ROYAL SOCIETY OF NEW ZEALAND RESPONSE TO MORST DISCUSSION DOCUMENTS: MORE STABLE FUNDING – BACKBONE SUPPORT FOR DATABASES AND COLLECTIONS, AND PUBLIC ACCESS TO RESEARCH INFORMATION

The Royal Society of New Zealand is the cross-sectoral, independent, national, academy of sciences for New Zealand. Instituted under an Act of Parliament, our role is to 'advance and promote science and technology'. Part of this role includes providing expert advice to government on scientific issues, or issues pertaining to the conduct of science and technology in NZ.

This paper was prepared by the staff of the RSNZ, with input of the Biodiversity Committee, which had discussed this issue at their recent meeting with MoRST officials invited (20 April 2007, minutes here.) This response has not been reviewed or ratified by the Council of the RSNZ; however, under delegated powers to the Biodiversity Committee, it may be considered official policy of the RSNZ.

SUMMARY

We commend the backbone funding policy development as important for New Zealand to enable proper support of nationally important collections and databases. Proper support for such assets is an important component of the 'more stable funding' policy development.

We recommend that a full scoping exercise be undertaken to determine the inclusion or exclusion of databases and collections, in a robust manner.

We suggest that funding needs to be comprehensive (covering all aspects of data collection, curation *and* information sharing) to fully ensure support to enable open access to data.

Finally, we suggest some ways to disseminate information freely, such as a web-based National Science Co-ordination Centre which would hold details of all National Collections and Databases, projects they support, their sensitivities/accessibility, stewardship and content. In making available data from collections and databases, there are international and regional (e.g. Australian) systems that could be used, to enable greater access to scientific knowledge for public good use. (See Appendix 1.)

RESPONSE TO SPECIFIC QUESTIONS

The answers below relate directly to the questions listed on the engagement paper "Backbone support for Databases and Collections" numbered 1 to 8.

Q1. We concur that the points on page 7 cover the issues of dealing with backbone funding for assets. We note, however, that there were 5 points, including the last one:

- Current lack of a clear framework for assessing needs of databases and collections
- Current arrangements for databases and collections are not up to date

- Tensions exist between owners and users
- Databases and collections across the science system are individual and varied, and have specific needs – one size of policy may not fit all
- Also that the level of funding has not kept pace with needs, and some annual adjustment is required.

We think it important to include activities relating to curation and maintenance of assets, even though this in some cases may appear to be 'research' by some (such as identification and systematics of specimens). Collections are not useful without names and/or identification. In addition, maintenance of an asset includes not only the fabric of the building in which it is housed, and the people involved in dataset maintenance and curation, but also the land used to grow (new) living specimens, and ongoing husbandry of live specimens. An exception to this might be genetic databases, such as Expressed Sequence Tag libraries, where it is historically accepted that identification of genetic sequences is a research activity, as opposed to a database maintenance or curatorial activity.

Q2. The suggested "progressive identification of assets for backbone funding that occurs alongside negotiations/contestable reinvestment rounds", may be practical (for investment agents) but it should be remembered that the purpose of the long-term funding is to remove the changeability associated with changing priorities year-on-year. Such a policy would make *funding of associated research programmes* the default criterion for support of the database or collection. This is a problem if there is no extra money appropriated for the assets, as money set aside to properly support such assets must come form the research areas that they exist to support. We believe there are other, more important, criteria to consider, such as *need* by non-research organisations, including regional or central government etc. (Criteria discussed below)

In addition, the progressive identification suffers from the following problem. Reducing the initial roll-out to just one type of investment may have inadvertent consequences due to the order in which investment-types are considered. For example if it started with RFI then went on to Health, and then considered those associated with Environmental Research, the latter may only have a turn at being considered at a time when the pot of money was 'used up', when it was considered that 'already plenty of databases' were being supported, and/or when the initial policy had evolved over time. Such a non-level playing field across the collections and databases could jeopardise a whole set of collections required for one area of research, e.g. Environment. It is better to 'risk' the databases and collections *equally* across all types of investments, by considering them all at once, using the same high-level criteria.

