records shows that the author reporting the fungus (Dingley) had found the fungus in association with sprouting tropical sweet potato in a market gardening area (Avondale) in the early 1960s. There are no further records in PDD, suggesting the species did not survive in Auckland at the time. This casts doubts on why it is classified as a non-regulated organism, as this occurrence record is not a natural occurrence.

So, because of a lack of understanding of synonymies and current names Samoa thought it didn't have something that it did. New Zealand on the other hand, because of a lack of awareness of, and correct interpretation of available presence/absence data, thought it did have something that it didn't. In cases like this, not having access to correct information can have significant implications for trade, biosecurity and biodiversity management.

Despite this, and many other similar examples, New Zealand has made significant advances in managing organism names. Landcare Research has developed a world leading framework for the management of organism names and taxonomic data and is already using this to provide up to date names in real time to a number of other organisations. Te Papa has sophisticated taxonomic data for a number of organism groups in its collection management system, and is making its collections data available online. The Species 2000 New Zealand project has almost completed a checklist of all known native and naturalised organisms in New Zealand. Staff from Landcare Research and NIWA are involved in international standards setting for the exchange of taxonomic data, and are members of the members of the Species 2000 Global Committee.

By building on these advances New Zealand has the opportunity to create a single organisms register for the country. This will assist greatly in national and regional biodiversity and biosecurity management, and will enable New Zealand to contribute to important global initiatives. This project is called the New Zealand Organisms Register (NZOR) and is the subject of this scoping report.

What is NZOR and why do we need it?

All biodiversity information systems use the names of organisms as a fundamental identifier. Names provide the essential vocabulary by which we discover, index, and link information relating to biodiversity. Access to an authoritative list of names and their relationships to species (taxa) is key to supporting information management across the conservation, biosecurity, and biotechnology sectors.

There is currently no single, definitive registry of the over 100,000 organism names relevant to New Zealand. Because of this many agencies currently each maintain their own lists of taxonomic names in isolation from each other, in different formats, and at different levels of depth and quality. The absence of a definitive source of taxonomic names means that resources are wasted through duplication of effort, there is increased expense to end-users in having to access multiple sources, and increased risk of confused decision making.

NZOR is a project to address this issue. The project's vision is to create an accurate, authoritative, comprehensive and continuously updated catalogue of taxonomic names of all New Zealand biota and other taxa of importance to New Zealand. This catalogue will be electronically available through one or more portals, and will be directly integrated into biodiversity and biosecurity systems used by central government ministries, departments, and agencies, local government, research institutes, NGOs and the wider community. The catalogue will be based on internationally agreed standards and will include organism names and synonymies, origin and occurrence data (presence/absence) and where possible alternate and historical synonymies. In the future it will link to information from other sources on aspects such as threats, ecology, distribution, use, management status, published material, keys for identification, and all collections, observation and survey data. As such it will form a key part of New Zealand's bioinformatics infrastructure, supporting scientific research and biodiversity and biosecurity management.

What will it include and how much will it cost?

The scope of the system includes all terrestrial, freshwater and marine organisms that are relevant to New Zealand. This includes native, introduced, and unwanted organisms. It will allow for scientific names, synonymies, and taxonomic hierarchy data, along with tag, vernacular, and trade names.

The initial three year implementation project will provide the technology platform, baseline data and governance structures to lay the foundations for achieving the above vision. The project will involve population of NZOR with existing digital sources of data from NIWA, Landcare Research, and Te Papa supplemented by data from the Species 2000 New Zealand lists, and from global sources (such as Species 2000/Catalogue of Life) relevant to New Zealand and of known provenance and quality. It will also include a gap and priority analysis for further building NZOR content through contributions from identified additional providers. The project will deliver tools to support initial and future data providers and tools to support end users to adopt and integrate NZOR information and services into their systems. It will also provide web based access to allow users to search current taxon concept information and view and download lists of organism names. The project will also deliver a functional governance structure with responsibility for ensuring quality of service and data.

While marine data is outside of the formal remit of TFBIS it is proposed that this is included in the initial implementation project. This is because it will be of real value to end user agencies such as DOC, MAF, and MFish. It is also because the cost of including marine data will only be marginally extra if it is included in the initial data population phase. This is due to bulk of effort in this phase being the development of transformation services to provide the data in a form compatible with the NZOR model, and the development of routines to automate the extraction of data from the Species 2000 New Zealand text documents. Because there are such significant efficiencies to be gained by addressing all organisms, the steering team for this Scope strongly recommend that if NZOR is funded that it include marine species.

There are a number of options for the base technology platform for NZOR, including the GBIF Electronic Catalogue of Life (ECAT), The Berlin Taxonomic Information Model, Landcare Research's Dynamic Taxonomic Framework, or a complete ground up development. None of these systems currently satisfy the full range of identified provider and end-user needs for NZOR and all would require some customisation. Based on a cost analysis of each of the above options, re-use of existing New Zealand technology appears to give the most cost effective option, while still ensuring that NZOR provides the concept based, authoritative taxonomic system that is required.

There are a range of potential funding sources for the ongoing maintenance of NZOR and the addition of new data, once the three year implementation project proposed in this scope is complete. These include MoRST funding for science 'backbone' infrastructure, FRST funding, and operational funding from agencies requiring additional data. During the development of this scope MoRST, DOC, MAF Biosecurity, ERMA and MFish have expressed their support and enthusiasm for this project.

The costs for the three year implementation project are $302,000, $337,000 and $305,000 for years one, two and three respectively. In funding this system, TFBIS is uniquely placed to take a leadership role by creating infrastructure and processes that support terrestrial and freshwater biodiversity management, and have benefit to marine and biosecurity concerns as well. In so doing TFBIS will have not only fostered, but followed through on real interagency collaboration, and will have contributed an important component to New Zealand's environmental information systems landscape.

1 Introduction

1.1 Project Context

This project was first envisioned at a Terrestrial and Freshwater Biodiversity Information System (TFBIS) commissioned workshop in August 2005. The purpose of this workshop was to clarify issues associated with taxonomic names and associated databases, identify options for resolving issues, and recommend a way forward. The workshop was attended by twelve people including leading biosystematics experts across the major fields of study, experts in biodiversity informatics from Landcare Research and NIWA, and representatives from end user agencies including as DOC, MAF, ERMA and NZERN.

The workshop concluded that biodiversity conservation, biosecurity management and many other sectors require access to a single digital authoritative source of the names of organisms of importance to New Zealand. It was proposed that the system be a comprehensive database, the New Zealand Organisms Register, and that it should include all NZ organisms – terrestrial, freshwater and marine. An application for the development of this system was made to TFBIS in March 2006. This application was put together with input from participants in the original workshop including representatives from the agencies listed above. The TFBIS committee determined that more detail was required in particular areas before funding for the project as a whole could be approved. Funding was provided to complete a project scope (this document).

1.2 Terms of Reference

1.2.1 Objectives

The purpose of this Scope is to clearly understand the size, complexity, coverage and implementation options for the New Zealand Organisms Register project.

The objectives of this scope phase are to:

- Ensure the boundaries around the project, in terms of what is included and what is not, are clearly understood and agreed by stakeholders
- Address the TFBIS Steering Committee’s concerns as raised in the feedback on the project bid from the TFBIS General Manager
- Generate buy in and support for the project at a senior level within relevant provider, funder and end user agencies
- Ensure there is a thorough understanding amongst relevant stakeholders of the relationship between this project, international initiatives, other whole of government initiatives and individual agency’s plans

1.2.2 What’s Included

The Scope phase for the NZOR project included:

- Identifying the required streams of work within the project (e.g. software infrastructure, getting the baseline data, taxonomic gap analysis, presence/absence data)
- Evaluating the business needs of potential end users including MAF Biosecurity, ERMA, the Pest Management community, DOC, Regional Councils and MFish, and the research community
- Identifying the business needs and high level functional requirements of data providers including the motivating factors for their ongoing participation

- Determining the framework for distribution of responsibility for different taxa e.g. Landcare Research, NIWA, Te Papa
- Identifying the relationship between providers and end users in terms of the model for data stewardship and hosting, support and maintenance
- Defining options for a project governance framework including ongoing system management and prioritisation processes for allocation of resources and effort. This will include the role of a governance group in addressing the prioritization of available funds against taxa, and putting in place the validation/scrutiny framework for taxon groups.
- Identifying and evaluating potential technology solutions including existing local and international systems
- Understanding and documenting the fit of the project with the wider international and national context
- Identifying expected costs and timeline for project implementation
- Identifying at a high level the potential models and sources for ongoing funding for the system
- Writing a high level communications plan for the project for progress and achievements

Specifically excluded from the Scope phase were:

- Detailed functional requirements for the system
- The technical design of the system
- The work of the governance body including determining the scope of organisms covered and the prioritisation process for data migration to the system
- Identifying particular taxa, or sources of taxonomic expertise of high importance to biosecurity and biodiversity

1.3 Process and Participants

This scope was written by Dr Jerry Cooper and Dr Aaron Wilton from Landcare Research, and Julian Carver of Seradigm Limited (with support from Marianne Vignaux). Significant contributions were made to the document by Dr Pat Brownsey from Te Papa, and Dr Don Robertson and Dr Dennis Gordon from NIWA.

The project was guided by a steering group including Dr Elaine Wright, Dr Rod Hitchmough, Peter Delange, and Andrew Townsend (all from DOC), Dr Geoff Ridley (ERMA), Dr Ilse Breitweiser, Dr Aaron Wilton, and Dr Jerry Cooper (all from Landcare Research), Dr Joan Breach (MAF) and Dr Don Robertson (NIWA).

The preparation of the scope involved interviewing staff from DOC, MAF, ERMA, MFish, MoRST, Regional Councils, NGOs, CRIs, Museums and Universities. In addition New Zealand taxonomists were communicated with through the Systematics Association of New Zealand (SYSTANZ), and through contact lists of contributors to Species 2000 New Zealand.

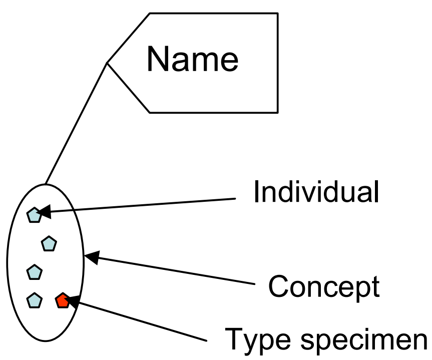
2 Business Case

2.1 It's Just a List of Names Right?

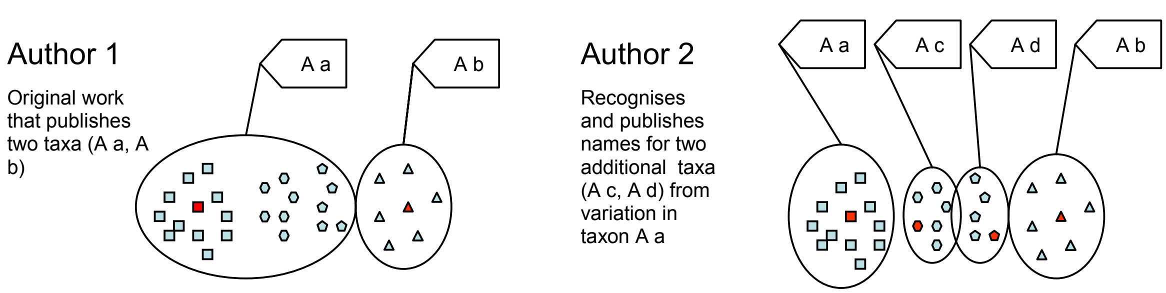
Why is NZOR a big project, when all we're doing is providing a list of the correct names for organisms that are relevant to New Zealand? Couldn't we just do that in a spreadsheet?

In order to answer this question, we'll need to understand something about how the science of taxonomy works. When a taxonomist sets about describing and naming a species, they look at many different individual organisms. They compare the way they look, where they grow or live, their DNA, and many other aspects. By doing this they form an idea about what the boundary around that species is. In other words, they develop a mental notion of what is included within that species, and what is not (and is therefore a different species). This 'idea' is called a taxon concept.

They then choose a single specimen that represents that concept and store it carefully within a well-managed collection. This specimen becomes the 'type specimen' for that species. The taxonomist then gives the species a name according to a set of formal rules, writes a description of what is included within that species boundary, and has the name and description published in a scientific journal. The relationship between these elements is shown in the figure below.

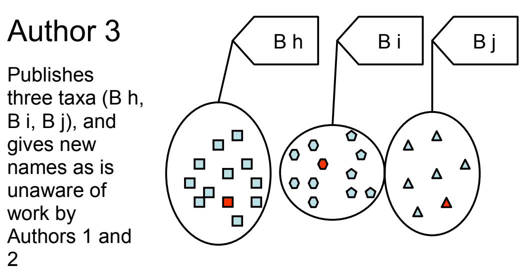


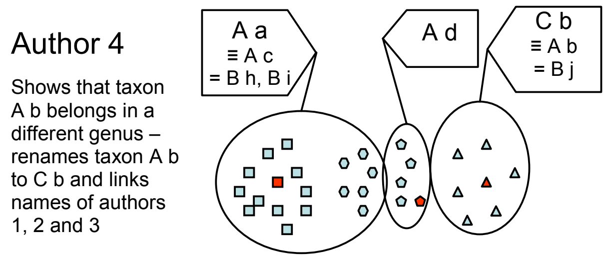
There are many taxonomists in the world, who spend much of their time at this task. Sometimes there is only one expert for a particular group of species, sometimes there are several experts, and they don't always agree. Over the decades new information is gathered, some groups of species are revised, and new species boundaries are drawn by successive generations of authors. This can be seen as follows:



NB In this figure the form 'A a' is a shorthand way of referring to the two parts of a scientific name, the genus and the species.

In this example Author 1 has described two taxon concepts that cover all the specimens they've observed. Some time in the future Author 2 makes a finer distinction, perhaps through looking more closely at the feeding behaviour, size, or genetics of the group of organisms, and decides there are two more species, giving a total of four. Unfortunately it doesn't stop there, as we can see below:

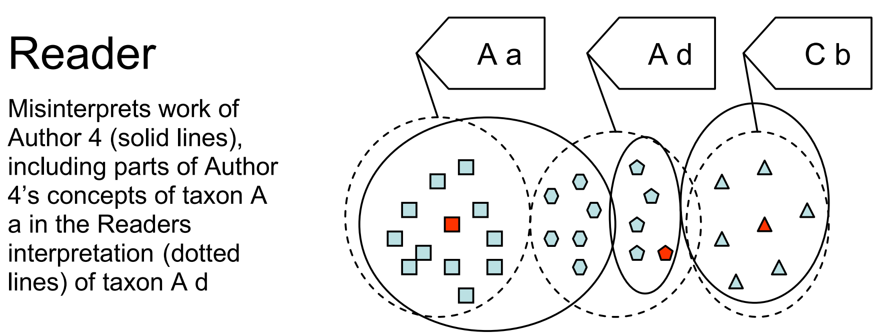




Author 3, without knowing that Authors 1 and 2 have been busy working on the same thing, comes up with not only a differing concept, but places them in a different genus and gives them a set of completely different names. Thankfully Author 4 comes along and sorts the whole mess out. In doing so he/she decides A b is actually in a different branch of the 'tree of life', and that Author 2 was wrong about A a and A c being different species. As a part of sorting all this out, he/she makes a list of the different published names these species have been called by. These other names are called 'synonyms'. NB in the diagram '≡' indicates the name following is based on the same type, and '=' indicates the name following is based on a different type.

When people involved in conservation or biosecurity use a name to refer to a species they want to preserve, or one they want to keep out of the country they are really referring to that 'idea' of a species, as described by a particular author. Where authors have disagreed or a concept has changed over time, the conservation or biosecurity managers need to be able to be clear about which author's definition is being used.

For example, a manager deciding on the conservation value of a region examines the available data. They find an old ecological survey publication that includes references to taxon A d. This is the only mention of A d in the region, so they double check with field staff and with a botanist who are able to confirm that A d doesn't occur in that region, and that the person conducting the survey (the Reader in the figure below) had misidentified some individuals of A a as A d. Because A d is rare within the region, making this discovery means that the manager does not have to implement an expensive and unnecessary management process.



In the future this common misapplication could be stored (and made available to NZOR), so that for other end users it is easier to find the correct answer and there are less risks of mistakes being made.

In order to effectively manage conservation and biosecurity therefore, it is crucial that all of this information is kept track of, is up to date, and is available to people who need to use it. An up to date list of current names is the tip of the iceberg. It can't be provided without all this other information supporting it. Managing the connections between all this information is why NZOR is somewhat complicated, and is more than just a list of names.

Please note, the definitions given above represent an initial overview of the levels of detail in NZOR. The terms used above are further defined and expanded in section 4.1. There are of course more complicated situations and levels of detail (it wouldn't be science if there weren't now would it?). These are expanded upon the Requirements and Systems Scope sections of the document. It should also be noted that not all data in NZOR will have full current synonymies, nor full historical synonymies. These data will be entered where they are available, and where there is a clear requirement for them. The system does however need to support this level of complexity to record and resolve taxonomic issues where they exist, and to provide an agreed list of preferred names.

2.2 Why We Need NZOR

All biodiversity information systems and almost all ecological/ biological research projects use the names of organisms as a fundamental identifier. Names provide the essential vocabulary by which we discover, index, and link information relating to biodiversity. Access to an authoritative list of names and their relationships to species (taxa) is therefore key to supporting research and information management across the conservation, biosecurity, and biotechnology sectors. Names of organisms underpin diverse areas of work including research, monitoring, surveys, management, planning and decision making in all these sectors.

Significant resources are made available to support standards and infrastructure for geographical information because spatial data are seen as key components of information. Taxonomic data are similarly important for supporting biodiversity information systems and similarly require sophisticated data management, standards, and curation by experts.

There is currently no single, definitive registry of the over 100,000 organism names relevant to New Zealand. Because of this many agencies currently each maintain their own lists of taxonomic names in isolation from each other, in different formats, and at different levels of completeness, depth and quality. This is often true across different systems within individual agencies as well. These lists may be updated on an ad hoc basis from information in scientific journals, consultation with internal and national and international experts, and through occasional exchange of data between data providers such as CRIs, and end user agencies such as MAF and DOC. The ad hoc and unstructured nature of these processes mean that lists are often out of date and inconsistent. For example there is currently conflicting information between the Landcare Research Fungi Names Database, the MAF Unwanted Organisms Register (UOR), and MAF Biosecurity Organisms Register for Imported Commodities (BORIC) data, and there are four major plant lists, two at DOC, one at Landcare Research, and one at MAF. The absence of a definitive source of taxonomic names means that resources are wasted through duplication of effort, there is increased expense to end-users in having to access multiple sources, and increased risk of confused decision making through different individuals from one agency accessing different sources.

Access to accurate, up to date, and agreed organism names and taxonomic data are critical for the threat classification, pest management and conservation work done by DOC. Not having access to these data increases the risk of inefficient use of resources in management programmes, and the risk of losing species through not adequately understanding or managing our biodiversity.

There is currently no definitive list of which organisms are present in New Zealand, and which are absent. This has significant implications for biosecurity management. It increases the cost of compliance and risk assessment for plant breeders and importers and may unnecessarily restrict economic innovation and financial returns for those industries. It also increases the risk to the export sector that misapplied names will imply that organisms are present in New Zealand when they are not, thus adding to the cost and/or restricting importation of New Zealand goods by other countries. There is currently a lack of definitive presence/absence data, and a lack of integration between biosecurity information systems and national and international sources of taxonomic data. This means that through changes in taxonomic concepts, misapplications of names, or misinterpretation of conflicting data, the risk of economically and environmentally damaging incursions of foreign organisms occurring is higher than it need be given the technology available today.

