MINISTRY OF BUSINESS, INNOVATION & EMPLOYMENT
HĪKINA WHAKATUTUKI

Scientific Collections and Databases Review

Terms of Reference

June 2018

New Zealand Government

Contents

1. Purpose

These Terms of Reference (ToR) set the parameters for a review of how government invests in scientific collections and databases for the benefit of New Zealand. They state the scope and purpose of the review, key questions the review will address, and the activities of the review team and advisory groups.

The review will inform advice to the Minister of Research, Science and Innovation on how the government should invest in scientific collections and databases.

The review includes, **but is not limited to,** collections and databases currently designated as 'nationally significant' and that receive direct Research, Science and Innovation (RSI) funding.

2. Background

Scientific collections and databases are a type of research infrastructure comprising systematically collected or compiled groups of items or data that underpin or contribute to a broad range of scientific research and science-based services.

Collections generally contain physical specimens such as animals, plants, fossils, rocks, soil, ice cores, and microorganisms and their associated metadata.

Databases are generally stored repositories of information on particular subjects of interest such as climate observations, levels of toxins or pollutants, genetic sequencing, a particular cohort of specimens (ie longitudinal studies), or natural hazards.

Both collections and databases may be expanded as more specimens or data are collected.

Examples of current uses of scientific collections and databases include scientific research, biosecurity, climate change modelling, land-use planning, designing health and social interventions, mining permits, conservation, and setting environmental standards.

Current funding models

MBIE funds 25 Nationally Significant Collections and Databases (NSCDs) through the Strategic Science Investment Fund (SSIF) Infrastructure appropriation. This investment is around $19 million per year. Annex One contains further background information on the NSCDs.

Examples of other scientific collections and databases that directly receive RSI funding through various mechanisms include NIWA's ocean bathymetry and greenhouse gas emissions data, OVERSEER, Antarctica New Zealand's Adélie penguin database, and Manaaki Whenua's Land Cover Database. These were funded as a result of individual funding decisions rather than through a strategic framework for funding scientific collections and databases.

Other scientific collections and databases are held by organisations such as Independent Research Organisations (IROs), universities, museums and government agencies. These have various funding and access models and although they do not currently receive direct RSI funding, they may receive it indirectly through full-cost research funding.

Drivers for the review

Stakeholders from many parts of the science system report[1] that collections and databases underpin critical scientific research and science-based services for New Zealand. The 2015 Royal Society report highlights the importance of and risks to New Zealand's taxonomic collections[2].

A substantive review of how the government invests in scientific collections and databases has not been undertaken since 1996, when several more collections and databases were deemed 'nationally significant'. Changes to the technological and funding landscapes since 1996 are key drivers for the review. Important changes to the science system since then include:

- more opportunities for long-term, strategic science investments
- increased collaboration across institutions and countries
- technological change, especially digitisation, genomics and ICT
- an international open science agenda and advances in best-practice data management
- the big data revolution.

There are also several other drivers for the review:

- A review of the role and scope of scientific collections and databases was signalled in the SSIF Investment Plan 2017-2024[3].
- We do not know if the current approach to investment is fit-for-purpose, if 'nationally significant' is a useful concept, or if there are any gaps or inefficiencies.
- Stakeholders report that a lack of appropriate investment and oversight is putting collections and databases at risk of deterioration and sub-optimal use.
- There is a lack of system-level understanding of scientific collections and databases in New Zealand beyond the current NSCDs.
- Government has a role in ensuring an efficient, effective, transparent and internationally best-practice funding model.
- Publicly-funded collections and databases may not meet best practice for openness and accessibility[4].
- There is increasing expectation to demonstrate value for money from public RSI investments.
- There may be opportunities to increase international connectedness in research infrastructure and data management, especially with Australia.

[1] For example, the Crown Research Institute (CRI) core funding review of 2015 (http://www.mbie.govt.nz/info-services/science-innovation/innovative-new-zealand/budget-2016-funding/pdf-library/cri-core-funding-review.pdf)

[2] https://royalsociety.org.nz/assets/Uploads/Report-National-Taxonomic-Collections-in-New-Zealand-2015.pdf

[3] http://www.mbie.govt.nz/info-services/science-innovation/investment-funding/how-we-invest/strategic-science-investment-fund/document-image-library/ssif-investment-plan.pdf

[4] Expectations and use of publicly funded data are changing and there are several groups and pieces of work across government tasked with meeting those needs.

3. Objectives

1. ***Outline the current role, scope and state of scientific collections and databases in New Zealand's RSI system***
    a. What public good scientific collections and databases currently exist in New Zealand (including but not limited to those currently designated as 'nationally significant')?
    b. How are they funded, accessed and managed (eg ownership models)?
    c. How are they used (by whom, what for, how often, trends, changing drivers)?
    d. What are the existing international connections (types of projects, uses, risks)?
    e. How open, accessible and interoperable are collections and databases currently and what are the barriers (eg cost, commercial, lack of demand)?
    f. What is the nature and extent of the reported pressures and risks created by underfunding (eg risks to capability such as taxonomists or infrastructure such as weather measurement stations)?
    g. What expansion and enhancement activities are currently occurring, and what opportunities are there to catalyse or improve these activities?