All types of investments could be considered to be critical, therefore it may be important to have the chance to support all their associated databases and collections in equal measure. Therefore, we recommend a scoping exercise below that includes all the databases and collections at once, using high level prioritised criteria, input from a wide range of stakeholders and an analysis of each database and collection, including those not currently supported (e.g. collections in museums for example). Such a system would enable that level playing field among the databases and collections and would be a robust manner with which to make the long-term decisions appropriate for backbone funding.

Q3. Funding and investment agencies are not necessarily best placed to make decisions about the national importance of backbone assets. Such decisions need to come from a higher level, namely MoRST, in conjunction with the rest of government, local and central, and including input from other potential 'customers' or those who currently pay for services.

This may require a 'scoping exercise' to uncover the current use and importance, potential use (i.e. if not constrained, for example, by a regional council's ability to pay for information) and future use – or 'needs on the horizon'. The scoping exercise must also include those collections, for example, that are currently funded by FRST and those not funded (due to a historical accident, or previous insufficient funding to maintain an excellent asset). Some smaller collections could be incorporated into a larger collection to make a more comprehensive and cost effective collection.

A scoping exercise, similar to that recommended in previous years as a Biosystematics Strategy, would need to examine all the collections and databases at once to prevent inadvertent consequences of investing in them within funding rounds. To use the biodiversity collections and databases as examples: their support needs, and associated long-term research needs must be evaluated in the context of 1) information requirements for science, biosecurity, conservation and protected areas, biotechnology and environmental monitoring; 2) the availability of existing expertise and 3) gap analysis with training and mentoring, asset support, and a research strategy aimed to achieve the needs identified in 1.

Potential uses, horizon scanning and utility of assets are really best informed by the scientists who are closely involved with them. There needs to be a certain amount of trust given to these well-informed people to get their input to a scoping exercise, and not to disregard their views as 'interested parties'.

A truly comprehensive, consultative, scoping exercise that covers all the databases and collections would engender buy-in from the taxpayer, the science community, and the end-users (e.g. government agencies, business, other scientists). Most importantly it would ensure robust decision making necessary for long-term commitments through the backbone funding.

Q4. The criteria based around National Importance (relating to established, long-term priorities; criticality to long-term operations of stakeholders; and quality to meet needs of stakeholders) and Best Funding Mechanism (whether backbone funding is in fact the best mechanism for the collection or database activity) seem to be useful starting points to decide the inclusion of databases or collections into the backbone mechanism.

In addition, these criteria have been suggested by our constituents:

- That the activity has a long-standing reputation for providing the information required or for being a comprehensive and excellent collection/ database.
- That the activity underpins a wide range of research projects/outputs /investments.
- That the activity has value to all users, both 'public' and 'public good'.
- The ability to demonstrate a long-term relevance of this kind of data or collection despite changes in the technological environment (e.g. herbaria or insect collections that have been standard use for centuries despite changing technologies).

- Where new technologies add to the utility of a collection or database (e.g. giving a collection more wide-ranging reach, or a database greater powers in its uses).
- The risk of losing the collection/database. What future needs (that may have been elucidated in a comprehensive scoping exercise) could be jeopardised by not maintaining a data series? Could a collection be re-collected and if so, at what cost?
- In the case of 'newer' databases and collections – the value of new paradigms enabled by the activity (might be used as a criterion to balance the inability to show retrospective long-term value and use).

In particular we caution against the following criteria for these reasons

- Ability of end-users to pay; historical commercial viability. (Because this is not necessarily a strong indicator of usefulness.) There are databases and collections that are currently used commercially but that do not demonstrate complete financial viability because of the inability of the clients to pay. Unfortunately this can occur alongside a lack of recognition by FRST of the scientific value of the collection, resulting in effective subsidy by the owner of the asset. End-users (e.g. risk analysts, local government, private businesses like farmers, Biosecurity NZ) are often willing to pay for direct answers or solutions to their immediate problems, but will not pay for the underlying capability or assets required to underpin and attain the answers.
- There are many collections that have been chronically under-funded historically, a fact reiterated in many government, including MoRST, reports over the past 20 years. Therefore historical success in obtaining funding should not be a criterion with greater weight than perceived need among stakeholders, and ability to meet that need.
- It is conceivable that if clients were able to obtain information without cost (through open, free access), then they would use the database /collection to a greater extent than previously, and the result would be improved monitoring, risk analysis, or work outputs for the benefit of NZ. This latter outcome is a beneficial result of open access fully supported through the backbone funding, that warrants the original investment, even if it runs against the philosophy of user-pays.