The lack of a definitive New Zealand registry of organism names also makes it harder for New Zealand to achieve its international reporting obligations. For example in the IUCN (World Conservation Union) Red Data Book the Chatham Islands pigeon is not listed as threatened because it was regarded as a subspecies rather than a species due to the use of outdated taxonomic authorities. Having a central registry of organism names would make it easier for New Zealand to contribute data to the Global Biodiversity Information Facility (GBIF) and could, in so doing, improve the quality of research into changes in global biodiversity.

2.3 Vision

The long term vision for NZOR is:

"An accurate, authoritative, comprehensive and continuously updated catalogue of taxonomic names of all New Zealand biota and other taxa of importance to New Zealand. This catalogue will be electronically available through one or more portals, and will be directly integrated into biodiversity and biosecurity systems used by central government ministries, departments, and agencies, local government, research institutes, NGOs and the wider community. The catalogue will be maintained by experts in each taxonomic group, and quality will be controlled through an agreed governance process.

The catalogue will be based on internationally agreed standards and will include organism names and synonymies, basic presence/absence data (whether the organism is present in New Zealand or not), and where possible alternate and historical synonymies. It will link to information from other sources on aspects such as threats, ecology, distribution, use, management status, published material, keys for identification, and all collections, observation and survey data. The lists will contain all names irrespective of taxonomic status and will provide rankings of reliability including transparent access to provenance of data, and the ability to filter by quality and currency."

Achieving this vision as a country would let us avoid duplication of taxonomic effort and conflicting interpretations. This national asset would reduce costs and the potential for errors inherent in each organisation maintaining their own lists. It would reduce 'log jams' at the border and the risks of new invasive species entering the country. It would save agencies time in finding authoritative names and remove uncertainty and guesswork from critical decisions.

As a result, it would be possible to share biodiversity and biosecurity information across organisations and national borders without ambiguity. Responsibility for the maintenance of lists of taxonomic names could be distributed cost-effectively whilst improving quality through a robust process for experts to resolve issues. Achieving this vision would enable conformance with international standards, enable more accurate and timely data exchange internationally, and improve New Zealand's ability to fulfil its international treaty obligations. It would encourage the maintenance and development of New Zealand's taxonomic capability, foster more collaboration both nationally and internationally, and maximise the return from investment in taxonomic research by enabling optimal biodiversity and biosecurity outcomes.

This vision builds on significant past work and successes in this area in New Zealand including the Species 2000 New Zealand project, the Landcare Research developed Dynamic Taxonomic Framework and Plant Names database, and the global checklist of Compositae project. The vision is however likely to take a number of years to achieve. The system will not be able to provide or link to all the data mentioned in the vision within the first year. Putting a technology platform and associated process in place now will however make this vision achievable in the long term.

2.4 The Role of TFBIS in NZOR

The TFBIS mission is to "*Support seamless accessibility of essential biodiversity data and information to achieve the goals of the NZ Biodiversity Strategy 2000*". TFBIS has made significant contributions to the state of biodiversity data and information, enabling the digitisation of many thousands of collection and observation records and of many books and journal articles, supporting the information management needs of volunteer-based NGOs, and contributing to the development of a number of nationally important systems and tools such as the National Vegetation Survey Databank (NVS) and the Freshwater Biodiversity Information System (FBIS). TFBIS has helped address a number of gaps in digitally available data and tools that would not have been filled by other funding mechanisms. In so doing TFBIS has also fostered interagency collaboration that would have not otherwise occurred.

In order to meet current and future biodiversity challenges it will be essential to harness and make best use of all relevant data and information. This can only happen in a 'federated' way, with data being shared across research providers, management and policy agencies in central and local government, NGOs and where appropriate the community at large. To achieve this some 'infrastructural' systems will be necessary to connect this diverse information together and help users navigate it. In its recent strategy draft TFBIS has signalled its intention to focus efforts on building this infrastructure. Given the size of the task TFBIS will not be able to build the entire required bioinformatics infrastructure itself. Because of the relative flexibility and autonomy of the fund however, TFBIS can take a significant leadership role in this area, catalysing change and encouraging others to follow.

In 2005 TFBIS realised that there were some challenges with taxonomic names and associated databases and commissioned a workshop to identify options for resolving issues and to recommend a way forward. This scope is a direct result of that workshop, as is the interagency collaboration it fostered between Landcare Research, NIWA, Te Papa, DOC, MAF and ERMA.

The NZOR system as described herein will have benefit not only to terrestrial and freshwater biodiversity research and management, but to the marine domain, and to biosecurity in general. These two areas are outside of the remit of TFBIS, and it may be justifiable for TFBIS to eschew contribution to these sectors. Constructing a system to manage terrestrial and freshwater names will however not require any additional technical effort to be able to support marine organism names, or to be useful for biosecurity purposes. Initial population of the system with currently available electronic data from Landcare Research, NIWA and Te Papa, and the data from the Species 2000 New Zealand lists would cost close to the same amount whether marine data was included or not. Increasingly end user agencies such as DOC, MFish and Regional Councils are addressing management of marine biodiversity, and there are significant efficiencies to be gained by addressing all organisms in one system.

Biodiversity and biosecurity are human created distinctions. In reality they overlap and have impact on each other. Biosecurity incursions can seriously threaten indigenous biodiversity, agricultural and aquacultural industries and human health. Healthy, diverse ecosystems are a powerful defence against destructive foreign organisms. Cooperation between agencies involved in biodiversity and biosecurity is very important. While the draft Biosecurity Science Strategy signals the need to improve access to taxonomic information through the use of information systems, that sector itself is unlikely to build NZOR in the immediate future. TFBIS is uniquely placed to take a leadership role here. By funding the initial establishment of the system, and the population of it with baseline terrestrial, freshwater and marine organism names and taxonomic data TFBIS will have created infrastructure and processes that support terrestrial and freshwater biodiversity management, and will also benefit marine and biosecurity concerns. The ongoing funding for the system and addition of extra data in the future is then likely to be picked up by a combination of MoRST science backbone initiatives, FRST research programmes, and end user agencies. By funding the establishment of NZOR, TFBIS will have not only fostered, but actually followed through on real interagency collaboration, and will have contributed an important component to New Zealand’s environmental information systems landscape.

3 Requirements

This section describes the business needs of potential end users including MAF Biosecurity, ERMA, the Pest Management community, DOC, Regional Councils and MFish, and the research community. It also identifies the business needs and high-level functional requirements of data providers.

3.1 Overview

In a formal sense, users of the system can be split into two major groups, providers of data, and consumers of data. Providers are primarily taxonomists working on particular taxonomic groups, or in the case of larger research organisations the informatics professionals that collect and manage the nomenclatural and taxonomic data in specialist systems. Consumers are primarily end user agencies who require access to correct organism names for their work in biosecurity and biodiversity policy and management. In reality many organisations are both providers and consumers of data. Landcare Research and NIWA for example both provide new taxonomic data, but they also have ecologists that use those names in their research and in systems in which they store their data. DOC is a large consumer of names data for conservation management purposes, (as is the Ministry of Fisheries). DOC also has taxonomists who are national experts on groups such as skinks and geckos. Even MAF who would almost exclusively be consumers of data may provide lists of non-present, undesirable taxa.

With that range of use in mind, the following is a high level overview of the types of business needs users will have for NZOR:

Consumers will need:

1. A list of correct, current names
2. Names, synonyms and taxonomic rank data for taxon concepts
3. Biostatus data (specifically whether an organism is present in New Zealand and its origin)
4. Publication information for nomenclature or synonymy
5. Information about the reliability and/or completeness of the data in NZOR
6. Related information such as publications, websites, distributional data, threat status, management status. NB these data are likely to come from other systems and NZOR will link to, rather than store such data.

For each of the above uses, there will be:

a) Use by an individual through a web site straight into NZOR

b) Use by a system (e.g. the PBI at MAF) through a web service and local cache to maintain a list of names (and full synonymies if required) in their application

c) Use by automated tools (for example to check an organisation's list of species names against information in NZOR)

Providers will need:

1. The ability to load nomenclatural and taxonomic datasets into NZOR
2. The ability to change data in NZOR including fixing errors, merging or splitting taxa, adding synonyms
3. To be able to see data usage statistics for the system

There will also be some special requirements that administrators of the system will have including viewing usage statistics on parts of the system, and synchronising data with international networks such as GBIF.

3.2 Who Will Use NZOR?

As described above, the organisations that will use NZOR fall on a spectrum from purely providers of data, to purely end users. A reasonable proportion of these organisations will perform both roles. This section provides short summaries of the organisations that may use NZOR, their current activities and systems of relevance, and the requirements they may have for NZOR.

3.2.1 Species 2000 New Zealand

Species 2000 New Zealand is related to, but distinct from, the global project known as Species 2000. Species 2000 New Zealand is a project led by Dr Dennis Gordon from NIWA, and involving 221 contributing authors. One of the major outputs of this project is a set of published volumes of kingdom/phylum-by-phylum reviews of all taxa, with checklists of all known species. This contains nomenclatural data for all known native and naturalised New Zealand organisms. It does not include synonyms.

Species 2000 is expected to be able to contribute a significant layer of data to NZOR on a once off basis. This will be particularly useful to fill gaps where taxonomic groups are not represented in easily accessible electronic databases (like those held by the major providers Landcare Research, NIWA and Te Papa).

3.2.2 Te Papa

Te Papa has significant collections of many groups of the New Zealand biota, estimated at over 1,300,000 individual collection lots. Of these, over 350,000 are now databased on KE Emu. Many of the collections complement the Nationally Significant Collections held at Landcare and NIWA and include groups not held at those institutions. The major collections are seed plants and ferns, bryophytes and lichens, marine algae, terrestrial invertebrates, marine invertebrates, fishes, birds, fossil vertebrates, reptiles and marine mammals.

Associated curatorial and taxonomic expertise extends in particular to ferns, bryophytes, lichens, marine algae, lice, spiders, molluscs, crustacea, fishes, seabirds, fossil vertebrates and marine mammals. Te Papa staff maintain taxonomic hierarchies of names for all the above groups in the KE Emu database. In some cases, such as ferns, the list of currently accepted names and synonyms has been incorporated into the Landcare Plant Names Database, or, in the case of bird names, to the OSNZ Checklist. In the case of molluscs, Te Papa staff contribute to the website list of currently accepted New Zealand molluscs. In other groups, such as fishes and marine algae, the database hierarchies are the principal list of species names for New Zealand.

Te Papa is therefore expected to be a significant provider of nomenclatural and taxonomic data and expertise to NZOR.

3.2.3 NIWA

NIWA is NZ's main research provider for all aspects of aquatic science, including research on biodiversity, biosecurity, and a wide range of ecological and population modelling studies across all of NZ's freshwater environments and the 200 nautical mile Exclusive Economic Zone. NIWA's responsibilities include maintenance of the Nationally Significant Marine Invertebrate Collection and database and the Nationally Significant Freshwater Fish Database. These are supported by significant taxonomic expertise in many of the main taxa of aquatic plants and animals. In addition NIWA maintains a number of other significant aquatic databases, including databases managed for the Ministry of Fisheries and MAF Biosecurity NZ. All of these and all research projects involving aquatic species are dependent on knowledge of their taxonomy/biosystematics and a definitive source of aquatic species names. NIWA's marine biological projects are underpinned by the FRST OBI programme on "Effective management of marine biodiversity and biosecurity". The NZOR concept if adopted will be, from NIWA's perspective, a huge step forward in the management of NZ's biodiversity and biosecurity.

NIWA has a long track record of managing all types of aquatic data. In the biodiversity context this includes the development of the web-map based Freshwater Biodata Information System (FBIS) with TFBIS funds. More recently as a part of the Global Biodiversity Information Facility (GBIF) programmes and the Ocean Biogeographic Information System (OBIS - the marine node for GBIF), NIWA has built and is now populating a regional OBIS Node for NZ and the South-western Pacific, covering the oceans from Antarctica to Fiji. This is serving around half a million species locality records to GBIF via OBIS.

Most of the national taxonomic data is managed by New Zealand's largest teams of systematics experts within NIWA, Landcare Research and Te Papa who receive funding under the FRST Defining NZ Biota Target Outcome ($6 million per annum terrestrial ecosystems and $1.4 million per annum for marine ecosystems).

3.2.4 Landcare Research

Landcare Research is the national focus of systematic research relating to the terrestrial biota under the Defining NZ Biota OBI (see above). Underpinning this research are four of the FRST Nationally Significant Collections and Databases managed by Landcare Research. These are the Allan Herbarium (CHR), the National Fungal Herbarium (PDD), the New Zealand Arthropod Collection (NZAC), and the International Collection of Microorganisms from Plants (ICMP). The taxonomic data associated with these collections are actively managed and made publicly available through web portals. These data are also essential for the management of the collections and in supporting our own research across systematics, ecology, conservation biology and biosecurity.

The web portal associated with PDD and ICMP is known as NZFUNGI. It contains comprehensive, actively maintained information on fungi, plant associated bacteria and viruses. This information includes appropriate synonymy and judgments on presence/absence. NZFUNGI has become valuable resource for MAF BNZ and ERMA in a biosecurity context.

Similarly the data associated with the Allan Herbarium is made available through the New Zealand Plants website and associated web service. The website provides taxonomic and nomenclatural information on seed plants, ferns, lichens, mosses, liverworts and freshwater algae. The taxonomic information on ferns and fern allies is managed in collaboration with Te Papa, and freshwater algae with the University of Canterbury. The service receives significant positive feedback and contributions from key end user agencies such as DOC. The web service is being used by EBOP, Horizon, QEII and NZERN to maintain local caches, and data extracts are used by DOC and the New Zealand Biodiversity Recording Network.

Landcare Research also has a potential role as a provider of technology supporting NZOR. Landcare Research has made substantial investment in developing data management systems to support taxonomic information and is prepared to make this suite of technology available to the project. These contributions include the Dynamic Taxonomic Framework (DTF) data management application developed in Microsoft DotNet, the back-end SQL-Server database structures and business process stored procedures, a preliminary web-service tool to facilitate remote end-user caching and updating and a basic web portal providing public access to the information. These have been developed over a number of years to support the Nationally Significant Databases maintained by Landcare Research and the work funded through FRST and Landcare Research funding. The tools have been developed in the context of developing new standards and with a wide engagement with the international community.

In addition Landcare Research is engaged in a GBIF-funded project in partnership with leading systematic research organisations around the world to develop a global checklist of Compositae (daisies and allies). This project has developed useful tools for integrating taxon concept data from multiple sources. (see relationships). Landcare Research is also currently funded under a joint TDWG/GBIF project to further develop Life Science Identifiers as an enabling technology for sharing taxon concept data. Landcare Research's Jerry Cooper has expertise in the area of taxonomic data management by helping to develop the international data exchange standard (The Taxonomic Database Working group's Taxon Concept Schema- TDWG-TCS), and is a member of both the Species 2000 Global Committee and the GBIF Electronic Catalogue of Names of Organisms (ECAT) Committee.

3.2.5 Universities and Regional Museums

Taxonomists working in Museums (other than Te Papa) and Universities are likely to provide some taxonomic data and some expertise in particular groups to NZOR. Of the 133 New Zealand authors for Species 2000 NZ, 22 worked in Universities, and 13 from museums (the large majority of whom were from Te Papa). Some of the data these specialists could provide is already in the large collection databases held by Landcare Research, NIWA and Te Papa and is likely to make its way to NZOR through that route. Nomenclatural data from these specialists would also be available from the Species 2000 New Zealand lists. Once these initial tranches of data were entered into the system, specialists from Universities and regional museums could be called in to help fill gaps. Staff from these organisations may also be a part of the data governance process for NZOR.

Universities and regional museums are also potential end users for NZOR. Researchers and collections managers from those organisations may choose to use the NZOR web site, download species lists for smaller research databases, and download more complete lists of names and taxonomic data for larger collections.

3.2.6 DOC

The Department of Conservation is a significant user of taxonomic names. Names are the basic currency for all DOC biodiversity data. Having accurate, up to date, and agreed names are critical for threat classification. They are also essential in determining the conservation work the organisation does. This is true not just for species DOC is trying to protect, but also for pest species.

DOC relies heavily on tag names as they often identify organisms for management that have not yet been scientifically described. Managing the relationship between tag names and scientific names once organisms have been described is important for the continuity of management, planning, and comparing distributional data across regions, and over time.

DOC has a number of databases and information systems that use taxonomic names. These include Bioweb, PestLink, Threat Classification Lists, and many smaller databases. The Natural Heritage Management System DOC is planning has taken the potential availability of NZOR into account in its requirements definition process.

DOC will also be a provider of names data for a small number of species, including herpetofauna.

3.2.7 ERMA

The Environmental Risk Management Authority (ERMA New Zealand) makes decisions on applications to introduce new organisms into the country. As such it is very important that they know, or can find out, what is present in New Zealand. They must balance the risks of letting organisms in that could cause environmental damage (under regulation) and the economic costs to the country of overly restricting the growth of new horticultural and biotechnology industries (over regulation).

Where applicants want to avoid an expensive risk assessment they must prove that the organism they wish to import is already present in New Zealand. The burden is on the applicant to provide evidence, and they do this through citing published papers, trade catalogues, unpublished reports, and expert witnesses. ERMA must verify that the claims applicants are making are true, are accurate, and are not based on misapplications, or names that have since changed. This is especially the case in groups where there has been a lot of taxonomic change, or there is high risk. Having access to a data on which organisms are present, and either by implication or specifically listed which are absent is essential in this process, as is being able to navigate the changes in taxonomy and synonyms over time.

Currently applicants and ERMA rely heavily on the Plant Biosecurity Index, maintained by MAF. ERMA has concerns about the PBI, in particular that it is incomplete, and there is no easy way for names to be kept up to date. There is also no data stored as to what evidence was used to justify including organisms on the PBI. The concern over incompleteness is shared by organisations such as the Plant Breeders Association as it makes over regulation more likely, and the process of application to ERMA more time consuming.

ERMA maintains a record of the decisions it makes on its Oasis database, also accessible through the ERMA web site. This records scientific names, but does not connect into taxon concepts or synonyms. As such ERMA has concerns about applicants and ERMA staff making incorrect judgements, based simply on the names used in past decisions, where the taxonomy and correct names may have changed. ERMA also maintains a spreadsheet to record the non-statutory advice it gives. This too does not have any way of connecting to synonyms and name changes over time.

ERMA is therefore supportive of NZOR in that it will help maintain the accuracy of the PBI, may provide more robust and extensive data to assist its staff in assessing applications, and will reduce cost and effort for applicants.

3.2.8 MAF

To minimise biosecurity risks MAF needs to keep organism names up to date, and it needs to understand the relationships between synonyms. This is currently difficult when it maintains multiple overlapping lists internally, the sources of those names externally come from multiple different providers, and the organisation does not have in-house access to all the taxonomic expertise it requires.

A range of people from MAF Biosecurity including staff in Business and Planning, Investigations and Diagnostics, Risk Analysis, Policy, and Strategic Science have expressed support for the NZOR project. The need for information systems to link databases and other science knowledge including taxonomic information is detailed in the draft Biosecurity Science Strategy. Strategic Science staff see this as taking a ‘federated’ approach where many different systems are connected together with effective standards and access agreements.

MAF has a number of internal systems and lists for biosecurity standards, risk analysis and expert certification that could make use of NZOR. Important ones include:

- **Biosecurity Risk Analysis Database –** lists organisms that could pose a threat to New Zealand. Includes relationships between pest, host plant (often commodities) and the plant parts it attacks, as well as the countries or regions of origin for the pest species. 34,000 taxa.
- **Plant Biosecurity Index** – lists which plant species are already present (and accepted) in the country and may therefore be imported. 27,000 taxa.
- **Unwanted Organisms Register** – lists organisms that have been officially determined unwanted by government agencies, and those declined importation by ERMA. 15,000 taxa, some overlap with BRAD.
- **PPIN -** provides a list of which plant pest species are, or have been established (and in some cases eradicated) in New Zealand. Approximately 30,000 taxa.
- **BORIC –** A list that provides border staff with information on how to treat pests found on specific commodities

There are also a number of other systems and lists that contain organism names, including Commodity Pest Lists, Country Freedom Lists, Global Invasive Species List, and STARS (A specimen tracking and reporting system). A short summary of the most relevant MAF Biosecurity information systems can be found in Appendix 3.

As part of the MAF Information Systems Strategy there is an imperative to consolidate organism and names data across the many systems MAF has. NZOR is seen as a useful tool in achieving this, as it will provide up to date data, and will save MAF from having to construct and maintain such a system itself. As such MAF information systems staff have also expressed support for the project. It seems likely that MAF would use a synchronised local cache of NZOR data, and connect their individual systems to that cache, rather than to NZOR directly.

3.2.9 Ministry of Fisheries

As described in the Marine section (4.3.3) above, The Ministry of Fisheries is potential end user of NZOR. While MFish is reasonably well served for taxonomic/nomenclatural data for fish species, having well managed organism names and taxonomies for other marine species could improve the quality of the science and advice given to MFish by its research providers on the environmental effects of fishing on ecosystems, including impacts on biodiversity. Senior MFish science staff have expressed support for the project.

3.2.10 Local Government

Local government agencies, in particular Regional Councils, manage biodiversity within their jurisdictions. This includes identifying and conserving threatened species, and managing pests. All of these agencies, at some level of sophistication or another, have information systems to support this. Having access to up to date organism names is important in doing this work.

A number of Regional Councils are already using external lists such as the Landcare Research Plant Names Database. Environment Bay of Plenty, and Horizons Regional Council have connected this directly into their own systems using the Plant Names Web Service. No such facility currently exists for invertebrates or for freshwater or marine flora and fauna however.

When it is available Local Government agencies are likely to use NZOR to look up taxa, import lists of names into their systems, and/or directly connect their systems to NZOR as consumers of names.

3.2.11 Non Governmental Organisations (NGOs)

There are a number of NGOs that will have some interest in NZOR. The Ornithological Society of New Zealand (OSNZ) are likely to be a provider of nomenclatural and taxonomic data for birds. Community organisations such as the New Zealand Plant Conservation Network (NZPCN), and the New Zealand Ecological Restoration Network (NZERN) are likely to be end users of NZOR data. NZERN for example already connects its systems to the Landcare Research Plant Names Web Service.

3.3 Use Cases

A set of example use cases have been developed to define provider, end user, and system requirements. Use cases describe a particular user goal (or trigger), a set of steps explaining the interactions between the user and the system (or in some cases between two systems), and an example of this occurring. The full use cases are included in Appendix 4 at the end of this document. These are not conclusive or definitive for NZOR, but rather have served to shape and refine the understanding of the scope of the system. The use cases will need to be refined further as a part of the formal requirements definition and specification of the system.

The following is a list of the use case titles:

Use Cases 1 Data Loading and maintenance

- Use Case 1.1 Registering as a data provider
- Use Case 1.2 Loading a set of data
- Use Case 1.3 Reporting an error
- Use Case 1.4 Updating Data

Use Cases 2 Data use

- Use Cases 2.1 Searching by taxon
  - Use Case 2.1.3 Taxon not found in New Zealand
  - Use Case 2.1.5 Taxon found under different name
  - Use Case 2.1.7 User uncertain of proper spelling of name
  - Use Case 2.1.8 Several taxa match
  - Use Case 2.1.10 Disputed or unresolved taxon concept
- Use case 2.2 Detailed information
  - Use Case 2.2.1 User wants reference information
  - Use Case 2.2.2 User wants biostatus information
  - Use Case 2.2.4 User wants a history of name changes
- Use Case 2.4 User wants to know the confidence level of data for a record in NZOR
- Use Case 2.5 User wants to print out a list of all taxa that meet some criterion
- Use Case 2.6 User wants a count of all taxa that meet some criterion
- Use Case 2.7 Parent relationships
- Use Case 2.8 Child relationships
- Use Case 2.9 Search on literature

Use Cases 3 Tools

- Use Case 3.1 Login with password
- Use Case 3.2 Information via web services
- Use Case 3.3 Regular updates
- Use Case 3.4 Comparing lists of taxa
- Use Case 3.5 Usage statistics
- Use Case 3.6 Data exchange of concepts

4 System Scope

4.1 Definitions

In order to define what is in scope for NZOR, and what is not, a number of definitions are necessary. Readers (especially experts) may not agree completely with some of these definitions or terms. Absolute consensus on these definitions is not necessary, however it is necessary to agree on these terms and definitions for the purposes of discussing them in this scope. The definitions are somewhat technical in nature, however they are required to fully convey the different types of information to be managed in the system. The definitions are:

Scientific Name – a formal scientific name conforming to the International Codes of Nomenclature (for plants (including fungi), animals, viruses, bacteria and cultivated plants). Scientific names typically take the form '*Genus species* Author (publication date)'.

Reference – an electronic record of the publication details

Publication – a peer reviewed scientific journal, or book

Type – is a single collection (i.e. a single collected specimen) upon which a name is formally established. Typification is the key component of the nomenclatural codes that provides a stabilization of the application of names to taxa. Type specimens are of particular importance because they provide the essential reference points in any taxonomic system of nomenclature. In reality the Type specimen is the only unambiguous use of a name. All other uses of the same name make the assumption of the equivalence of taxon concepts.

Taxon – (plural taxa) a classificatory unit of any rank within the taxonomic hierarchies of organisms. Thus, Animalia, Chordata, Vertebrata, Carnivora, etc. are each taxa, as is the great Siberian Tiger, *Felis tigris*, figuratively lying in the shade, at the bottom of the 'tree'. It is important to be clear that a taxon is a subjective mental concept about a group of organisms, not an actual specimen or observation.

Nomenclature – the objective method of assigning a 'correct name' to a taxon from a list of alternative names. It operates within the legalistic framework of the Codes of Nomenclature. It is dependent on the declaration of type specimens.

Taxonomy – this is the process of assigning organisms to taxonomic groups based on shared/differing characters, and the arrangement of taxonomic groups into a hierarchical classification. This is typically a subjective process resulting in a circumscribed taxon concept (a taxonomic opinion).

Rank – a position within the divisions in the Linnean hierarchy (Domain, Kingdom, Phylum/Division, Class, Order, Family, Tribe, Genus, Species, Subspecies, Variety, Forma)

Vernacular Names – informal names in any language that are used to refer to taxa, a taxon, or a part or developmental stage of a taxon or taxa. The use, format, language and application of these names are not governed by a set of rules (e.g. the Codes of Nomenclature for scientific names). As a consequence a single vernacular name is often applied to a different taxa depending on the geographic regions, societal groups (even in the same region), and language applying the name. Similarly, a single taxon may have several vernacular names that vary according to the geographic region, the language, the societal group, and the stage or part of the organism being referred to. NB vernacular names are frequently referred to as 'common names'. However, this term should be avoided because it is unlikely that a common, single vernacular name will be used across all geographic regions, societal groups and languages to refer to the same taxon. Vernacular names need be considered to consist of two components (1) the name, including information on its geographic and linguistic origins, translation and transliteration, (2) the use of the vernacular name - i.e., the taxon concept to which the name has been applied within a particular context of language, geography etc.

Synonyms – a taxonomist's opinion of the different scientific names used to refer to a single taxon, together with the 'correct name' arrived at by applying the rules of nomenclature. There are two major kinds of synonyms:

> Those that refer to the same type specimen – in botany these are referred to as 'homotypic synonyms', and in zoology they are referred to as 'objective synonyms'. For example the Linnaean name Pinus abies L. has the same type as Picea abies (L.) H.Karst. When the latter is taken to be the correct name (there is almost complete consensus on that), Pinus abies is a homotypic synonym of Picea abies.
>
> Those that refer to different type specimens – these occur where a taxonomist believes the two type specimens are really representatives of the same taxon. In botany these are referred to as 'heterotypic synonyms', and in zoology they are referred to as 'subjective synonyms'. For example some botanists split the dandelion into many, quite restricted species. The name of each such species has its own type. When the dandelion is regarded as including all those small species, the names of all those species are heterotypic synonyms of Taraxacum officinale F.H.Wigg.

A Synonymy – A set of synonyms including the correct name. This may include both heterotypic and homotypic synonyms.

Taxon concept – an attempt to provide a communicable definition of the mental concept of a taxon. A taxon concept is the subjectively asserted relationship between a name or group of names (a synonymy) and that taxon, made by a taxonomist. It is formally defined by a published 'circumscription'. When many taxonomists use the term 'synonymy' they are referring to what we define here as 'taxon concept' i.e. the list of synonyms including the circumscription.

Circumscription – the traditional, and ungoverned mechanism by which a taxonomist expresses their definition of a taxon concept in a published work. It is characterised by a Correct Name, a list of Synonyms and a placement in a Taxonomic Hierarchy. For example, the species *Raoulia australis* is a child of the genus *Raoulia*, which in turn is a child of the family Asteraceae, and *Raoulia lutescens* is a synonym of *Raoulia australis.* In addition, and perhaps more importantly, a circumscription usually references a number of specimens that have been examined, together with a description. The taxonomist's concept is encapsulated by their description, which combines the envelope of characters covering the specimens examined (i.e. the observable characteristics such as morphological, chemical, molecular/genetic traits that are used as evidence to delineate the taxon), together with their understanding of the envelope of characters covered by the set of synonyms (and specifically the characters associated with the Type collections of the set of synonyms).

Correct Name – the name used by a taxonomist to refer to a taxon concept. It is arrived at by applying the rules of nomenclature to the set of names the taxonomist considers to be synonyms, in their taxonomic opinion.

Concept Correlation – a representation of the relationship between two taxon concepts. For example, a concept correlation may record the difference in taxonomic opinion between the circumscriptions in two taxonomic treatments (published papers or books such as monographs, revisions etc) by indicating that a given concept (a) is equivalent to, contained within, overlaps, is excluded from (etc) a given concept (b) in a second paper. A concept correlation will include reference to: concept (a), concept (b), the relationship, the name of the person asserting the correlation, and the publication the correlation is published in, or if not explicitly published, the publication the correlation assertion is based on.

Active concept – Because taxonomy is subjective there can be very closely related taxon concepts (from different taxonomic treatments), which are likely to be referring to the same taxon. When this occurs a decision on which one will be used is required, in particular so that the single 'correct name' for that taxon can be provided. NZOR will follow an agreed process to determine which of the competing taxon concepts is the 'active concept'. It should be noted that the term 'preferred name' is often used by taxonomists to mean the active concept, when pedantically, it means the correct name of the active concept.

Usage concept – This is typically made as an identification by a non taxonomist (e.g. an ecologist, a DOC field worker, or a collections staff member) and is a subjective interpretation of a taxon, which results in the application of a name (typically to an observation or specimen record). In the case of observations the formal definition of the non-circumscribed concept is less easy to discern as there is usually 'little to go on' other than the personal reputation of the recorder amongst peers. There may thus be differing interpretations of the taxon being referred to. This results in 'third party' assertions (see below) which can generally only be correlated through broad categories of 'overlapping', 'contained within' etc (similar to a concept correlation, but in this case at the non expert level). In the case of specimens in a collection it may be a taxonomist who makes the identification, there is typically more rigour in application of usage concepts, and more data is recorded to allow the identification to be validated. A usage concept will usually include reference to: person (i.e. observer), place, date, object/observation, name (invoking a concept and a circumscription), and publication. The link to a concept and circumscription does not always happen explicitly in observations, it does occur more commonly with specimens in collections.

Usage assertion – a third party interpretation of the relationship between usage concepts. Usage assertions can apply to observation, survey and collection data, and may be restricted by specific geographic regions and time periods. There are some cases where usage assertions can be considered safely, and some cases where they should be used carefully. This is essentially a quality measure where the assertions are based on who made them and how expert they are. A usage assertion will include reference to: A usage concept (including possibly a specific geographic area and temporal range), the correct concept (and therefore name), and who made the concept assertion. For example, a researcher using early records of *Gentianella divisa* from north of Lees Valley could use the revision by Glenny (2004) to apply the usage assertion that they represent *G. luteoalba* when from the Lookout and Hope Ranges, *G. magnifica* when the records were collected from the Amuri and Rachel Ranges, but would need to find more information for observations from other places north of the Lees Valley.

Misapplications – a special case of a concept correlation, or usage assertion, where the misapplication of a name is well understood, but needs to be recorded to prevent poor decisions being made based on that misapplication. For example, the name *Libertia pulchella* was used in New Zealand until 2002 for the New Zealand taxon now named *L. micrantha*, while *L. pulchella* correctly refers to an Australian taxon that does not occur in the wild NZ flora. *Libertia pulchella* as used in New Zealand prior to 2002 is a misapplication, and it is therefore possible to make the usage assertion that all non-cultivated records of *L. pulchella* collected in New Zealand prior to 2002 may be interpreted as *L. micrantha*. These will be recorded in NZOR while many other usage assertions will not.

Tag names – these occur when something considered a new taxon is found, but it has not been formally named using the code of nomenclature and the publishing process. The informal name given to it is called a 'tag name', e.g. the Remarkables Carabid Beetle. Tag names often occur in ecological studies where taxa are compared in ecosystems without needing to be formally described, they are also used relatively extensively by DOC. Often a tag name may be used for some time before resources become available to complete the proper nomenclatural and taxonomic work on that species. Once that work is done, the tag name effectively becomes a synonym (albeit nomenclaturally invalid) of that taxon concept.

Specimen – a part, whole or group of individual organisms that have been collected and permanently preserved as part of a natural history collection.

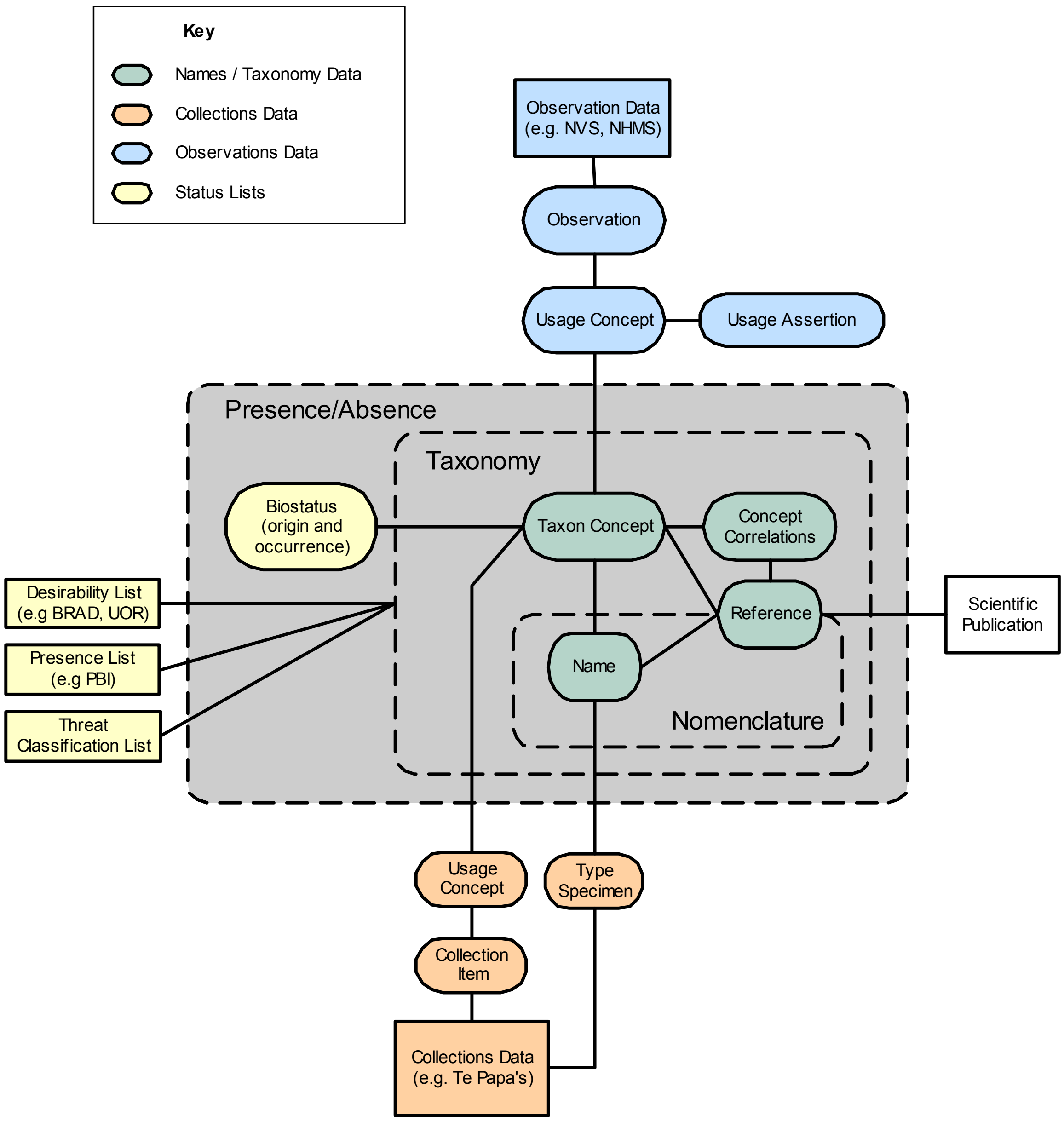
Observation – a record of the occurrence of an organism(s) at a given time and place, as seen and identified by the observer.

Biostatus – a number of measures of the status of a taxon. In the case of NZOR biostatus is particularly focused on that taxon's origin (e.g. indigenous, endemic, foreign, cryptogenic) and its occurrence (i.e. its presence or absence) within the specified region. Regions may be geographic or political, with political regions likely to be of most interest to users of NZOR. To be useful for NZOR's purposes biostatus must be verifiable by a scientific publication ideally by reference to a documented specimen in a publicly accessible collection. Some other aspects could possibly be considered part of biostatus, such as threat classification, management status, control status as a pest. These however are outside of scope for NZOR as they are most appropriately the responsibility of other systems and information stores.

LSID – Life Science Identifiers are a uniform way to name and locate pieces of information on the web. An LSID is a unique identifier for some data, and the LSID protocol specifies a standard way to locate the data (as well as a standard way of describing that data). An LSID is represented as a Uniform Resource Name (URN) with the following format.
URN:LSID:<Authority>:<Namespace>:<ObjectID>:<Version>

4.2 Scope of NZOR

This diagram represents the scope of NZOR. Entities and attributes inside the grey dotted line box are in scope and will be stored in NZOR. Those outside the grey box will be kept in other systems held within provider or end user organisations.



NZOR will contain New Zealand relevant organism names, references to publications, names grouped together as synonymies within taxon concepts, tag names, vernacular names (including Maori names), concept correlations (including common misapplications), and biostatus (origin and occurrence). While tag names and vernacular names are not strictly scientific, they are in scope for reasons discussed in section 4.3.2 below. For simplicity's sake the diagram only contains one box for each entity (e.g. a name), however there will of course be many names, and many taxon concepts in NZOR. Mapping the relationships between names and between taxon concepts is not conveyed in the diagram however these are in scope.