Q5. Collections and databases are assets set up by people, used by people and run or maintained by people. Therefore it is important that *people* associated with backbone assets are well supported.

This issue may be solved through the Backbone Funding instrument, or it may require additional instruments. The use of a long-term research strategy, e.g. for biosystematics, may enable another solution to this issue. A funding regime for backbone assets must include some support for people involved in curation, for example, taxonomic identification and related systematics of stored specimens in a collection. Otherwise, it's a bit like having a library full of numbered books with no catalogue of authors and titles. Unless this work is explicitly supported by backbone funding, it may be considered 'research' activity and be left to be supported by other contestable instruments, which have traditionally shown poor ability to respond to this need in the science system. (In addition, if that research activity is left to be funded out of a fixed pool from which the asset was funded, then there may be insufficient to support the research/curation required to make the catalogue for the backbone asset.)

It is obvious that any asset requires maintenance, rather than just a setting-up cost. However, the grey scale along which one considers an activity to be solely maintenance, (e.g. keeping a containment facility complying with ERMA rules; cleaning, or maintaining an IT server on which a database lies) or curation or research needs to be explicitly expounded. Then owners would then be clear as to the intent of MoRST regarding the level and *extent* of support for the core activity.

Lastly, when unforeseen issues arise, there needs to be some flexibility to provide additional resources to keep a backbone asset running legally. An example would be the new ERMA rules associated with containment facilities which came in to effect in the late '90s. A possible future rule might be security arrangements around a scientific facility to prevent biosecurity breaches or bio-terrorism.

Q6. Mechanisms to encourage end-user engagement in a backbone asset's development are unlikely to be required, if the support is truly long-term and non-contestable. The major barriers to collaborative discussions were contestability and/or the policy of user-pays; and the proposed policies aim to alleviate both of these. Free access to a completely-supported collection or database would enable free and frank discussion between the scientists and the end-users to improve the provision of information. If it were considered that the owner of an asset were belligerently preventing development of an asset as required or requested by end-users, then the latter should approach MoRST to mediate negotiations around the contract for services associated with that asset.

To encourage greater use by end-users of the National Databases and Collections, the development of a web-portal would enable knowledge about data available. This might also extend to more than just those supported through backbone funding, (with permission of the owners of the other assets). There may be synergies to 'advertising' databases and collections which can provide monitoring data, land surveys, marine ecosystems, genetic sequences etc that can be accessed either free (in the case of backbone assets) or by user-pays (in the case of private collections and databases).

If a scoping exercise were to be undertaken, then such a web portal and the background information would be easy to compile, as much of the information would have been collected from a wide range of database types and collections in advance of deciding which ones were to be funded through the backbone instrument.

Q7. Processes to decide how and when to stop funding a backbone asset need to be decided in advance, and have robust logical backing, with buy-in from owners (organisations), scientists involved with, and users of, any collection or database. Some ideas we suggest are:

- Even when an asset is nearing the end of its life, it still needs support (albeit at a lower level), right to the end. Even if the amount of new information or number of organisms coming in slows, then methods of access will still require updating as technology changes.
- If an alternative resource exists and is available/accessible, e.g. an international collection or a digitised collection or database, then the requirement for a local duplicate should be evaluated and potentially the asset could stop being funded. However, long-term security must not be jeopardised by closing down a local facility in favour of an international one, in case we should lose access to the international resource which might risk our ability to maintain our own priorities (e.g. biosecurity).