Out of scope for NZOR are storage of usage concepts and assertions used within collections and observations, and other data about taxa such as distributions, threat status, desirability and control status. These are out of scope because within the federated bioinformatics setting that is currently emerging, for many practical and technical reasons those types of information are more appropriately managed in other systems (e.g. NVS, NHMS, FBIS, BRAD), rather than kept in NZOR itself. Providing links to other information such as distributions, literature, identification keys and images held in other places is in scope for NZOR, however it is likely to be some time before this information is prioritised for inclusion.

4.3 Data Scope

4.3.1 New Zealand Relevant

The scope of NZOR includes all terrestrial, freshwater and marine organisms that are relevant to New Zealand. This means:

- All organisms that are native to New Zealand
- All organisms that have been introduced to New Zealand (whether they are naturalised or in cultivation)
- All organisms that have been determined as absent from the New Zealand political region, but are specifically considered as undesirable to New Zealand from a biosecurity perspective
- All organisms that have been determined as absent from the New Zealand political region, which are considered undesirable by countries that may import New Zealand goods (i.e. country freedom lists)

The inclusion of absent, undesirable taxa are necessary to meet the requirements of MAF Biosecurity. Basic biostatus information will be recorded for all taxa, specifically each taxon's origin (e.g. indigenous, endemic, foreign) and its occurrence (present or absent).

The approximate number of taxa that could be in NZOR is:

Type of taxa	Number of taxa
Known native and naturalised taxa	55,000
Estimated cultivated taxa	40,000
Absent, undesirable taxa	20,000
Total	**115,000**

Of the known native and naturalised taxa approximately 20% are undescribed, and it will take some years before this research is complete. Of the cultivated taxa, only 27,000 are currently documented in a form accessible to NZOR and so it is likely to be some years before the full 40,000 taxa could be included. In the future something like another 100,000 taxa could potentially be added due to the estimated numbers of undiscovered New Zealand species. It should be noted that approximately 60,000 of these are nematodes (worms) and 23,000 are arthropods (insects and spiders etc), and it will take many, many years before these are described. A more detailed breakdown of numbers of New Zealand taxa by environment (terrestrial, freshwater, marine) and by kingdom (and for Animalia by phylum) is included in Appendix 2.

Eventually NZOR could include all these taxa. It should be noted that while NZOR is likely to contribute some funds to the digitisation of records of currently described taxa, it is not expected to fund any taxonomic work to describe the known undescribed, and undiscovered taxa.

4.3.2 Tag, Vernacular and Trade Names

Tag names and vernacular names are somewhat contentious in the biosystematics and biodiversity management community. This is because they are not truly 'scientific'. That is, they are not created or used according to any agreed set of rules (such as the International Codes of Nomenclature). Their use is typically ad hoc, and not verifiable by reference to type specimens in formal collections, nor to peer reviewed publications. Their relationship to actual taxon concepts is often varied, and somewhat vague. They do however have great utility, for managing suspected new taxa that have not been formally identified, and for acting as a reference point into scientific names respectively.

Trade names, variety names and cultivar names are similarly complicated. These, along with tag and vernacular names are included within the scope and architecture of NZOR. The reasons for this are summarised in the following subsections.

4.3.2.1 Tag Names

As explained in the definitions section above, tag names occur when something thought to be a new taxon is found, but it has not been formally named using the codes of nomenclature and the scientific publishing process. The informal name given is referred to as a 'tag name'. Tag names are of particular importance to DOC, as often they must manage and protect a taxon for some time before formal nomenclatural and taxonomic work can be done. Sharing access to tag names between Regional Councils, NGOs such as QEII Trust, and DOC using unique LSIDs for particular tag names, while not as rigorous or exacting as scientific names, can be very useful in coordinating efforts to protect threatened taxa. It can also help raise awareness of and prioritise potential new taxa for proper scientific identification, circumscription and naming.

Some rigour can still be applied to tag names, and this is suggested for NZOR. For example, tag names could be included, but only where a voucher substantiates them. For example, "Gingidia aff. baxterae (White Rock Station; AK 299889)" means that there is an entity, it "may" be distinct, it occurs at White Rocks Station (or was first found there, or recognised as potentially distinct there) and the specimen lodged to account for it is at the Auckland Museum Herbarium, 299889. Using this method the specimen can be checked in the future to ascertain the distinctiveness of the taxon. This also means that it is possible for the tag name to become a synonym (albeit nomenclaturally invalid, and marked as such) of the correct name, once that taxon has been formally described.

It is important to be clear that entities represented by tag names are 'possibilities' rather than scientific 'facts'. As such they should be carefully distinguished in NZOR so there is no confusion between them and scientific names.

Tag names are therefore in scope for NZOR, however they are lower down the priority list than scientific names for population of the system.

4.3.2.2 Vernacular Names

Vernacular names are an imprecise method of referring to a taxon concept. As described in the definitions section above, vernacular names are often applied differently depending on the geographic regions, societal groups, and languages. Despite their ambiguity they provide an important access point for users to access the scientific names for the taxa. As such they are in scope for NZOR in terms of the system architecture and data structures, however vernacular name data will not be high priority for inclusion initially.

Their inclusion in NZOR serves a number of purposes. Firstly they improve access to scientific names (and NZOR) for a wider range of end-users by allowing them to search on a name they are familiar with. Secondly, they show users all possible taxa (and hence available data) the vernacular name may refer too, reducing their frustration and increasing the chance of them finding what they are looking for. Finally, they educate users in the ambiguity of use of vernacular names and the need for more precision in many contexts.

4.3.2.3 Trade Names, Variety Names and Cultivar Names

A "cultivar" is a taxon arising in nature or cultivation that is formally recognised and named. The names of cultivated plants are frequently difficult to verify and apply, however the naming of cultivars, cultivar Groups, and graft-chimaeras is now covered by the International Code of Nomenclature for Cultivated Plants (ICNCP) and requires deposition of a voucher specimen to help identify a new cultivar. The ICNCP does not regulate trade names (or trademarks) nor the naming of plant varieties.

A "plant variety" is a legal term governed by the International Union for the Protection of New Varieties of Plants (UPOV) Convention and the legislation of each country. It is distinct from the rank "variety" recognised by the ICBN. Registration of plant variety provides the breeder with some legal protection for the production and sale of the protected 'variety'. In this context variety may mean any cultivar, clone, hybrid, stock or line of plant.

A tradename, or trademark, is legally defined word(s) or that are used for the selling and marketing of a plant variety or cultivar. Tradenames are usually indicated by the symbol "™", however when this is omitted there may be confusion between tradenames and cultivar names.

These names fall within the scope of NZOR as they provide an important access point to scientific data for a variety of users, and will also enable linkage between NZOR and registers maintained by other agencies (e.g., the Plant Variety Rights Office). The inclusion of these names will be considered within the technical design of NZOR, however, considerable effort will be required to mobilise these data, and this may not be prioritised in the initial stages of NZOR.

4.3.3 Marine

A New Zealand Organisms Register would not meet its vision of being a catalogue of all New Zealand biota and other taxa of importance to New Zealand if it did not include marine organisms. From an information management perspective there is no substantive difference between marine organisms and terrestrial or fresh-water organisms. As such no additional expenditure on the software, and none, or very little extra on hardware and storage infrastructure would be required if marine organisms were included.

A number of potential end user organisations of NZOR, in particular the central government agencies such as MFish, MAF, ERMA and DOC have responsibilities that extend out into the marine environment. MAF for example currently send organism samples of importance in biosecurity decision making to the Marine Invasives Taxonomic Service at NIWA, where they are identified by marine taxonomists. Having ready access to definitive marine names, synonymies and other taxonomic data is likely to improve these taxonomists ability to provide this service to MAF. Having a connection between NZOR and MAF databases could also improve the quality of MAF marine risk assessment work, in much the same way that BRAD does for plant pests.

For DOC, having access to marine organism names is important for exactly the same reasons as it is in the terrestrial and freshwater environments. Being able to accurately identify, name, and refer to marine taxa in DOC information systems is crucial in doing the research needed to establishing marine protected areas, identify and manage threatened species, and understand the impacts of invasive species.

The Ministry of Fisheries is another potential end user organisation for which understanding organisms in the marine environment is fundamental. MFish has good existing sources of taxonomic/nomenclatural data for fish species, and the problems of name and synonymy changes for fishes are not large. Increasingly, marine researchers (working on projects for MFish, and others) need to identify large numbers of other marine organisms (such as plants, invertebrates and other animals) for biodiversity and ecosystem analyses. Over time, having well managed marine organism names and taxonomies could improve the quality of this science, and therefore the advice given to agencies such MFish and DOC.

The advantages of NZOR for organisations working in the marine area are very similar to those working in the terrestrial area. An authoritative list of taxa would be used:

- As an on-line reference source for names information
- To ensure data quality of names used in marine research projects and databases
- To ensure that databases containing names are updated when names change
- To improve communication between agencies on marine biota and ecosystems matters

In an international context, the regional node for OBIS (the Oceans Biogeographic Information System) has a requirement for an authoritative list of names of organisms in the New Zealand marine environment. NZOR could meet this requirement if it includes marine organisms. It is envisaged that the OBIS data loader would link directly to NZOR for definitive names information, hence ensuring the data quality of the information loaded into the OBIS regional node. Also, NZOR would provide taxonomic hierarchy information for OBIS users, eliminating the need for OBIS to duplicate this.

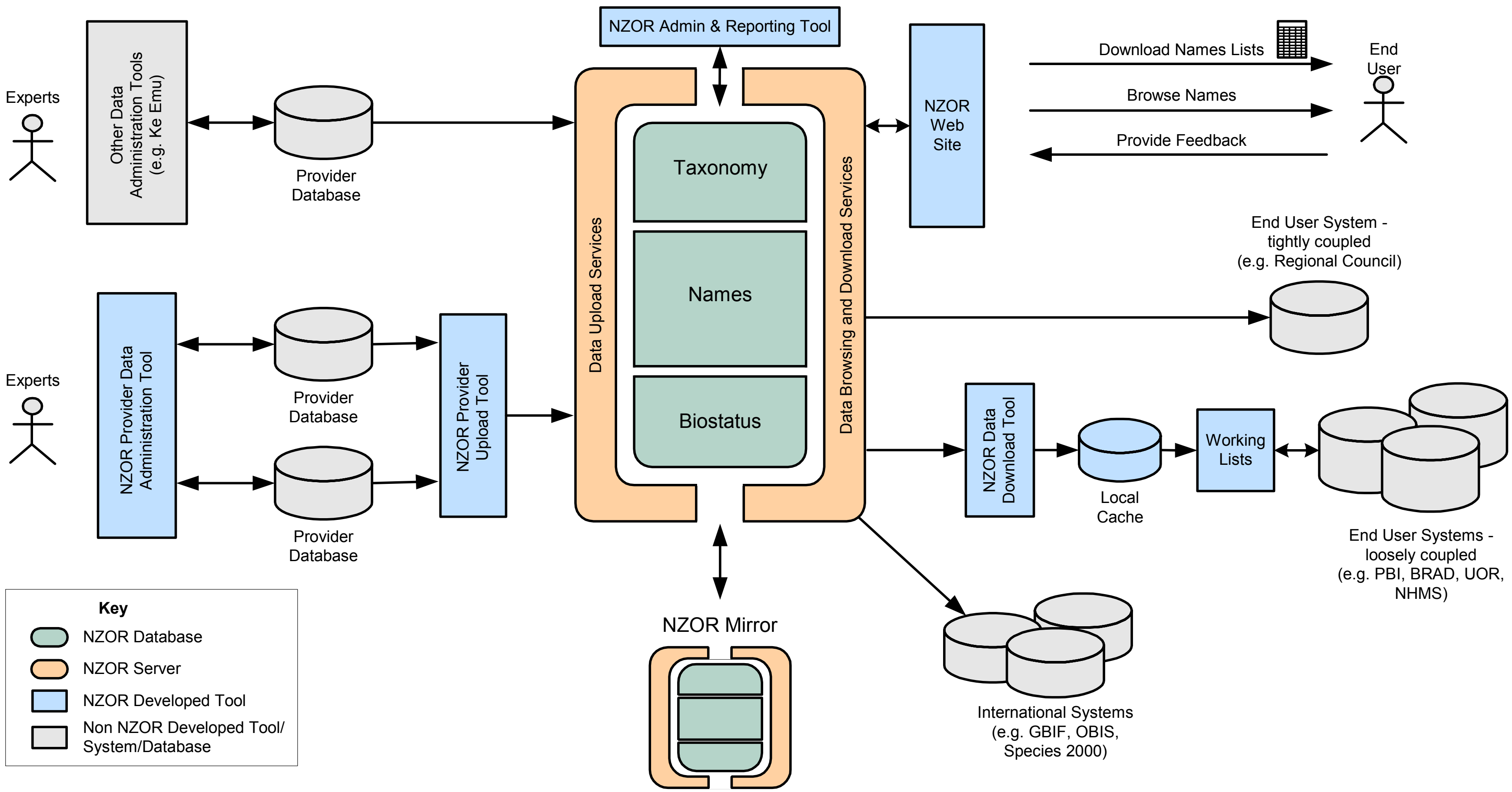
There are two components to the inclusion of marine organism data into NZOR. The first is the initial population of the system with available electronic data from NIWA and Te Papa, and with the Species 2000 New Zealand lists. If this were done at the same time as importing the terrestrial and freshwater data there would only be a marginal difference in cost. For this reason it is proposed that this component be funded as a part of the initial project to create the system and populate it with baseline data.

The second component is the addition of other marine data in the future. As with the ongoing funding for addition of other terrestrial and freshwater data, this would be expected to be a part of end user driven and funded priority setting processes.

The steering team for this Scope acknowledge that including marine organisms in the project as proposed could be seen as being out of scope for the TFBIS programme. All participants at the initial workshop convened to assess departmental stakeholder buy-in for the concept, agreed that it was important that NZOR address all organisms and that there was not another obvious source of funds that could support separate treatment of a marine organisms register. There are significant efficiencies to be gained by addressing all organisms, and as such steering team strongly recommend that if NZOR is funded that it include marine species.

4.4 System Boundaries, Services and Tools

The following diagram shows the components of the NZOR system, and the relationship between NZOR and provider and end user systems.



The NZOR server will contain the database storing all NZOR data, and a set of software services providing upload, download, and browsing facilities. Also hosted on the NZOR server will be the NZOR web site, and a set of administration and reporting tools.

Providers will be able to interact with NZOR in a number of ways. If they already have sophisticated nomenclatural and taxonomic data administration tools and their systems can send data in the correct formats to NZOR, they may connect directly through the NZOR upload services. If they do not, they may choose to use NZOR developed taxonomic data administration and data upload tools to organise their data and transform it into NZOR approved formats.

End users will also have a number of methods by which they can interact with NZOR. Individual users may browse the NZOR web site, search on taxa, and download lists of names. Some end users may wish to connect their systems directly to NZOR in a 'tightly coupled' fashion, using NZOR effectively as a 'data dictionary' for lists of organism names. Other end users will be able to connect in a 'loosely coupled' fashion. This will involve downloading data into a local cache and using a set of working lists to control which data they integrate with their various systems and datasets. This is the kind of approach that is most likely to be taken by larger organisations such as MAF and DOC who have multiple systems that require access to organism names.

5 Achieving Consensus

Nomenclatural and taxonomic data will be provided to NZOR by a number of different individuals and organisations. The quality of the data and relative 'qualifications' of the provider to contribute those data must be managed. For the large majority of the data in NZOR (excluding for example tag names), these data must be supported by valid scientific publications. Taxonomy is a subjective activity however, and as discussed earlier in the document there are at times different taxonomic treatments that create competing or at least overlapping taxon concepts. These must be resolved to determine the 'active concept' in order to provide end users with a list of correct names for political, legislative, and biosecurity and biodiversity management purposes. The way data quality will be managed, and the way differences of taxonomic opinion will be resolved is the subject of this section.

NZOR is, at its core, an informatics infrastructure development project. It is a tool for data providers to more efficiently make their nomenclatural, taxonomic, and biostatus data accessible to end users. The way data quality and disagreements between conflicting taxon concepts are managed however has implications for the success of, and costs for the project in the long term. In achieving the vision for NZOR there are a number of options ranging from providers bearing primary responsibility for data quality, and the NZOR governance structure only being invoked by exception where there are significant issues, through to a formal NZOR funded set of committees that review and approve all data entering the system. These options are explored in the following subsections.

It is important to note that the order that data will be entered into NZOR during the initial implementation and first year or two of operation means there are unlikely to be significant amounts of data overlap or conflict in the early stages. This is because priority will be given to nomenclatural data, and primary taxonomic data from the major providers. Additional taxonomic data, including alternate concepts, vernacular names, and concept correlations will not be entered initially.

5.1 What Is Being Decided

One of the primary objectives for NZOR is to provide a consolidated list of New Zealand relevant taxonomic names for end users including MAF Biosecurity, ERMA, MFish, DOC, Regional Councils, the Pest Management community, and the research community. Some users will simply want a single list of preferred names while other users will require synonyms and other taxonomic and biostatus data. In order to deliver on these requirements it will be necessary to a) ensure that data being put in the system are correct, accurate and come from a reputable source, and b) to be able to resolve differences between competing taxon concepts.

Decisions are therefore necessary on:

1. Who can be a data provider to NZOR
2. Which, out of a set of competing taxon concepts, is the 'active concept'

There are a number of models for making these decisions. These are explored in the following subsection.

Where decisions or changes are made on 'active concepts' the alternate concepts will not be removed from NZOR. For reasons of technical continuity, and historical accuracy once a name or concept is in NZOR it will always be available, or a link to the new record will be available if a record was deprecated by a provider. Agencies therefore still have access to names and 'non active' concepts and could use these at their discretion in making decisions about importation licences, biosecurity risks, or conservation management programmes.

There are a significant proportion of taxa for which only basic nomenclatural data exist, and complete synonymies have not been compiled. In these cases it is important to note that any NZOR committees or experts will not be tasked with 'doing taxonomy on the fly'. In the cases where these data do not exist these names will be flagged, the need for nomenclatural work on that group will be noted, and conveyed in some way to data providers.

5.2 How Decisions Might Be Made

There are at least two models that could be used. The first involves having little or no 'certification' of data providers, and the responsibility for data quality and resolving disagreements being held by NZOR as an organisational entity, using a hierarchy of expert committees. The second involves having more rigorous provider 'certification', data quality being the responsibility of the provider, and having a transparent process for resolving conflicting data only when it is absolutely necessary.

5.2.1 The Committee Approach

In this approach all data that are entered into NZOR must be approved by an expert or committee in the NZOR organisation. This organisation would of course be virtual, NZOR is very unlikely to employee staff, however these individuals and groups would have roles and authorities defined within the NZOR governance structure. Their activities may be funded by NZOR, and/or their own organisations.

Under this model any person or organisation may contribute taxonomic data to the system. There will be particular formats they will be expected to submit data in, however the primary responsibility for checking the quality and veracity of the data will be NZOR's. To undertake this there would be a number of committees, each tasked with responsibility for a different taxonomic group. In this approach it is estimated that somewhere between ten and twenty committees might be necessary. A possible list includes committees for Bacteria, Protozoa, Chromista, red and green algae, vascular plants, Bryophytes and lichens, Fungi, vertebrates, hexapods (insects, etc), other arthropods, worms (more than one), molluscs and gastropods, other marine invertebrates (more than one), other land invertebrates (more than one).

Committees would review data submitted to NZOR and approve it being made live on the system. They would also decide which out of a set of conflicting taxon concepts was to be flagged as the active concept. It would not be the role of the committees to generate new data.