- If the research that the backbone was 'for' became redundant, this would be signalled well in advance by declining use of the asset. The asset should be maintained until it is truly no-longer needed, and then funding should be withdrawn in a manner that enables transfer of any *people* involved to other science activities. In addition, it may be deemed appropriate to give (e.g. via international aid) or to sell the asset to another country that still needs it, (in which case it may also be appropriate to support the other country's capability development through training by NZ staff).

Q8. Risks, threats and opportunities arising from implementation of the proposed model are outlined below:

Risks:

- That the activity will not actually be fully supported in the long term, thereby putting at risk the financial viability of the owner /sponsor.
- Alternatively, an owner may use additional funds from the contestable pool to provide support for the asset (e.g. through subsidised research overheads, or associated research activities) which may hide the true cost of maintaining an asset, just to make it appear more efficient and therefore preferable to a competitor's asset.
- New databases may not get a look-in on the funding for a long time. An organisation may need to support a nationally-important new dataset for a long time (several years) before it is deemed to be 'permanent' and worthy of backbone funding. In the mean time, the business-case for keeping the asset may be marginal due to market failure on the part of agencies who need the information but can't afford to buy it on a full-cost basis, (i.e. they may afford the answers or solutions to problems, but not the underlying asset required to provide those solutions, as mentioned before.)

  This is an important issue for a government that sees itself as supporting an information-based pathway to economic development. If that were truly the case, then a commitment to the growth in number of information-based products, and their ongoing expansion, would need to be made to bridge the market failure-gap.
- There is a risk that a database or collection is only supported if it aligns with a government priority in this or next year. However, biological specimen-based databases and collections are key long-term activities (undertaken sometimes over centuries, not just decades). They underpin almost any environmental objective, natural resource management and so on. In particular, it is not known now, how much we will depend on certain types of information in the future, as mankind comes to depend more and more on valid evidence to make decisions enabling sustainable practises.

Threats:

- That it will not gain buy-in from the sector, because some important factors, such as criteria, are chosen wrongly.
- That it will fail to deliver on its goals, by failing to be implemented well in the first place.

- That assets will be funded through existing funding rounds, using no additional money, threatening the research areas (also funded through those rounds) that they exist to support, or (worse) threatening the research required to underpin maintenance, management and curation of a collection or database. This could make the asset unable to function properly, or with no users to benefit from it.

Therefore we recommend a scoping exercise to comprehensively review the need, future need and all aspects of each database and collection, including a look at collections that may not (yet) feature in the proposed backbone funding instrument. This exercise needs to include government agents who use the information, and scientists (and not to disregard their views as 'interested parties'). It should not be subject to current government research priorities (which are annually, or 5-yearly changeable), but to more long-term strategies that encompass the whole of government's needs (central and local).

Finally, one of the criticisms of the discussion documents was that the 'method for prioritisation' had not been clearly stated (mainly because this is still being consulted on). Therefore a further step is required - the whole system needs to be made clear, including criteria for choosing, processes of implementation, robust systems for withdrawing funding etc., and *then this information should be provided for further consultation* to garner any unforeseen pitfalls and to improve buy-in of the roll-out procedure. (Note that following the most recent changes to funding instruments the greatest criticisms from the science sector were around the poor implementation processes. This was also revealed through independent review, in the case of the Outcome Based Investments round.)

Opportunities in the backbone funding system include:

- That the government can enable an information-based pathway to economic development by increasing each year the funding levels, thereby enabling development of new databases that support important research areas.
- That monitoring data required for National Policy Statements on the environment are more likely to be maintained if there is core infrastructural funding for it. This will enable the government to improve its Initiatives on Sustainability by developing policies around a strong evidence-base. (Examples include freshwater quality, air quality, ecosystem classifications, land use classifications, and such potential new databases as carbon sequestration rates in varying models/systems.)
- If there is conflict between the maintenance of a collection and research using that collection, then it could be resolved by funding the work necessary for public access to information – such as providing correct naming, keys, drawings, and other data online – through the backbone, while leaving the activity of learning about the specimens in the research realm of contestable funding. (However, one would want to avoid the risk that the latter does not get done, thereby minimising the utility of a collection.)
- By avoiding progressive identification of backbone activities, and, rather, undertaking a scoping exercise, the government will have a comprehensive picture of the full range of activities that could be incorporated in future into the backbone funding.