There are a number of constraints to this committee based process. These are expertise, workload, and trust. People involved must have enough expertise to be able to make informed decisions about competing taxon concepts. For many taxonomic groups there are very few people who have this expertise and in some cases the only expert may be the data provider themselves. Where there is only one expert they are likely to make the decision, and have the committee simply validate it. There may also be instances where the required experts are not on the committee, and the committee must commission a written opinion and validate that. In other cases the decision may be an informed consensus from the whole committee. Where there are larger numbers of people who could contribute to a decision there is the desire to keep a committee small enough to be able to make decisions efficiently. If too small a committee is tasked with a large taxonomic group, with a lot of competing concepts, they may struggle to manage the workload. The workload for committees could be moderately high on an ongoing basis, however it is likely to be significant during the initial population of NZOR with data. Finally those involved in the decision-making must be trusted by their peers so that the decisions will be accepted by providers and end users alike.

The committees may use a set of guiding principles to aid in decision-making. In Australia a set of such principles, and a decision-making method have been developed by the plant systematics community to resolve alternative taxonomies within the Australia Virtual Herbarium. These guiding principles include:

> *the need for monophyletic taxa, minimising taxonomic change, understanding that some taxa have strong 'interest groups', making it clear that 'preferred name' does not necessarily imply 'best name' on all criteria, avoiding epithets used in possible congeners, and the concept of 'majority rules' when states and territories have differing views.*[1]

The last point deals with the fact that while there may be consensus within particular states (e.g. Victoria, New South Wales), there may be differences of view between states. While New Zealand does not have this issue, the differences in opinion between organisations (e.g. between Landcare Research and DOC) are perhaps similar in nature. It is unlikely that a 'majority rules' approach would work as determining what comprises a majority would be difficult. It seems more useful to simply have representatives on the appropriate committees from organisations across which there may be differences of opinion.

For some groups such as marine invertebrates the sheer number of taxa and sparse expertise may mean that the workload needs to be divided up amongst more, smaller committees, or over a longer time period.

There may be a hierarchical relationship between committees, for example each committee could have a representative that served on an overall committee that ratified all decisions made.

[1] Entwisle, T.J. & P.H. Weston. 2005. Majority rules, when systematists disagree. Aust. Systematic Botany 18:1-6.

Committees could be appointed by the NZOR governance committee, or they could be democratically elected committees by their peers.

Decisions as to actual committee membership are outside of the task of this scope.

5.2.2 The Provider Driven Approach

In this approach data quality and veracity is primarily the responsibility of each provider. Each provider must demonstrate certain capabilities and meet certain standards in order to have registration approved by the NZOR System Governance team, or by a set of peer referees. Over the longer term, decisions on which of (and what order in) a providers' data were put into the system would be made through a prioritisation process largely driven and funded by end users.

Where there were conflicts in the data this would be resolved through a set of escalating measures:

1) Initially contact would be made between providers to see if the conflict was a result of misinterpretation
2) If this did not result in resolution a third party expert opinion would be sought
3) If this still did not provide a resolution an expert committee would be formed to deliberate and if necessary to vote on the issue

A similar process would be used where a provider failed to address quality concerns expressed by end users.

Under this model the issue of workload is not as significant, as committee decisions on which taxon concept should be 'active' will only be necessary where there is unresolved conflict.

As above guiding principles such as those used in the Australian situation might be used to inform decision making.

It seems likely that in order to gain buy in from the systematics community it would be necessary for any expert committees to be democratically elected or peer nominated rather than simply appointed by the NZOR System Governance team.

5.2.3 Which approach is best?

Experience in Landcare Research in managing taxonomic data and the success of Species 2000 Project led by NIWA suggest that solutions to these issues are achievable, and the scale of the problems should not be overestimated. It is therefore likely that the Provider Driven approach will be an appropriate model, and more sustainable than a process relying heavily on committees. However, it is outside the scope of this document to determine the final structure for the decision making process. It will be the responsibility of the NZOR steering committee during the implementation phase to develop an appropriate policy for governance and decision making, and to prepare guidelines for these processes.

5.2.4 Technology Mediated

Whichever approach is taken it is suggested that the quality assurance and conflict resolution process be supported by technology. Automated systems will be able to detect and report where there is potential overlap and conflict between data.

In addition a number of discussion forums could be incorporated directly within the NZOR system, and could be tied directly to particular taxonomic groups or levels. This would mean that any interested parties, from both research organisations and end user agencies could focus specifically on the groups that interest them. They could move in and out of the discussions based on their expertise and interest without having to have formal, sanctioned membership in particular committees.

These interested parties could raise issues of quality and competing taxon concepts, and could in some cases suggest and achieve resolutions themselves. Any unresolved issues would then be handled through one of the models above.

6 Ownership and Funding

This section describes the expected relationships between providers and end users in terms of data stewardship, system hosting, support and maintenance, and data curation responsibilities. It also outlines potential models and sources for ongoing funding for the supporting technology for NZOR, and for ongoing entry and curation of data.

6.1 Stewardship, Custodianship and Curation

NZOR will be an important piece of national informatics infrastructure supporting research and management in biodiversity and biosecurity. Ideally the ownership of the system should vest with the Crown rather than any particular agency. In practice the underlying software is likely to be created under an open source licence as it may also be used and contributed to by other countries.

The data that will be stored in NZOR are all in the public domain. While not all data will be visible to all users (in particular in relation to biosecurity and export issues), none of the data will have any intellectual property constraints attached. Data providers will be responsible for the data they contribute. They will be able to withdraw as a provider from NZOR, however the data they have contribute to that time will remain within the NZOR system.

NZOR is likely to be hosted at Landcare Research and at NIWA as mirrored systems. Mirroring of NZOR is necessary because it will become a very important system over time, and hosting at sites in Lincoln and Wellington will provide an adequate level of disaster recovery.

A distinction is therefore drawn between stewardship, custodianship and curation. Landcare Research and NIWA would be custodians of the data, hosting it and maintaining the software and hardware systems that support it. Individual providers would be the stewards of the data, responsible for its provenance and quality. They will also have curation responsibilities in maintaining the data they contribute. Depending on the governance model used NZOR committees may also have stewardship and curation responsibilities.

It should be noted that Landcare Research and NIWA will also be significant providers of data, and for the research other than taxonomy they in engage in (e.g. ecosystems research) they will also be significant end users of the system.

6.2 Funding

Section 9 defines three major workstreams for the project: the technical platform, data content, and governance. Each of these has some initial establishment costs, and some ongoing costs. The funding sources, structures and models for each of these workstreams and phases could be different.

Potential sources of funding for NZOR include TFBIS, MoRST/FRST, MAF, DOC, Landcare Research, NIWA, Envirolink, and GBIF. The way these sources of funding could be applied to the workstreams and stages for NZOR is as follows:

Work Stream and Component	Potential Funding Sources
Technology Platform	
Detailed requirements definition and base NZOR technology design and enhancement	**TFBIS** **MAF** – capital expenditure through budget for Biosecurity Science Strategy initiatives **Landcare Research** and **NIWA** - in kind contribution of existing technology frameworks, IT infrastructure and technical support
Data import and editing tools, NZOR website and web services, end user tools	As above **Envirolink** for tools for Regional Councils
Additions to TCS standard to facilitate data exchange, and validation as a nationally accepted standard as part of eGIF	**TFBIS** **SSC** – contribution of labour and expertise in assisting in standards certification process
Ongoing system maintenance and support	**MoRST/FRST** – science backbone funding **MAF/DOC/CRIs** – through operational IT expenditure
Data Content	
Initial population with names and taxonomic data from major providers, names data from Species 2000 NZ, and easily available biostatus data	**TFBIS**
Harvest additional 'end user' repositories of names from DOC/BNZ etc	**TFBIS** **MAF/DOC** – capital or operational expenditure
Gap analysis reports for informing priority setting for additional data	**TFBIS** **MAF/DOC** – capital or operational expenditure **FRST/Landcare/NIWA** – Defining NZ Biota OBI
Addition of other priority data	**TFBIS** **MAF/DOC/others** –operational expenditure (based on their priorities for data)
Governance	
Project governance (project steering committee, project management)	**TFBIS** **MAF** – capex through budget for Biosecurity Science Strategy projects
On-going system governance	**MoRST/FRST** – science backbone funding **MAF/DOC** – capital or operational expenditure
Development of policy and process for provider certification, data governance	**TFBIS**
Data governance, committees to resolve conflicting taxon concepts	**MAF/DOC** – operational expenditure **FRST/Landcare/NIWA** – Defining NZ Biota OBI

The nature of these potential funding sources and their relationship to project workstreams and components suggests the following approach to funding the system.

1. TFBIS funds the establishment of NZOR including the initial technology platform and tools, the population of the system with existing digital data from the key players (Landcare Research, NIWA, Te Papa), baseline data from Species 2000 NZ, and relevant data from global sources (such as Species 2000/Catalogue of Life), along with initial work to establish data policy and project/contract management processes. This may be supported in part by funding from MAF for Biosecurity Science Strategy initiatives if that is available and NZOR is seen as a priority use of those resources.
2. Ongoing addition of priority data is funded through operational expenditure by MAF, DOC and other end users based on their needs. Some support for that priority setting process may be available through relevant FRST OBIs. Future OBI and other appropriate government funded projects could contractually require deposit of data from taxonomic research into NZOR. Additional data could also of course be funded by TFBIS as separate projects run through the normal TFBIS funding rounds.
3. Ongoing system governance and maintenance could be funded through MoRST/FRST backbone science (NZOR could for example become a new nationally significant database), and through contributions from key end users such as MAF and DOC

7 National and International Context

7.1 Related New Zealand Strategies and Programmes

The NZOR project is related to a number of other initiatives and strategies, both national and global.

In the national context the project aligns closely with the objectives of the Digital Strategy, the e-Government Strategy, and the policy work MoRST is undertaking on research data saving and sharing and backbone science infrastructure. It is also aligned with the objectives and actions in the Biodiversity Strategy, the draft TFBIS Strategy, and the draft Biosecurity Science Strategy.

The recently released Digital Strategy states that:

> *"to become a true Knowledge Society, New Zealand needs to focus on information-rich activities: those in which we create, collect, manage, process, store, move, or access information via networked environments."*

The Digital Strategy contains three enablers, Content, Connection and Confidence. Within the Content enabler, the National Library is the lead agency for the National Content Strategy, currently in draft. This involves:

> *"Bringing New Zealand online by mapping New Zealand's information assets and developing a framework and policies for national access. Identifying criteria for what should and shouldn't be digitised from existing holdings. An information architecture will be developed to preserve, share, and manage digital objects."*

Within the biodiversity and biosecurity context, the NZOR will be a very important part of this architecture. Staff at the National Library have been made aware of TFBIS and this project and have expressed enthusiasm for it as an early example of the kinds of systems and behaviour they hope the National Content Strategy will support.

The e-Government strategy says that:

> *"By June 2007, networks and Internet technologies will be integral to the delivery of government information, services and processes. By June 2010, the operation of government will have been transformed through its use of the Internet."*

The ICT unit of the State Services Commission is responsible for guiding the implementation of the strategy. A number of initiatives are involved including the e-Government Interoperability Framework (eGIF). As a part of the eGIF the ICT unit has developed a standards definition process and some discussions have already been initiated between the NZOR project team and the SSC, on this topic.

NZOR will also contribute to objectives of the Biodiversity Strategy, the draft Biosecurity Science Strategy, and the TFBIS Strategy, and is closely aligned with a number of MoRST and other agency initiatives as detailed in the following subsections.

7.1.1 Biodiversity Strategy

The NZOR directly contributes to achieving numerous actions identified in many themes throughout the NZBS. It is perhaps most directly relevant to Theme 9: Information, Knowledge and Capacity*, where NZOR will contribute to specific actions.*

Objective 9.1 Expand the research frontier

Identify and fill critical gaps in scientific knowledge, including applied research, and prioritise and coordinate future research to address key issues and threats to biodiversity.

Objective 9.2 Use ecosystem-based methods to map our indigenous biodiversity

Develop and implement effective approaches to map indigenous biodiversity at ecosystem scales and inform management actions and research.

Objective 9.4 Reporting and adaptive management

Ensure that local, regional and national reporting on the state of indigenous biodiversity informs ongoing priority setting for biodiversity management and research as a key part of an adaptive management approach.

Objective 9.5 Share information and best practice

Consolidate and share existing and new information, methods, technologies and management experiences so that others can benefit from relevant knowledge about indigenous biodiversity.

Objective 9.6 Build capacity

Enhance the capacity of people and organisations to fulfil their responsibilities to conserve and sustainably manage New Zealand's indigenous biodiversity.

And also, Theme 10: New Zealand's International Responsibilities

Objective 10.1 International forums and treaties. Contribute towards the international effort to conserve and sustainably use global biodiversity through participation in relevant international forums and treaty systems.

7.1.2 A Biosecurity Science, Research And Technology Strategy For New Zealand

The draft of this strategy, published for consultation on 8 November 2006 has a goal and set of actions that related to NZOR. The relevant excerpt from the strategy is as follows:

9.3 Goal 3: Science Uptake: Ensuring that science is responsive to biosecurity needs and priorities and that uptake is timely and effective

Objective 3.1: To ensure ease of access to biosecurity science outputs so that uptake is timely and effective

It is important that people working in the operational end of biosecurity are aware of the research which is currently underway and the outcomes of finalised projects. This ensures that potential benefits and applications of that research are identified and can be integrated with a number of tools and systems.

Action area 3.1.1 Develop systems to enable rapid and efficient access to up-to-date biosecurity science information

Biosecurity science information is currently distributed in numerous unconnected databases and in both peer reviewed and "grey" literature such as project reports. Effective uptake of science into all biosecurity activities from operational through to international standard setting and trade negotiations requires that this information is easily accessible at international, national and regional levels so the best science can be used to inform biosecurity decision making. Information systems must enable efficient reporting and data retrieval relating to biosecurity risks to New Zealand.

Key actions:

1. Develop and adequately resource an information system which links relevant databases and other science outputs and knowledge systems to ensure easy access to biosecurity science information including:
    - international pests and diseases which are a risk to New Zealand;

- pests and diseases already present in New Zealand;
- which host species are susceptible or already affected;
- associations of pests and diseases with commodities;
- impacts of the pest or disease in particular areas;
- relevant literature including published reports;
- management options, including surveillance and monitoring options.
- published sources of Mātauranga Māori and published information relating to Māori and biosecurity science.

2. Review smaller local focus databases on the distribution of diseases and exotic species and link these where appropriate.
3. Develop processes for agencies with responsibilities for biosecurity to review and monitor their use of science and share science results.
4. Improve access to taxonomic information through maintaining and enhancing national and international networks and through the use of information systems.

Action area 3.1.1, and key action 1 are listed as high priority items in the draft strategy. The similarity of Key action 1 to the role TFBIS plays for biodiversity is noted. Through discussions with MAF Strategic Science it is apparent they intend to take a cross agency, federated approach to key action 1. As such there are likely to be potential synergies between this and TFBIS initiatives in approach, architecture and infrastructure, and in particular in relation to key action 4, with the NZOR project.

7.1.3 TFBIS Strategy

The draft TFBIS Strategy also supports the idea of shared informatics infrastructure and says:

> *"Many components in this 'meta-system' can be created by individual agencies. While serving the greater good, they are often targeted at solving particular problems, or improving research in a particular area. Some components, however, are essentially 'infrastructure' – they benefit the whole system in a 'diffuse' sort of way. They become very valuable, but there is often the lack of individual gain for any one organisation to 'just build it'. It's a bit like asking every driver to build all the roads in their suburb. If there is sufficient shared vision, however, TFBIS may be able to fund some of these 'infrastructure' components through projects involving collaboration between a number of agencies."*

In regard to the funding profile for TFBIS for the next two to three years the draft strategy says there will be:

> *"A significant increase in spending on providing underlying and connecting infrastructure, along with supporting standards, privacy, data access, and intellectual property policy"*

7.1.4 MoRST Initiatives

Over the last two years MoRST has undertaken significant internal work in the area of research data saving and sharing. In June 2005 MoRST held a national workshop on these issues. The workshop brought together around 50 people from research funding agencies, Government departments and research providers in the social and environmental sectors. The following is an excerpt from the workshop summary report:

> *The workshop strengthened the vision of a federated, interoperable approach to access and sharing of publicly funded research outputs and data. This approach is consistent with initiatives to facilitate a 'whole of research, and science' approach across the*

tertiary, science, government and where appropriate private sectors to support the vision of public access to publicly funded research output.

In late 2006 this federated approach was used in planning cross agency data sharing agreements, standards and supporting infrastructure for the Ocean Survey 20/20 programme.

There is now an increasing acknowledgement by central government that databases are a key part of science infrastructure. In the Cabinet paper entitled "A More Effective and Stable Funding Environment for Science" published in April 2006, the Minister for Research, Science and Technology proposed five key changes to the science system, including:

iv. enabling non-competitive investment to support the "backbone" of New Zealand science (essential infrastructure, databases, and collections)

Cabinet tasked MoRST to report back to the Minister of Research, Science and Technology on the policy settings, criteria and funding options for these "backbone" investments during 2007. This is an important step in ensuring that primary data are preserved and well curated for future needs. This policy work is currently underway. It has focused in particular on the nationally significant databases and collections, and how the criteria for, and funding for those and similar important resources could be better structured, and more stable in the future. Through informal discussions between members of the NZOR project team and relevant MoRST staff it is apparent that there is a potential case for NZOR becoming one of these 'backbone' systems in the future. Further work would be required on both sides to advance this, and potential funding from this source for aspects such as ongoing maintenance and infrastructure support costs for NZOR are by no means certain. MoRST is however enthusiastic about the role NZOR could play as a core component of the federated environmental informatics system, and the benefits this could have for improving science outputs.

The project is also likely to support the objectives of the MoRST/FRST Envirolink programme in that it will increase the availability of accurate taxonomic names, and associated research data and findings to regional councils.

Envirolink is a funding scheme initiated by MoRST and delivered through FRST. It is specifically targeted at supporting Regional Councils to access advice and management tools from CRIs and some not-for-profit research associations. This is done through small advice grants, medium advice grants, and tools development. Tools development has some relevance to NZOR. From the FRST website:

Tools development funding is used to support development and adoption of natural resource and environmental management tools for use by all regional councils and Unitary authorities. These tools may be physical technologies or something more conceptual, such as a formalised or systematic approach to problem solving or analysis.

It is possible that tools to allow connection of taxonomic names to Regional Council information systems could be developed from, or at least part funded by the Envirolink programme.

7.1.5 Other Initiatives

The NZOR project is aligned with NIWA and Landcare Research FRST-funded Outcome Based Investment programmes within the Resilient, Functioning and Restored Natural Ecosystems Portfolio, and in particular the Defining NZ Biota OBI. It will help support taxonomic research, and assist in priority setting processes within this research programme.

The project could also play a role in a proposed Landcare Research Kiwi Advanced Network (KAREN) capability project. This project will utilise GRID-based technologies developed in the US to support the Geospatial Informatics community (through the GEON project) and the

Ecological Informatics community (through the SEEK project). Advanced network infrastructures will be used to develop linkages between ecological, spatial & taxonomic information resources.

7.2 International Context

In the international context the project supports the objectives of the Ocean Biogeographic Information System (OBIS), and the GBIF Electronic Catalogue of Life (ECAT) programme. It will assist in meeting New Zealand's international obligations including the Global Strategy for Plant Conservation (GSPC) Target 1 by 2010, "A widely accessibly working list of names of known plant species, as a step towards a complete world flora", the Global Pollinator Initiative, Global Taxonomy Initiative and others.