Public Access to Research Information

In general we support the proposed public assess principles, including the exemptions.

It is important to recognise that in some cases public good research is subsidised by an organisation's 'own' investment. This may occur through adjusting overheads (in-kind support through not charging overheads), use of CRI capability funds (to maintain capability considered to be of value in the future, but not currently supported) and other methods (e.g. private funding). In these cases, the government needs to take care in enforcing public access principles, as the public purse is not necessarily fully supporting the activities. N.B. If a science organisation were to provide such support because it believed the knowledge to be of sufficient value to warrant the subsidy (as a 'public good' provision) that suggests that the activity should be fully supported via the backbone funding.

In some cases the raw data or primary results are of little value to anyone without expert interpretation, and the cost of providing that can be larger, time consuming and simply too difficult, because the 'expert' may simply not have sufficient time to provide interpretations for those wanting access to the primary data in a timely, meaningful way. Consider that a scientist is employed to undertake research, often for a particular FRST-funded contract of a few years, with quarterly milestones on which to report. This scientist is racing to be the best to win the next contest in a fully contestable system. Such a person may not have the time to deal with official requests for information. (It is therefore not surprising that agencies requesting information are suffering delays in receiving it.) Therefore, for true access to information, that part of the provision needs to be paid out of the backbone funding instrument alongside maintenance of an asset.

It may be necessary for FRST or MoRST to further define 'primary results' to make clear exactly what research information a member of the public may request. This could vary from access to individual's lab books, to the data behind graphs in publications (to the publications themselves).

Summary

We reiterate that the 'cost of dissemination' may be quite large, if hourly rates including overheads are included. In the case of National Databases and Collections funded by the Backbone instrument, this cost should be part of the support for the activity, thereby making access truly free.

We thank MoRST for the opportunity to be involved in this consultation.

For further information, please contact

Dr Kathleen Logan

Manager, Policy and Evaluation
Royal Society of New Zealand
P.O. Box 598
Wellington
Kathleen.Logan@rsnz.org; tel 04-470 5754

Attached: APPENDIX 1.

APPENDIX 1
EXAMPLES OF COORDINATED INFORMATION DELIVERY SYSTEMS

In the Royal Society of New Zealand's response to MoRST's discussion document on backbone funding for databases and collections, we mention the idea of a National Science Co-ordination Centre, which would operate as an information centre for the public about databases and collections. It also might offer 'portals' to the information sources, and eventually manage a certain amount of inter-operability between similar datasets that add utility by being 'joined-up'.

Some collections and databases are already delivering information to the public (end-users) in a co-ordinated manner with standards to assist with making information useful without demanding too much cost by those collecting the data. Standards are particularly important for numerical data that may be delivered in aggregated form and therefore hard to independently verify or quality-control.

It is also important to have standards for delivering information based on natural biotic collections, such as herbaria, animal collections, and ecosystem data. These may include illustrations or photographs, including microphotographs using scanning electron microscopy, keys, naming and location of samples etc. Therefore the delivery of such information predicates upon having sufficient understanding of the organisms in a collection. There is a subtle difference between what research (or study) is required to underpin the proper running and delivery of a collection (naming, key development, loading information onto web-based portals etc.) and what research can then be performed *using* a collection.

Benefits of robust information delivery will become especially important as New Zealand manages its resources in a sustainable way, including land use (forestry versus dairy), fisheries management, soil protection, water management and all the services we take from our surrounding ecosystems.

Another benefit of providing information via portals is the imagery able to be used to quickly inform non-specialists of highly complex scientific data. (For an example, refer to www.nabis.govt.nz for fish distributions, for example.) A risk, however, of such portals is over-simplification of the contexts in which data are obtained, and uncertainties surrounding their use.

Some examples of data delivery systems are mentioned below with references.