Landcare Research is actively involved in related international work. NZOR is directly aligned with a US$200k GBIF-funded project recently awarded to LCR to prepare 'A Working Checklist of the Compositae". This 18-month project involves collaboration with nine international partners including Kew and the Smithsonian Institute to develop an "organism register" specifically for the daisy family, which constitutes 10% of the world's flora. LCR is also currently funded under a joint TDWG/GBIF project to further develop Life Science Identifiers as an enabling technology for sharing taxon concept data.

LCR's NZFUNGI fungal names database was originally constructed using the IndexFungorum global fungal nomenclator as a backbone resource and is thus already globally linked. IndexFungorum is a nomenclatural database of over 300,000 names of fungi developed at CABI in the UK. IndexFungorum is managed for the global public good under a cooperative agreement between CABI, Centraalbureaux Schimmelcultures (Netherlands), and Landcare Research. The IndexFungorum Partnership and the subsequent development of NZFUNGI has already demonstrated the effectiveness of global taxonomic data sharing and is an example of the potential of NZOR.

Through a TFBIS project to digitise New Zealand entomological literature LCR and Canterbury University have collaborated with the UBio (Universal Biological Indexer and Organizer) group at the Wood's Hole Marine Biological Laboratory in the US. UBio has developed tools to automatically extract organism names from scanned literature. UBio currently provides one of the largest global organism name indexes, and is the lead agency in a new, ambitious and substantial project to develop a complete Encyclopedia of Life.

The international work carried out in this area by LCR arises through collaborative networks developed through active involvement in the GBIF, Species 2000 & TDWG communities. See section on Landcare Research for more details.

7.2.1 GBIF (Global Biodiversity Information Facility)

New Zealand, through MoRST, is a signatory to GBIF. This international program was launched in response to the Convention on Biological Diversity and aims to make biodiversity information globally accessible. Participants provide biodiversity information through the country node(s). In effect GBIF is developing the capacity to access the vast numbera globally federated network of databases around the globe and currently serves over 100 million species-locality records. supported by GBIF's informatics infrastructure. The GBIF network currently provides access to 120 million primary collection/observation records from over 1,000 sources.

GBIF provides a metadata registry of the available global, primary biodiversity data and has already provided its own central portal that enables simultaneous queries against biodiversity databases worldwide. It is currently developing open web service interfaces for end users to construct their own thematic portals and specialised search facilities. In the long term, molecular, genetic, ecological and ecosystem level databases will be linked to the system.

GBIF has a number of active work programmes including Database Access and Database Interoperbility (DADI), Digitisation of Natural History Collections (DIGIT) and Electronic Catalogue of Names of known organisms (ECAT). The DADI and DIGIT work programmes are well developed. The ECAT work programme, which is most closely aligned with NZOR, is currently being developed in collaboration with Species2000/COL, UBio and a number of other partners. Jerry Cooper, Landcare Research, is a member of the GBIF ECAT and 'Nodes' Science Committees. David Penman, University of Canterbury, is the Chairman of the GBIF Governing Body, and Don Roberston, NIWA, is the Head of Delegation for New Zealand.

New Zealand does not currently fund a national GBIF Node to provide a clearing house for national biodiversity data, but does provide over 1.8 million observation/collection records to the GBIF network from both terrestrial and marine databases held by Landcare Research and NIWA.

7.2.2 Ocean Biogeographic Information System

OBIS is an on-line atlas that allows a user to find information about marine species or localities. For example, a user can ask where any marine species is found, or what lives in any part of the ocean and. users can overlay species and environmental information to discover the relationship between species distribution and other factors. OBIS is the marine node for GBIF and NIWA has built the South-western Pacific Regional OBIS Node which serves up around 500,000 species locality records.

7.2.3 Species 2000 and the Catalogue of Life (COL)

Species 2000 is a global initiative intended to identify global species databases which may be linked together in a federated structure to provide a consensus taxonomic database for all described species. It is registered in the UK (Reading University, led by Frank Bisby) and has received significant support from both the EU and GBIF. EU funding was awarded to develop Species 2000 Europa which is intended to provide a consensus list of organisms in Europe. The US based Integrated Taxonomic Information System (ITIS) is a federal organisation with a remit to carry out the same function as NZOR in North America, and has partners in Canada and Central America. ITIS and Species 2000 collaborate under the banner of the "Catalogue of Life". COL provides both an annual checklist distributed on CD and a dynamic checklist derived from the federated database linkages. It is the Catalogue of Life which currently provides the taxonomic backbone for the GBIF portal which in turn provides access to linked collection/observation data. These initiatives differ from NZOR in some important respects. For example, COL provides a single taxonomic view which is a reflection of the non-overlapping global species databases used to construct it. Species 2000 does not include cultivar names of plants. ITIS is not a federated system and uses editorial staff to develop the national catalogue. At a recent COL meeting director of ITIS expressed an interest in deploying Landcare Research's DTF management tool to facilitate data provision to ITIS.

In March 2007 Species 2000/COL will celebrate the delivery of 1 million species records, accounting for about half of all described species. Incorporation of the remaining 50% will be harder to achieve for a number of reasons. Significant residual data relates to organism groups where there is little global expertise or little agreement on a consensus taxonomy. In addition, there is a significant gap in Species 2000/COL relating to groups with regional endemism where those groups are not covered by existing global species databases. New Zealand is therefore relatively poorly covered in Species 2000/COL and NZOR will fill an important regional gap. It is for this reason that Species 2000 invited Jerry Cooper and Dennis Gordon to become members of the Species 2000 Global Committee to encourage the establishment of a Species 2000 regional hub for New Zealand. Species 2000/COL is also an important baseline dataset for NZOR particularly with reference to introduced taxa.

8 Technology

In order to achieve the vision, the NZOR requires technology infrastructure. This can be broken up in a number of ways, but for the purpose of this discussion, we identify the following 6 essential components:

1. A data model complex enough to store concept/biostatus data
2. A database to store concept/biostatus information for all New Zealand taxa
3. A data provider data management tool, guidelines and standards to allow data providers to prepare and send data to NZOR
4. A data administration tool at NZOR to manage the data once it has reached NZOR
5. A web application to allow users to query and report from the data in the database
6. Some end user tools to facilitate full use of the data (eg to compare lists, maintain a local cache etc)

These six requirements for NZOR are not satisfied by any existing system. However, if each of these components of NZOR is considered separately, it may be possible to modify parts of existing systems.

For example, the GBIF Electronic Catalogue of Life (ECAT) programme already aims to make available over the Internet an international database of biodiversity information including names of organisms. It is developing tools to allow federation of that data. In principle, one option would be for NZOR to develop data administration tools that made it easy for local data providers to prepare data for provision directly to GBIF, and let this internationally funded system host New Zealand's data and make it available to users over the existing GBIF website. This would however, have obvious disadvantages because the current GBIF portal cannot easily provide the required national view, with national control over content or quality, and cannot provide the required biostatus information. An advantage of this approach is it would, without additional effort, contribute New Zealand's national data to this international effort.

A second approach would be to use a locally managed version of an internationally developed taxonomic model. The Berlin Taxonomic Information Model is a European model and database structure developed by the Botanic Garden and Botanical Museum at the Freie universitat of Berlin, specifically for storing taxonomic concept data. This model has been used for the German regional node for GBIF, by the European Union for Euro+Med PlantBase (the information resource for Euro-Mediterranean plant diversity) and a small number of other international taxonomic databases. However, it would need modification for the particular requirements of the NZOR system, and there is little local knowledge of the model or the computer systems that it is built on.

A third approach would be to base NZOR on existing New Zealand technology. Landcare Research currently maintain extensive databases of taxonomic data for New Zealand plants and fungi. They also operate a concept based data administration tool (known as the Dynamic Taxonomic Framework) to allow Landcare Research staff to enter taxonomic data into these databases and provide website applications (such as nzflora.landcareresearch.co.nz) to make this information available over the Internet. The NamesWebService makes information from the plant names database available via web services to other organisations. Another Landcare Research initiative (the Global Compositae Checklist) has also developed sophisticated models and technology components similar to those required by NZOR. Again however, if NZOR were to use Landcare Research's existing models, they would need modification for the specific requirements of NZOR.

In particular, because NZOR expects to receive information from a number of data providers, the existing data administration tools would need to be generalised into distributed tools that would allow a number of providers to prepare data for loading into the database. This would, for example, include the need for tools for transferring data over the Internet, and standards and protocols for data transfer which are currently not required for Landcare Research's internal systems. There is an opportunity to re-use and help improve existing international standards for data exchange such as the Taxonomic Databases Working Group's XML based Taxonomic Concept Schema (TCS) or the Access Protocol for Information Retrieval (TAPIR).

The modifications required to bring existing systems up to the requirements of NZOR may be significant and expensive, but a fourth option, development of the model, database and tools from scratch, without re-use of existing components would be much more expensive.

An evaluation of approximate costs, advantages and disadvantages for each of the four options is given below.

8.1 NZOR facilitates data provision but data is stored in an international system

In this logical possibility, NZOR develops data administration tools that made it easy for local data providers to prepare data for provision directly to GBIF, data is stored in GBIF and made available to local users over the existing GBIF website. The provision of tools to assist users in accessing this data is an additional option.

Component	Approximate cost
Re-use of model underlying existing Landcare systems	$10,000
Re-use of existing Landcare database infrastructure	$60,000
Reuse of modified version of Landcare's DTF (improved)	$90,000
Users use GBIF website to view data	$0
Data users tools based on existing Landcare tools	$120,000
Indicative Total	**$280,000**

Advantages:

- Low cost
- Perception that New Zealand is participating fully in a global initiative

Disadvantages:

- New Zealand has little or no say in what the model or the website look like
- New Zealand data would be mixed in with international data.
- There is no control over who contributes data to GBIF so data quality may be patchy
- No biostatus information (a critical failure)
- The software for providing/consuming taxon concepts within the GBIF network is not yet deployed (a critical current failure)
- The web portal software for querying/reporting taxon concepts is not yet fully developed (a critical current failure)

8.2 Re-use of international taxonomic model

In this approach an internationally developed model such as the Berlin Taxonomic Information Model is used but the NZOR database is developed and managed within New Zealand.

Component	Approximate cost
Modification of Berlin model	$10,000
Development of database based on Berlin model	$80,000
Development of data management tools	$105,000
Development of NZOR data administration tool	$130,000
Development of web site application	$150,000
Data user tools from scratch	$240,000
Indicative Total	**$715,000**

Advantages:

- Local control over data quality

Disadvantages:

- High cost, lost opportunity to re-use local components and expertise
- Berlin model is primarily botanical, may not be appropriate for all-of-taxonomy database
- No local expertise in Berlin Model or in related technology
- Existing data management tools based on the Berlin model are “thin-client” web browser-based and currently inefficient for data providers.

8.3 Re-use of existing New Zealand technology

This approach would be to base NZOR extensively on existing New Zealand technology such as that developed by Landcare Research, with appropriate modifications to meet the specific needs of NZOR.

Component	Approximate cost
Re-use of model underlying existing Landcare systems	$10,000
Re-use of existing Landcare database infrastructure	$60,000
Reuse of modified version of Landcare’s DTF (improved)	$90,000
Data administration tool for NZOR (simple)	$100,000
Modification of existing Landcare websites	$80,000
Data users tools based on existing Landcare tools	$120,000
Indicative Total	**$460,000**

Advantages:

- New Zealand gets exactly what it requires (not reliant on overseas platforms, none of which are provided or maintained as commercial systems)
- Local control over data content and quality

- Biostatus information can be made available in required form
- DTF data management tool is “thick-client” and efficient.
- Local knowledge of technology facilitating development and maintenance
- Local development of NZOR by groups already involved in international work ensures international systems and standards are informed by the needs of New Zealand.

Disadvantages:

- Requirement for on-going locally supported maintenance of DTF.

8.4 Development of technology from scratch

In this option, all components for the NZOR technology are built from scratch, without re-using existing models or technology.

Component	Approximate cost
Development of an entirely new model	$25,000
Development of an entirely new database	$90,000
Development of data management tools	$1,020,000
Development of NZOR data administration tool	$130,000
Development of web site application	$140,000
Data user tools from scratch	$240,000
Indicative Total	**$1,645,000**

Advantages:

- New Zealand gets exactly what it requires (not reliant on overseas)
- Local control over data quality
- Biostatus information can be made available

Disadvantages:

- High cost, lost opportunity to re-use components and expertise

NB. The estimated cost to develop a new data management tool capable of implementing the vision of NZOR as described in this document may appear high, however it is realistic. The existing DTF and supporting systems developed by Landcare Research represent an investment of 6 FTE years of work from programmers and analysts. This has been funded both from FRST funding aligned with the Nationally Significant Databases, and significant internal investment from Landcare Research through Non Specific Output Funding (NSOF – now Capability Funding). Discussions with colleagues developing the Berlin Model suggest an equivalent investment funded through EU framework infrastructure projects.

8.5 Summary

There are no existing international systems that currently satisfy the full range of end-user identified needs. Re-use of existing New Zealand technology appears to give the most cost effective option, while still ensuring that NZOR provides the concept based, authoritative taxonomic system that is required.

9 Implementation

There are three major of streams of work within the NZOR project. Each of these has initial setup components, and ongoing components. The workstreams are:

1. Technical Platform
2. Data Content
3. Governance

The implementation process will run in a number of stages, each stage including components from each of the workstreams.

9.1 Technical Platform

This workstream involves providing and maintaining the database and software technology to run the NZOR system. It includes enhancing the available software frameworks (as discussed in section 8) to fit the exact requirements for NZOR, developing the NZOR web site, developing web services for direct interaction between NZOR and systems maintained by data users and data providers.

In addition, every software system requires maintenance. This includes fixing errors in the software, managing and fixing hardware, and supporting users in accessing the system.

Specific components are listed as follows.

9.1.1 NZOR Platform

- Platform hardware, database software based on industry standards
- NZOR database implementation, including communication and rules for harvesting/integrating provider data
- Tools for supporting data quality and administration of NZOR

9.1.2 Data Management tools and services supporting providers

- Data providers will require a tool to manage their contribution of NZOR. Current candidates include Landcare Research's DTF, Berlin Model and K-Emu.
- The service by which data are incorporated into NZOR will use an agreed data transfer mechanism based on the international Taxon Concept Schema data exchange standard.
- Some data providers may already have tools to manage their contribution to NZOR that are appropriate to their applications, processes and technology platforms. NZOR will not be dependent on providers using exactly the same technology platforms as those on which NZOR is implemented. Within the NZOR work programme, platform independent connectors will be developed and key data providers will be supported in integrating these with their own taxonomic data management systems.

9.1.3 Tools for end users to interrogate and report from NZOR

End users will require a number of different kinds of tools to access NZOR data. For some users this will be as simple as accessing the NZOR web site and looking up a name, or perhaps downloading a spreadsheet with a list of names. For other users more sophisticated tools will be needed. These include client side tools for maintaining a local cache, establishing the initial link/ correspondence between local datasets and NZOR, subsequent data integrity analysis and reports.

9.1.4 Training and support for both data providers and data users

The reasons why NZOR is not a simple list of names and requires a level of technical complexity that is not often encountered in information systems for managing taxonomic data, or in end user systems such as collection/observation databases have been outlined earlier in this document. Both data providers and data users will therefore require a level of support from the developers and expert users of NZOR to enable them to provide and consume NZOR information appropriately. This workstream, also referred to as provider and end user 'uptake' provides this training and support.

9.2 Data Content

Of the three workstreams, getting data into the system is the largest. The priority order for initially incorporating these data is:

1. Nomenclatural data
2. Primary taxonomic data (i.e. data supporting 'active' concepts, including synonyms)
3. Biostatus data
4. Additional taxonomic data, including alternate concepts, vernacular names, concept correlations

These priorities span the following activities:

- Harvesting and integrating existing digital sources of NZOR data. Currently identified sources are the data available from NIWA, Landcare Research, and Te Papa's systems supplemented by data from global sources (such as Species 2000/Catalogue of Life) relevant to New Zealand and of known provenance and quality.
- Identifying critical NZOR data gaps not represented by digital sources and identifying priorities for digitisation by data providers. This would include sources such as publications, specimen collections and surveys.
- Engage key end-users to link to NZOR. Currently identified key agencies include DOC and MAF BNZ.

It should be noted that funding is being sought from TFBIS to establish NZOR with baseline data provided by key agencies, and demonstrate proof of concept through uptake by key end users. On- going funding to support filling of priority gaps identified by end users should come from other sources (e.g. specific operational funding).

9.2.1 Data Quality Management

A registration process for data providers which will require them to demonstrate the ability to provide and maintain appropriate data quality is recommended.

NZOR tools will provide a level of automatic checking of data for internal nomenclatural and taxonomic consistency, and adherence to standards

In both the implementation and post-implementation phase it is likely there will be a probable need for an Administration/Editorial function. This role will act as the primary point of contact for providers and users. The role will include following up issues of data quality with the System Governance Team for action.

9.3 Governance

In this context governance refers to the process by which strategic, financial, and data quality decisions are made in regard to NZOR. This involves:

9.3.1 Implementation Project governance

It is recommended that a small steering group with across agency representation (as used during the scope phase) be confirmed as a project steering committee to provide governance for the implementation project. This steering group would have responsibility for major decisions and guidance on how the project is run including monitoring budget expenditure, major decisions on functionality and technology, communications with stakeholders, and reporting. It is distinct from day to day project management which is the responsibility of the lead agency contracted by TFBIS.

NB. The development of a communications plan was within the terms of reference for this scope, however this task has been deferred to the first part of the implementation phase due to dependencies on final technology selection, decisions on implementation timings, and confirmation of the terms of reference for the project steering committee.

9.3.2 On-going System governance

Once the system is established and the initial development and set up stages are complete there will need to be an ongoing process for making major decisions in regard to NZOR. This should be a steering group with multi agency representation, ideally from the key providers, end users and funders. This is likely to have a very similar composition to the project steering group, and will take over responsibilities at the completion of the formal implementation project. A specific task of this group is overseeing the on-going approval and registration of data providers.

9.3.3 Data governance

In the preferred model data quality would be determined by data providers for whom on-going registration is approved by the System Governance team. Where there is conflict in data between data providers, or a provider fails to address quality concerns expressed by end users, then this will be referred to the System Governance team for resolution. This team may seek opinion from appropriate experts to support their decision making. This workstream is explored in detail in section 5 above.

9.4 Implementation Process

Given the varying states of data available, the costs of data acquisition, and the varied ways and levels through which providers and end users are expected to want to connect with NZOR, a 'big bang' approach to implementation is not recommended. Rather, NZOR should be run as an incremental project to maximise provider and end user engagement, manage expectations, and manage risks.

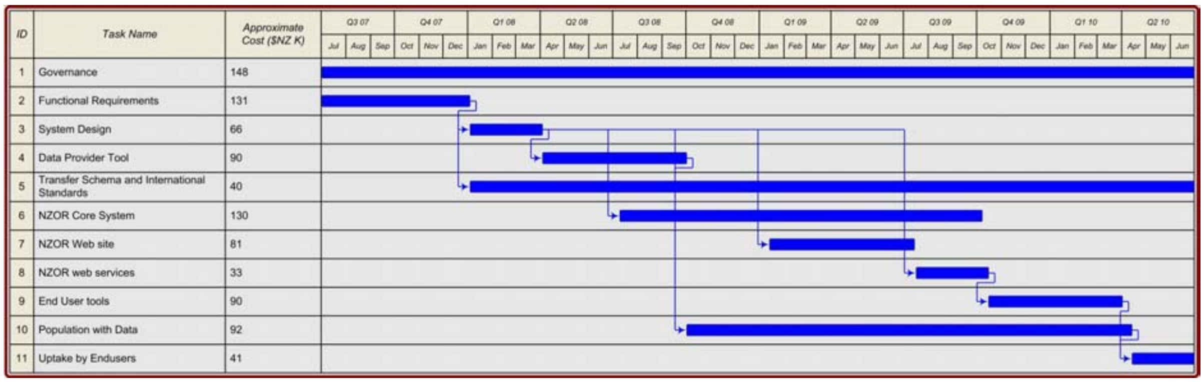
The project will be run over a number of stages. Each stage will incorporate components from each of the work streams outlined above.

Work Stream	Stage One	Stage Two	Stage Three
Technology	Requirements analysis. Technical Design Required enhancements to TCS data exchange and data management protocol	Implement NZOR Platform Develop data provider tools Develop NZOR website prototype	Develop data user web service tools Public release NZOR website
Data		Populate NZOR with electronic data Prepare other baseline datasets	Populate NZOR with digitised baseline data Perform gap analysis
Governance and Administration	Establish System Governance Team Development of operational guidelines Initiation of training programmes Communications	Registration of data providers Training program Analysis of costs and strategies for ongoing maintenance Communications	Strategy developed for ongoing maintenance. Engagement of end users for uptake Communications

10 Costs

It should be noted that costs identified in section 9 Technology above refer specifically to the technology components of NZOR. The costs associated with data population, governance, and uptake are additional to whatever technology platform is adopted. The complete costs for the implementation of NZOR are itemised in this section, including those for data population, governance, uptake and costs for enhancement of the preferred technology platform.

The principle components within the workstreams indicated in Section 10 are identified in the following GANNT chart. This provides indicative total costs and duration of each component over the three years of the implementation phase of NZOR. This analysis assumes that existing New Zealand developed technology is adopted, as identified in section 9.



These costs (in $1,000s) may be broken down by year:

Component	Year 1	Year 2	Year 3	Totals
Governance	50	50	50	**150**
Functional Requirements	131			**131**
System Design	66			**66**
Data Provider Tool	45	45		**90**
Transfer Schema & International Standards	10	15	15	**40**
NZOR Core Systems		100	30	**130**
NZOR Website		81		**81**
NZOR Web services			33	**33**
End User Tools			90	**90**
Population with Data		46	46	**92**
Training, support and uptake by End Users			41	**41**
Totals	**302**	**337**	**305**	**944**

Data mobilisation and incorporation into NZOR during this implementation phase includes both terrestrial and marine organisms where the latter constitute approximately one third of the total. There is no practical justification in separating these two data streams and significant benefit to NZOR from doing both. Digital data will be mobilized from the existing resources of Landcare Research, Te Papa, NIWA and Species 2000/COL. Costs for mobilizing these data are associated with transformation services to provide the data in a form compatible with the NZOR model. Whether data refers to marine or terrestrial organism does not impact on the required effort. Non digital baseline data will be extracted from the Species 2000 New Zealand publications. Here the concentration of work is around the development of routines to automate the process of recognising components within the structured text documents relating to taxon concepts and parsing these data into an NZOR compatible form. Again, the type of data do not impact on the required effort. There will inevitably be a small amount of manual processing of data where costs could potentially be divided between the marine and terrestrial sectors but this is considered to be a relatively minor component.

The original NZOR proposal estimated costs of $700k associated with Phase 1 of NZOR to be completed over 2 years. That proposal also identified additional Phase 2 components requiring further clarification and negotiation of funding, and which extended the project to 4 years in total. Phase 1 components identified in that submission are now fully incorporated into the identified workstreams in this document. Further analysis carried out during this scoping study has now more clearly identified necessary components of Phase 2. This implementation phase of NZOR can be undertaken within a 3 year timescale. The costs identified in this analysis are broadly in-line with those identified in the original TFBIS proposal.

This three year implementation will provide the core systems, services and data for NZOR to become operational. Specifically, it will provide:

- A robust core technology platform compatible with international standards, interoperable with existing and newly emerging global networks.
- A system populated with baseline data and integrating the most current taxon concept information from a number of key national and international data providers.
- Tools supporting data providers, including those in key agencies explicitly included in the implementation phase, and also available to future NZOR data providers.
- A gap and priority analysis for further building NZOR content through contributions from identified additional providers.
- Tools supporting end users to adopt and integrate NZOR information and services into local systems.
- Web based access to current taxon concept information providing functionality for queries, reports and complex taxon concept navigation.
- Support and training for both data providers and end users
- A functional governance structure with responsibility for ensuring quality of service
- An operational system that is delivering benefit to key end users, demonstrated through uptake (specifically within MAF-BNZ, DOC, MFish and ERMA)

Once the system is operational it will require ongoing maintenance, additional to the costs defined above. It seems likely that the ongoing maintenance will be supported by MoRST funding for science ‘backbone’ infrastructure, although this is not yet guaranteed. Ongoing funding for addition of new data is expected to come from a range of sources including FRST funded science outputs, and operational funding from agencies requiring those data.

11 References

Biosecurity Working Group, 2003: Databases in the Biosecurity Environment. Internal MAF report, 2003.

Carver, J.; Davidson, C.; Pyle, E.; Munro, M. 2005: Research e-Data Saving and Sharing Workshop Summary. Proceedings of a Workshop held in Wellington on 15 June 2005. Prepared for the Ministry of Research, Science and Technology, unpublished report available on request from MoRST.

Clinehans, S. 2006: Improving the information available for determining new organism status of plants. Draft report prepared for the Sponsor and Steering Committee of the Terrestrial and Freshwater Biodiversity Information System (TFBIS) Programme. MAF, Wellington, 17 August 2006

Cooper, J.; Carver, J. 2005: A Survey of Regional Council Biodatabases. Report prepared for the Sponsor and Steering Committee of the Terrestrial and Freshwater Biodiversity Information System (TFBIS) Programme. Landcare Research, 14 July 2005

Entwisle, T.J. & P.H. Weston. 2005. Majority rules, when systematists disagree. Aust. Systematic Botany 18:1-6.

Glenny, D. 2004: A revision of the genus Gentianella in New Zealand. New Zealand Journal of Botany 42(3): 361-530.

Gordon, D.P. (Ed.) 2007: In press. The New Zealand Inventory of Biodiversity. Volume 1. Kingdom Animalia: Radiata, Lophotrochozoa, and Deuterostomia. Canterbury University Press, Christchurch.

Gordon, D.P. (Ed.) In prep. The New Zealand Inventory of Biodiversity. Volume 2. Kingdom Animalia: Chaetognatha, Ecdysozoa, and Ichnofossils. Canterbury University Press, Christchurch.

Gordon, D.P. (Ed.) In prep. The New Zealand Inventory of Biodiversity. Volume 3. Kingdoms Bacteria, Protozoa, Chromista, Plantae, and Fungi. Canterbury University Press, Christchurch.

MAF. 2006: A Biosecurity Science, Research And Technology Strategy For New Zealand. Draft strategy for circulation. MAF, Wellington, 8 November 2006

Appendices

Appendix 2. Numbers of Taxa

The estimated number of taxa for known native and naturalised New Zealand organisms are as follows:

Taxa	Marine	Terrestrial	Freshwater	Total	% known & undescribed	Estimated undiscovered species
Bacteria	400	300	330	**1,030**	1%	unknown
Protozoa	1,660	670	460	**2,790**	9%	6690–7570
Chromista (brown and golden algae, aquatic moulds etc)	850	160	920	**1,930**	7.5%	1295–1305
Plants	630	5,160	1,110	**6,900**	2%	265
Fungi	3	6780	250	**7033**	10%	14,960
Animals	12,640	20,340	2,410	**35,390**	23%	60,500 – 89,800
Chordates (Fish, Amphibians, Birds, Mammals)	1,310	260	100	**1,670**	13%	830
Molluscs	3,660	940	90	**4,690**	38%	430
Arthropods (insects, spiders etc) &	2,680	18,250	1,480	**22,410**	20%	22,880
Nematodes (roundworms)	170	500	20	**690**	9%	30,300 - 60,000
Other (sponges, corals, worms etc)	4,990	890	740	**6,620**	29%	11,250
Total	**16,183**	**33,410**	**5,480**	**55,073**		**83,710 - 113,900** (excluding bacteria)

These figures have been taken from the Species 2000 New Zealand set of publications. It should be noted that to make the table easier to read taxa numbers have been rounded to the nearest ten. Taxa have been grouped by Kingdom, and grouped further (somewhat arbitrarily) within Animals to illustrate where the large numbers lie.

NB. This does not include 25,000 – 40,000 cultivated plant names which will be an important component for many end-users, nor does it include the number of animals in captivity (zoos, aquaria, agriculture, etc).

Appendix 3. MAF Systems

A short summary of the main MAF Biosecurity information systems that would make use of NZOR follows.

Biosecurity Risk Analysis Database (BRAD)

BRAD contains information on plant pest species (i.e. organisms that are pests to plants) and the relationship between these and commodity species, as well as the countries or regions of origin for the pest species. It presently contains approximately 32,000 pest species and 800 commodity species.

It is used to generate formatted pest lists for inclusion in Import health standards. These lists provide information on the organisms that the overseas country must ensure are not present with the commodity prior to export to New Zealand. Information is also extracted and imported into BORIC for border control staff.

BRAD stores scientific and vernacular names, synonyms, and some taxonomic information.

BRAD has recently been developed as a replacement for the PAQIS system. In the future there is a desire to have it directly linked with the PBI and the UOR. It also has some potential data overlaps and potential linkages with Quancargo, Ecert, and PMS (Pest Management System).

Plant Biosecurity Index (PBI)

The PBI is a tool used by Biosecurity New Zealand, MAF to manage the importation of terrestrial and freshwater plant species within the higher plant kingdom. The database lists plant species recognised as being present in New Zealand before 29 July 1998. These species are considered to be *not* "new organisms" for the purposes of the Hazardous Substances and New Organisms Act 1996 (the HSNO Act), and thus may be imported to New Zealand without approval under that Act.[2] Plant species approved for importation since the HSNO Act came into effect are also listed in the PBI as are species that are prohibited for importation (unwanted organisms, pest plants etc.). ERMA makes decisions about what can be included on the PBI.

MAF is currently planning the redevelopment of this system in response to some concerns about its incompleteness, and accuracy and level of detail of taxonomic plant names listed.

There are approximately 27,000 taxa in the PBI. There are estimated to be 15,000 introduced cultivated plants that are not listed in the PBI because formal collections and identifications of these taxa have not been done.

Plant Pest Information Network (PPIN)

This is a key MAF database, providing a list of which plant pest species are, or have been established (then eradicated) in New Zealand. The database is managed by the National Plant Pest Reference Laboratory on behalf of the Biosecurity Authority. As the database is widely used to address queries regarding species present in New Zealand, some of which can have far-reaching implications on for example exports, data integrity is critical. It therefore contains independently verified records. The data is mainly provided by the National Plant Pests Reference Laboratory, but is also supplemented with verified records from some of the CRIs.

---

[2] Under the HSNO Act "new organisms" are defined as plants, animals, micro-organisms, and genetically modified organisms (GMO's) that were not present in New Zealand prior to 29 July 1998 (HSNO, 1996, s 1, 2).

The database was established in 1995 with an initial plant biosecurity focus. Since then it has been expanded to include records on forestry hosts. Data is sourced from actual surveillance data, but over time historical records (literature and internal reports) are being added to expand it. The database provides information about established species, including temporal and spatial data, as well as host information and information on the bio-region where it was encountered. It currently contains more than 27,000 records, with more than 12,000 records still being verified, or pending entry.

Unwanted Organisms Register (UOR)

The unwanted organisms register is a requirement of the Biosecurity Act 1993. It is a register of organisms that have been determined unwanted by Chief Technical Officers of government departments with biosecurity interests. It also contains organisms declined importation by the Environmental Risk Management Authority (ERMA NZ), and organisms listed in the second schedule of the Hazardous Substances and New Organisms Act 1996.

Organisms are recorded in categories according to the type of “unwantedness” (e.g. prohibited, regulated, unwanted, etc.). Note that organisms can belong to more than one category.

The register currently contains around 15,000 taxa. Plants Biosecurity place organisms on the UOR from BRAD based on risk assessments, so the large majority of the taxa on the UOR overlap with those on BRAD.

Biosecurity Organisms Register for Imported Commodities (BORIC)

BORIC records organisms that may be associated with plants or plant products that are imported into New Zealand. The quarantine status for each species is indicated i.e. regulated or non-regulated. The list provides provides border staff with information on how to treat pests found on specific commodities. This list is updated frequently, usually at 1-2 week intervals.

Appendix 4. Use Cases

Use Cases 1 Data Loading and maintenance

Use Case 1.1 Registering as a data provider

Priority: Essential

Trigger:

An organisation that has information of interest to NZOR wants to register as a Data Provider organisation

Steps:

1. The Data Provider organisation makes contact with NZOR
2. NZOR and the Data Provider organisation work out what data will be provided, how it needs to be validated and in what agreed format it will arrive.
3. The Data Provider starts to provide data to NZOR.

Example:

A research organisation contact the staff at NZOR to find out what information they have that NZOR could use. They discuss the areas of expertise of the research organisation (so that there isn't duplication of effort and to ensure NZOR data quality) and work out what format the data should be provided in, and how it should be provided. The question of who will check the data to ensure data quality is discussed. When this is all worked out and service level agreements or similar are signed, the Service Provider organisation starts to provide data to NZOR.

Alternative Example:

A research organisation wants to register a resource as available to NZOR. They send web services messages to NZOR to gain information about the format of data that NZOR requires and the manner in which it should be provided. If this is agreeable to them, they send a web services message that attempts to register the organisation as a Data Provider with NZOR. If their registration is acceptable to NZOR then the organisation is registered as an NZOR Data Provider and can start to provide data to NZOR.

Use Case 1.2 Loading a set of data

Priority: To be Decided

Trigger:

A registered Data Provider organisation has a list of species that they are expert in and they want to load them into NZOR so that NZOR can contain a complete list for that taxonomic group.

Steps:

1. The Data Provider organisation prepares the information required for loading into NZOR
2. The data is checked for quality and approved for loading.
3. The Data Provider passes the information to NZOR

4. NZOR reads and checks the new information and stores it in its database.
5. NZOR allocates new identifiers within NZOR for the new taxon concepts
6. NZOR makes the new data available for users to access.

Example:

A research organisation that has registered as an NZOR Data Provider holds a resource in their local databases. NZOR accesses this resource and harvests the agreed information, validates it and stores it in its databases, making the information available to other users.

Use Case 1.3 Reporting an error

Priority: To be Decided

Trigger:

A Data User has pointed out an possible error (for example a spelling error) in information on NZOR and the Data Provider needs to fix it.

Steps:

1. The Data User reports an error to NZOR
2. NZOR identifies the potential error as being within the data provided by a specific Data Provider
3. NZOR informs the Data Provider of the problem within a specific taxon record
4. The Data Provider organisation checks and, if necessary, corrects the data in their database, and completes the necessary verification processes
5. The Data Provider passes the new information to NZOR
6. NZOR reads and checks the new information and stores it in its database updating the old NZOR record.
7. NZOR makes the new data available for users to access.

Use Case 1.4 Updating Data

Priority: Essential

Trigger:

A registered data provider is reading journals relating to his or her field of expertise and finds that a taxon has been re-named. They realise that this information needs to be added to NZOR so that users will be able to find this taxon under either the old or the new name, and be aware that the correct name has been changed.

Steps:

1. The Data Provider organisation documents the information required for loading into NZOR, and completes necessary verification processes
2. The Data Provider passes the new information to NZOR, with an NZOR record identifier
3. NZOR reads and checks the new information
4. NZOR stores the changed data as new records, and stores information about who made the change, who reviewed it and when the information was changed.
5. NZOR makes the new data available for users to access.

Example:

A research organisation that has registered as an NZOR Data Provider holds a resource in their local databases. When the expert finds that a species in their area of expertise has had a name change they change it in their database. When NZOR next accesses this resource and harvests the agreed information it notices the change, validates it and stores it in its databases, making the information available to other users.

Use Cases 2 Data use

Use Cases 2.1 Searching by taxon

Use Case 2.1.3 Taxon not found in New Zealand

Priority: Essential

Trigger:

A data user wants to know if a particular taxon is found in New Zealand. This particular taxon is known to NZOR but is not believed to be found in New Zealand.

Steps:

1) The data user logs into NZOR
2) The user selects the "Taxon query" option
3) The user enters the information they have about the taxon. They may enter a scientific name, a vernacular name, or an NZOR identifier or several of these.
4) NZOR searches its database and finds that there is a record for this taxon. NZOR displays basic information about the taxon.
5) The user confirms that this is the taxon that is of interest and asks for further information
6) NZOR displays more detailed information about the taxon, including what is known about its presence in New Zealand. The record shows that the taxon is absent in the New Zealand region. More detailed information about the taxon shows that it was once thought to have been found in New Zealand, but that occurrence is now thought to have been a different taxon.

Example:

The user searches under the scientific name "*Dissostichus mawsoni*". NZOR searches its database and finds that there is a record for this species, Antarctic toothfish. The record states that the species is not believed to be found in the New Zealand EEZ (this is a biostatus record of "Absent" in the region "New Zealand EEZ" with a "Geo-Schema" of "NZ Political".

Use Case 2.1.5 Taxon found under different name

Priority: Essential

Trigger:

A data user wants to know if a particular taxon is found in New Zealand. This particular taxon is found in New Zealand, but is properly known by a different name.

Steps:

1) The data user logs into NZOR
2) The user selects the “Taxon query” option
3) The user enters the information they have about the taxon. They may enter a scientific name, vernacular name, or an NZOR identifier or several of these.
4) NZOR searches its database and finds that there is a record for a taxon that was once known by that scientific name, but is now properly known by a different name. NZOR displays basic information about the taxon, including historic names.
5) The user confirms that this is the taxon that is of interest and asks for further information
6) NZOR displays more detailed information about the taxon, including what is known about its presence in New Zealand

Example:

The user searches under the scientific name “*Gracilaria secundata*” and the common name “Sea Moss”. NZOR searches its database and finds that there is a record for a species that was once known as *Gracilaria secundata*, but is now properly known as *Gracilaria chilensis*. NZOR displays basic information about the species *Gracilaria chilensis*, including historic names, with reference to *Gracilaria secundata*.

Use Case 2.1.7 User uncertain of proper spelling of name

Priority: To be decided

Trigger:

A data user wants to know if a particular taxon is found in New Zealand. They couldn’t find it under the spelling they thought they knew, so they search for it again using wildcards.

Steps:

1) The data user logs into NZOR
2) The user selects the “Taxon query” option
3) The user enters the information they have about the taxon. They may enter a scientific name, a vernacular name, or an NZOR identifier or several of these.
4) NZOR searches its database and cannot find a taxon that matches that name, and suggests ways for the user to rephrase the query in a more generic way.
5) The user is sure about part of the scientific name, so queries again using a wildcard for the other part of the name.
6) NZOR displays the record for the taxon that matches the query.
7) The user confirms that this is the taxon that is of interest and asks for further information
8) NZOR displays more detailed information about the taxon, including what is known about its presence in New Zealand

Example:

The user searches under the scientific name “*Gracilia Chilensis*” and the common name “Sea Moss”. NZOR searches its database and finds that there is no record for this species. The user searches again under the name “*Grac* Chilensis*”. NZOR finds a record for a species *Gracilaria Chilensis* that is also known as Sea Moss. NZOR displays basic information about the species *Gracilaria Chilensis*.

Use Case 2.1.8 Several taxa match

Priority: To be decided

Trigger:

A data user wants to know if a particular taxon is found in New Zealand. There are several taxa that match, some of which are more likely than others to be the one the user is interested in.