GBIF - GLOBAL BIODIVERSITY INFORMATION FACILITY

The Convention on Biological Diversity, to which New Zealand is a signatory, has set up useful systems for sharing and disseminating data. Emanating from the OECD is the Global Biodiversity Information Facility. The New Zealand representative championing the development and use of this facility is Professor David Penman of University of Canterbury. This facility incorporates Biodiversity Informatics for delivery encompassing these strands:

ECAT: Electronic Catalogue of Names; DIGIT: Digitisation of species and observational records; DADI: Data Access and Database Interoperability; and OCB: Outreach and Capacity Building.

New Zealand has a challenge to be fully involved in this process, but in addition, the process makes a good case study for how multiple collections or databases can be brought together under agreed governance arrangements for delivery to users across the world wide web. (see:
http://www.rsnz.org/advisory/biodiversity/archive/biodiv_gbif_2006.pdf)

We advocate for being involved in international processes where they exist, rather than duplicate them in NZ, providing access is enabled free to the users. (This may mean the government paying a membership fee.)

NZ GEOSPATIAL STRATEGY

In New Zealand, an example is the Geospatial Group which has been working for a decade to develop systems for delivering geological, and other sets of information in a manner that has utility for a wide range of users. This has been championed by Dr Dave Loubser currently seconded to Land Information New Zealand (from Ministry for the Environment). The Geospatial Strategy was signed off in April 2007 and encompasses such information as location information, features above and below ground level, ecosystems and use, cadastral information and other data-streams. In particular the strategy aims to develop governance for the system and to increase interoperability between datasets collected by different agencies about similar things.

(See: http://www.linz.govt.nz/publications/geospatial-strategy-2007/nz-geospatial-strategy-2007.pdf)

Benefits of the Geospatial Strategy will be delivered through such portals as the National Aquatic Biodiversity Information System (see http://www.nabis.govt.nz) used by Ministry of Fisheries for resource management.

Some in the Biodiversity Committee have suggested that we could provide a similar service as the LINZ geospatial service, in terms of a taxonomic or systematics 'group' or institute that dealt with identification and phylogenetic issues in relation to biodiversity. Such a group would serve a broad stakeholder group such as Regional Councils, Biosecurity NZ, Department of Conservation Conservancies, fishing industry, Ministry of Fisheries, museums and other science researchers or their organisations.

Australia benefits from a National Collaborative Research Infrastructure Strategy (NCRIS) that is aimed at the application of knowledge through systems of delivery across 12 priority areas. The projects include a national network of medical imaging facilities, an online Atlas of Living Australia and facilities to support gene discovery and genome analysis

ALA ATLAS OF LIVING AUSTRALIA

The Atlas of Living Australia will use standards developed by a Taxonomic Database Working Group (TDWG), now called the Biodiversity Information Standards. These need not be re-developed in New Zealand, but we could piggy-back off the Australian effort to date. TDWG is also liked with the GBIF (see above) and OGC (Open Geospatial Consortium). The importance of these connections is enabling inter-operability between different organisation's databases. In fact now is an opportune time to be involved with the Atlas of Living Austral(as)ia since New Zealand has such a good understanding of its biodiversity. Becoming involved now, rather than later, would enable New Zealand to have more influence on the manner of open-access and inter-governmental support of the facility.

At an institutional level, some collections already make up super-collections for example several herbaria in New Zealand belong to the New Zealand National Herbarium Network, of which the Allen Herbarium at Landcare Research is the largest. In turn, the New Zealand National Herbarium Network is a member of the Council of Heads of Australasian Herbaria (CHAH), a major project of which is Australia's Virtual Herbarium.

SUMMARY

The purpose of this Appendix is to outline some existing and potential examples of data access by cooperation across New Zealand and the region, (and even internationally). To coordinate such memberships and activities, we proposed that an analysis be undertaken of all the collections and databases, to determine the needs, future uses and appropriate support mechanisms for all. This would remove the risks associated with an ad-hoc approach to determining funding, such as the progressive identification of backbone activities by funding and investment agencies during their funding round reviews.