Steps:

1) The data user logs into NZOR
2) The user selects the "Taxon query" option
3) The user enters the information they have about the taxon. They may enter a scientific name or vernacular name.
4) NZOR searches its database and finds that there are several taxa that match this query, including taxa which were once known by the same scientific name. NZOR displays basic information about each of the candidate taxa, including the historic names, and when they applied.
5) By looking at the information, the user can get an idea about which of the taxa is most likely to be the one they had in mind.
6) NZOR displays more detailed information about the taxon, including what is known about its presence in New Zealand

Example:

DOC are using NZOR to verify observation data. The DOC user searches under a scientific name. There are a number of species that are known, or were once known by the same scientific name. NZOR displays all of these species. The DOC user does not have the expertise to decide which of the species was intended, so "parks" the observation record until the source of the observation record can be contacted and the issue resolved. Once this is done they save the NZOR identifier with the observation record so that future users of the observation record will know exactly what was meant.

Use Case 2.1.10 Disputed or unresolved taxon concept

Priority: To be decided

Trigger:

A data user wants to know if a particular taxon is found in New Zealand. The name they use relates to a taxon that is thought by some taxonomists to exist in its own right, but the most authoritative taxonomic opinion is that it does not exist separately. Or NZOR may contain several taxon concepts, but none have been indicated as the active concept.

Steps:

1) The data user logs into NZOR
2) The user selects the "Taxon query" or "NZOR LSID query" option
3) The user enters the information they have about the taxon. They may enter a scientific name, vernacular name, or an NZOR identifier.
4) NZOR searches its database a taxon which matches the criteria.

5) NZOR displays information about this taxon record with a flag to show that it is not a taxon record in the authoritative taxonomy.
6) NZOR also displays information about the taxon that the name refers to does not have a single active concept by indicating that there are multiple taxon concepts available.

Example:

The user searches under the name “South Island xyz”. The most authoritative taxonomy suggests that this is not a separate species, and should be included within “North Island xyz”. However there is one taxonomic opinion that suggests that it is a separate species. When the user searches for “South Island xyz” NZOR will return the record for “North Island xyz” with a note that “South Island xyz” is another name sometimes applied to this species and another record showing that the “South Island xyz” does not exist in its own right in the authoritative taxonomy, although there is debate about this.

Use case 2.2 Detailed information

Use Case 2.2.1 User wants reference information

Priority: To be decided

Trigger:

A data user wants to know more about a specific taxon, for example why it has been classified as it has been.

Steps:

1) The data user logs into NZOR
2) The user selects the “Taxon query” option
3) The user enters the information they have about the taxon. They may enter a scientific name, a vernacular name, or an NZOR identifier or several of these.
4) NZOR searches its database and finds a match and gives basic information about the taxon so that the user can confirm that it is the one of interest.
5) The user clicks on the record for further information.
6) NZOR displays further information about the taxon, including a list of publications relating to the allocation of this name to this taxon.

Example:

The user enters the name xyz. NZOR finds a record for the species xyz and displays basic information. The user drills down to more detailed information about the species and finds the journal reference for where the nomenclature was published.

Use Case 2.2.2 User wants biostatus information

Priority: To be decided

Trigger:

A data user wants to know more about a specific taxon, for example, whether it is present in New Zealand.

Steps:

1) The data user logs into NZOR

2) The user selects the “Taxon query” option
3) The user enters the information they have about the taxon. They may enter a scientific name, a vernacular name, or an NZOR identifier or several of these.
4) NZOR searches its database and finds a match and gives basic information about the taxon so that the user can confirm that it is the one of interest.
5) The user clicks on the record for further information.
6) NZOR displays further information about the taxon, including a list of biostatus records for this taxon, setting out whether it is present or absent in one or more areas in New Zealand.

Example:

The user enters the name *Orosius argentatus*. NZOR finds a record for the species *Orosius argentatus* (Insect - Homoptera: Cicadellidae) and displays basic information. The user drills down to more detailed information about the species and finds a number of biostatus records.

Geo-Region	**Geo-Schema**	**Occurrence**
South Island	Geographic	Absent
North Island	Geographic	Absent
Kermadec Islands	Geographic	Present
New Zealand Territorial	Political	Absent
New Zealand EEZ	Political	Present

The user is able to conclude that for their purpose (for example, MAF’s responsibility is the inhabited area of New Zealand), the species is Absent. However for some other purpose (for example, Ministry of Fisheries responsibility extends to the New Zealand EEZ), the species is Present.

Use Case 2.2.4 User wants a history of name changes

Priority: To be decided

Trigger:

A data user wants to understand how the names for a particular taxon have changed over time.

Steps:

1) The data user logs into NZOR
2) The user selects the “Taxon query” option
3) The user enters the information they have about the taxon. They may enter a scientific name, a vernacular name, or an NZOR identifier or several of these.
4) NZOR searches its database and finds a match and gives basic information about the taxon so that the user can confirm that it is the one of interest.

5) The user clicks on the record for further information. NZOR displays further information about the taxon, including a complete history of different names that this taxon concept has had historically.

Example:

The organisms referred to as *Cuscuta campestris* were correctly known as the *Cuscuta pentagona* var. pentagona until some date when the *Cuscuta pentagona* var. pentagona taxon concept was split into two taxon concepts – the *Cuscuta arvensis* and *Cuscuta campestris*.

Use Case 2.4 User wants to know the confidence level of data for a record in NZOR

Priority: To be decided

Trigger:

A data user needs to make a decision based on the information in NZOR and therefore wants to know how certain the information is, so that they can back up the decision. In fact this taxon has only been very poorly identified with this name (the taxonomy is considered outdated but has not been re-done).

Steps:

1) The data user logs into NZOR
2) The user selects the "Taxon query" option
3) The user enters the information they have about the taxon.
4) NZOR finds a match to a taxon
5) The user drills down to find out further information about the taxon and finds that there is a flag indicating that the identification of this taxon to this name, is considered "Very poor" because the taxonomy is considered outdated.
6) The user takes this "Very poor" rating into account in making the decision.

Example:

The user enters the name xyz. NZOR finds a record for the species xyz and displays basic information that the species is known by a particular scientific name. The user drills down to further information about the species and finds that the identification of this species to this name is considered "very poor" because the taxonomy is outdated. The user decides that instead of making a decision at this time, they will ask for some additional research to be done in this area, so that the identification can be confirmed or corrected.

Use Case 2.5 User wants to print out a list of all taxa that meet some criterion

Priority: To be decided

Trigger:

A data user wants a list of all species that meet some criterion, for example, all taxa that are recorded in NZOR as being present in New Zealand.

Steps:

1) The data user logs into NZOR
2) The user selects the "Advanced query" option

3) The user enters their criteria.
4) The user indicates which fields they are interested in for each record returned.
5) NZOR searches its database to find all records that meet the criteria
6) NZOR produces a report which for all records that meet the criteria, gives the information that the user requested.

Example:

The user wants a list of the current correct species name for all fungi species that NZOR has listed as being "Present" in the New Zealand geographical area. They select the Advanced query option and enter their criteria as being that the biostatus record for the geographical region of "New Zealand" have value "present", that the record be for a species (rather than a family) and that the record fall within the taxonomic group of fungi. NZOR queries its database and returns a list with one record for each of these fungi species and reports the preferred name for each of these taxon concepts.

Another Example: – DOC do not choose to access NZOR in a live link when they want to query NZOR information, but prefer to keep a cache of information such as species names and NZOR identifiers within DOC, and then keep this list updated on a daily or weekly basis.

Another Example:- a user wants a list of all the taxon concept records that refer to a particular published review paper.

Use Case 2.6 User wants a count of all taxa that meet some criterion

Priority: To be decided

Trigger:

A data user wants a count of all taxa that meet some criterion, for example, all taxa that are recorded in NZOR as being present in New Zealand.

Steps:

1) The data user logs into NZOR
2) The user selects the "Advanced query" option
3) The user enters their criteria.
4) The user indicates that they are only interested in a count of records.
5) NZOR searches its database to find all records that meet the criteria
6) NZOR produces a report that provides a count of all records that meet the criteria.

Example:

The user wants to know how many fungi species that NZOR has listed as being "Present" in the New Zealand geographical area. They select the Advanced query option and enter their criteria as being that the biostatus record for the geographical region of "New Zealand" have value "present", that the record be for a species (rather than a family) and that the record fall within the taxonomic group of fungi. NZOR queries its database and returns a count of these fungi species.

Another Example:- a user wants a list of all the taxon concept records that refer to a particular published review paper.

Use Case 2.7 Parent relationships

Priority: To be decided

Trigger:

A data user wants to know about the taxon that is a parent to another taxon, that is, that contains this taxon within it. For example they want to know more about the Genus that a particular species belongs to.

Steps:

1) The data user logs into NZOR
2) The user selects the "Taxon query" option
3) The user enters the information they have about the taxon. They may enter both a scientific name and a common name.
4) NZOR searches its database and finds that there is a record for this taxon. NZOR displays basic information about the taxon.
5) The user confirms that this is the taxon that is of interest and asks for further information
6) Amongst other information, NZOR displays information showing a link to the parent taxon that contains this taxon within it.
7) The user selects the option asking for more information about this parent taxon.
8) NZOR shows information about the parent taxon.

Example:

The user searches for hoki under the scientific name "*Macruronus novaezelandiae*". NZOR searches its database and finds that there is a record for this species and gives information that the parent record for this taxon concept is the Family named Merlucciidae. The user asks for more information about this Family, and can then find that there are several other species in this Family in New Zealand, including hake *Merluccius australis*.

Use Case 2.8 Child relationships

Priority: To be decided

Trigger:

A data user wants to know about the taxa that are thought to be within this taxon (that is, are contained within this taxon). For example they want to know more about all the species that belong to this Genus.

Steps:

1) The data user logs into NZOR
2) The user selects the "Taxon query" option
3) The user enters the information they have about the taxon. They may enter both a scientific name and a common name.
4) NZOR searches its database and finds that there is a record for this taxon. NZOR displays basic information about the taxon.
5) The user confirms that this is the taxon that is of interest and asks for further information

6) Amongst other information, NZOR displays information showing a link to the child taxa that are contained within this taxon.
7) The user selects the option asking for more information about one of these child taxa.
8) NZOR shows information about the child taxon.

Example:

A user is looking at the record for the Family Merlucciidae, wanting to know whether there are any species in this Family in New Zealand. The detailed information for this Family includes a number of species, including the taxon with the preferred (current correct) species name of "*Macruronus novaezelandiae*". The user requests further information about this species and finds that it is present in New Zealand, going under the common name of hoki.

Use Case 2.9 Search on literature

Priority: To be decided

Trigger:

A data user wants to search the literature or bibliographic records independently of the taxon concept records. For example, they want to know what papers there are in NZOR that relate to their area of interest.

Steps:

1) The data user logs into NZOR
2) The user selects the "Literature query" option
3) The user specifies what papers or documents in the referenced literature they are interested in
4) The user specifies what information they would like about the papers or documents
5) NZOR searches its literature records to find all literature records that meet the criteria.
6) NZOR produces a report indexing all of the relevant literature records.

Example:

A user is interested in all literature that relates to a certain Order of fungi (including the species and other taxa within this Order). The specify what they are interested in to NZOR which searches its database and returns a report containing publication references for all the papers that have been referenced in NZOR (for example nomenclature publication) that are linked to by taxon concept records in this Order.

Use Cases 3 Tools

Use Case 3.1 Login with password

Priority: Essential

Trigger:

A data user wants to access NZOR and see information that is not available to the general public.

Steps:

1) The data user contacts NZOR to register and request a login ID
2) NZOR checks that this request is acceptable
3) NZOR issues a username and an initial password to the registering user
4) The registering user logs in using this initial password and changes their password to something secure
5) The user logs in to NZOR using their own username and password. This gives them certain privileges that are not available to other users.
6) The user accesses records that are not marked as generally available to the public
7) NZOR stores information about which records were accessed and when.

Example:

A user at MFish logs in to NZOR as a guest user but finds that they cannot access the information they need as it has been set up as unavailable to the general public. They apply to NZOR for a username and password and log in again under their own username. Now they can see the information they require.

Use Case 3.2 Information via web services

Priority: Essential

Trigger:

A data user or data user computer system wants to know what information can be made available via NZOR, so that they can properly phrase requests for information.

Steps:

1) The data user contacts NZOR via a web service message.
2) NZOR checks that this request is acceptable
3) NZOR sends a message back via web services that describes the information that NZOR offers and how to access it
4) The user can correctly formulate requests for information.

Example:

MFish want to access NZOR information via web services. The MFish computer sends a web services message like "GetSearchableFields()" and NZOR replies by sending an XML message that sets out all the fields that are available.

Use Case 3.3 Regular updates

Priority: Essential

Trigger:

A Data User organisation wants to keep their own databases up to date, so they want to be kept informed about changes in the NZOR database. However they are only interested in certain groups of species in their area of interest or authority.

Steps:

1. The Data User organisation arranges with NZOR to set up a regular update service (perhaps using technology such as RSS)
2. The Data User specifies a (or more than one) family or genus that is of interest to them.
3. When the NZOR database for a record that falls within this family or genus is changed or deprecated, a web service message is generated describing the change, identifying at least the identifier for the concept that has changed and a flag to identify the type of change.
4. The Data User organisation receives this web service message and they consider the information, making appropriate changes to their database to reflect the change if required.

Example:

The Biodiversity Risk Assessment Database is interested in whether species in certain families or orders are already in New Zealand. They arrange with the NZOR data administrator that they want to get a regular feed of information about changes to species that fall within their area of interest. Every day they receive a web service message in XML that tells them about:

- All new species that have been added to NZOR
- All the species that have been reclassified as “Present” rather than “absent”
- All the species that have been renamed
- All the species that have had the species concept changed (for example, one species split into two or two species combined into one)
- Any species that have been reclassified as “Absent” rather than “Present”

A scientist at BRAD reads and interprets these messages, in conjunction with the detailed information on NZOR to decide whether these changes are ones that should be copied into BRAD, or whether further investigation is required on their part.

The Natural Heritage Management System stores information about observations of New Zealand organisms. The observations are submitted using scientific name which are then matched to an NZOR identifier. NHMS stores the scientific name that the observation was reported under as well as the NZOR identifier for the matching taxon concept. NHMS also stores the preferred (correct current scientific) names for the NZOR identifiers. If a change is made in NZOR that changes these names then NHMS needs to be updated as soon as possible.

Use Case 3.4 Comparing lists of taxa

Priority: To be Decided

Trigger:

A Data User organisation wants to maintain the integrity of their own databases, so they want periodically to check how the information on their database compares with that of on NZOR. However they are only interested in certain groups of species in their area of interest or authority.

Steps:

1. The data user prepares a list of taxon that they are interested in

2. This list is passed to NZOR
3. The list is compared with the NZOR database. NZOR searches for each taxon on the list in turn, to determine whether a unique match for that taxon can be found.
4. A report is generated that specifies for each taxon:
    a. Whether any match could be found
    b. Whether a unique match could be found, and if not why not (for example, there may be several taxon that have been known by this name at different times, and it is not 100% certain which is meant)
    c. If a unique match could be found, the report would include some information about the taxon such as its biostatus (present/absent)
5. The data user checks any taxon where a unique match could be found but the information (eg present/absent) information does not match
6. The data user checks any taxon where a unique match could not be found and decides which of the several possible taxon is relevant
7. The data user checks the taxon where no match could be found, and reports these to NZOR for possible inclusion.

Example:

The Biodiversity Risk Assessment Database is interested in whether species in certain families or orders are already in New Zealand. They pass a list of species to NZOR. NZOR runs a query over the NZOR database and generates a report. This report lists a number of species as not in NZOR. On closer examination some of these turn out to be spelling mistakes in one or other system, which are then corrected. Some turn out to be where NZOR has not been kept up to date in this area and a research programme is initiated. The report lists a number of species as being un-matchable because there are several species in NZOR that have gone by this name. Some of these turn out to be because of misapplications in the distant past, but it is clear which species is being referred to in BRAD. Other species turn out to be less clear cut. The report also lists a number of species where NZOR and BRAD disagree as to whether the species has been found in New Zealand. Although NZOR has a record that one particular species has been found in New Zealand, the data administrator of BRAD decides not to update the record in BRAD until further investigation has been done, because the evidence does not appear to make this certain. There is a greater risk of damage to New Zealand if a species is wrongfully imported into New Zealand on the basis of incorrect information, than if a species is incorrectly banned while the information is clarified.

DOC want to send some data to be loaded into the National Vegetation Survey database. They have been told that they can only use species names that have unique matches to NZOR taxon concepts. They send a list of species names that they intend to use to NZOR and NZOR checks the list. Some of the species names are recognised as unambiguous in NZOR, so they know that these are OK to submit to the NVS. Others are returned as ambiguous and they need to go into NZOR on-line to investigate further. They work out what species concept on NZOR they meant, and use a name (or NZOR taxon concept identifier) that identifies them uniquely.

Use Case 3.5 Usage statistics

Priority: Essential

Trigger:

A Data Provider organisation wants to know how much information about a particular taxonomic group is being used and by whom, so they can work out how to target more work in this area.

Steps:

1. A data provider accesses NZOR (either on-line or via web services)
2. The data provider chooses the "Usage Statistics" option
3. The data provider specifies a (or more than one) family or genus that is of interest to them.
4. NZOR searches its database to generate a report containing information about how often the taxon concept records within this group have been accessed, when and (if available) by whom
5. NZOR passes this report to the data provider.

Use Case 3.6 Data exchange of concepts

Priority: To be Decided

Trigger:

A Data User organisation wants to maintain the integrity of their own databases, so they want periodically to check how the information on their database compares with that of on NZOR. Like NZOR, the organisation maintains a set of taxon concepts, and the NZOR taxon concept identifier is stored alongside the organisations's concept identifier if known.

Steps:

1. The data user prepares a list of taxon concepts that they are interested in, referred to where possible by the NZOR taxon concept identifier and otherwise by the scientific name.
2. This list is passed to NZOR
3. The list is compared with the NZOR database. NZOR searches for each taxon concept on the list in turn.
4. Where the NZOR taxon concept identifier is known, NZOR checks whether there have been any changes to the taxon concept record since the check was last done. Where there have been changes NZOR reports on the change.
5. Where the NZOR taxon concept identifier is not known, NZOR tries to find a match against the scientific name.
6. A report is generated that specifies for each taxon:
    a. Whether any match could be found
    b. Whether a unique match could be found, and if not why not (for example, there may be several taxa that have been known by this name at different times, and NZOR is not 100% certain which is meant)

c. If a unique match could be found, the NZOR taxon concept identifier

7. The organisation updates its database with the NZOR taxon concept identifiers where there was a unique match
8. The data user checks any taxon where a unique match could not be found and decides which of the several possible species is relevant
9. The data user checks the taxa where no match could be found, and reports these to NZOR for possible inclusion.

Example:

The National Vegetation Survey database maintains its own list of taxon concepts. Where possible it stores the NZOR taxon concept identifier as well. The NVS regularly compares its list of taxon concepts with NZORs list to check for discrepancies and to synchronise. Where there is a new taxon concept in the NVS, and there is a unique match from this taxon concept to a taxon concept in NZOR, NVS will be updated with the NZOR taxon concept identifier. Where there is not a unique match some work will be required to decide how they should be linked. Once these links have been made, the NVS includes both the NZOR taxon concept identifier and the most up to date scientific name for the taxon in all data extracts for its users. This means that if the user wants further information about the taxon, for example how it fits into the taxonomic hierarchy, they can obtain this information from NZOR and NVS does not need to reproduce it.