2. ***Identify international best-practice and collaboration opportunities***
    a. How are scientific collections and databases defined, funded and managed in other countries and international research consortia[5] (including drivers and trends, funding criteria, degree of centralisation and funding devolution, and accessibility/openness issues)?
    b. How do other countries and international research consortia currently interact with New Zealand's scientific collections and databases?
    c. Are there international opportunities we could utilise, such as increasing collaboration, impact or efficiency?

3. ***Develop investment principles and recommendations***
    a. What is a useful framework (eg criteria) to determine which collections and databases to fund at each level: 1) with public money, 2) with RSI funding, and 3) through an MBIE-managed RSI infrastructure investment?
    b. Should we continue using the concept of national significance?
    c. How can we determine how much to invest?
    d. What investment and operating principles would ensure that collections and databases can contribute optimally to government priorities and broad, public-good impacts (including utilising and enhancing Mātauranga Māori) both now and in the future?
    e. What conditions could be attached to inclusion in an RSI investment in scientific collections and databases (eg free access, digitised, reporting requirements, income generated must be reinvested in the collection or database)?

[5] For example the CGIAR (formerly the Consultative Group for International Agricultural Research)

4. Scope

For the purposes of this review, 'scientific[6] collections and databases' means those that were created by or are used for a range of scientific activities including but not limited to research, routine monitoring, testing, or well-managed citizen science. They contain general purpose specimens or data (ie that are re-usable for multiple purposes), are medium to large scale, and have a wide range of users that are predominantly focused on public-good outcomes.

More detailed definitions and criteria for the kinds of collections and databases government should invest in will be developed as part of the review.

In scope

- Existing 25 NSCDs and others already receiving direct RSI funding.
- Other general purpose, public good-focused scientific collections and databases held/owned primarily in New Zealand that might be appropriately included in an MBIE-managed RSI infrastructure investment.
- Existing and future scientific data that should be incorporated into an MBIE-managed RSI infrastructure investment in scientific collections and databases (gathered through research, monitoring, testing or other scientific activities).
- Data infrastructure and points of access.
- Capability required to support curation, maintenance, and enhancement.
- Developing high-level principles and recommendations for how to fund scientific collections and databases.

Out of scope

- Whether to fund scientific collections and databases (the policy rationale for a government investment in some form is already clear).
- Developing a general policy for managing all research data and government-held scientific data (ie all research data, including that not generated for or from a publicly funded collection or databases).
- Long-term training and capability strategy (ie skills policy, predicting or addressing any impending capability needs or gaps).
- Collections and databases not used primarily for scientific research (ie those in non-scientific fields such as arts and law).
- Collections and databases held primarily overseas and funded by overseas organisations.

Other

An MBIE-managed RSI infrastructure investment should complement other public investments in scientific collections and databases. Examples of other public investments include:

- RSI-funded but not MBIE-managed (eg health-focused collections and databases and longitudinal studies)
- non-RSI publicly funded (eg fisheries data)
- mixture of public and private funding (eg collections and databases held and managed by organisations such as museums and universities).

[6] 'Science' is defined in the Crown Research Institutes Act 1992: "**science** includes the physical sciences, the biological sciences, and the social sciences; and also includes technology; and **scientific** has a corresponding meaning."

The review will consider these three categories to develop an investment framework for public investments.

Once the principles and recommendations have been developed, the relevant organisations will be responsible for applying them to their collections and databases separate from this review (but guided by MBIE if required). This review will not seek to address issues or provide advice specific to other collections and databases.

5. Government investment in scientific collections and databases

The rationale for government investment in some scientific collections and databases is clear. Scientific collections and databases generate broad and sometimes unpredictable positive externalities, many of a public good nature. Alongside scientific research that contributes to a wide range of impacts, scientific collections and databases have many science-related uses such as:

- providing evidence and reference material for monitoring, policy, and decision-making (examples include biosecurity, resource management, health and social interventions, Environment Court cases, or hazard planning)
- education from pre-school through to post-graduate levels
- citizen science
- commerce (eg informing risk assessments for insurance or mining companies).

Scientific collections and databases can contribute to research in unpredictable ways long after their inception, usually due to technological developments opening up new uses. For example, genetic technology has allowed extraction of genetic information of flora and fauna collections that existed before genetic technology could even be conceived.

Furthermore, scientific collections and databases often contain irreplaceable historic or longitudinal information (eg taxonomic collections or climate data) that have been added to over many years and are unique or specific to New Zealand.

The review will test the following assumptions about how the government should invest in scientific collections and databases:

- Publicly funded scientific collections and databases should be readily available as per New Zealand's expectations for open data[7].
- Broad public benefits of scientific collections and databases should be maximised.
- The use and impact of scientific collections and databases should be appropriately tracked.
- Maintenance and curation should be adequately supported and to international best practice.
- Scientific collections and databases should be safeguarded against deterioration or destruction.
- Management of scientific collections and databases should be coordinated, efficient and cost-effective with an optimal amount of redundancy.

[7] https://www.ict.govt.nz/guidance-and-resources/open-government/new-zealand-data-and-information-management-principles/ and the complementary https://opendatacharter.net/principles/

- Ongoing maintenance, enhancement, expansion and adaptation of scientific collections and databases should occur in a financially sustainable way including adding data and specimens, and exit to other forms of management where appropriate.
- Any commercial value of scientific collections and databases should be gained through value-add scientific research or services and appropriately funded by the benefitting entity or sector.
- Regular reporting on all of the above should be required.

6. Timeline and Deliverables

The key deliverable is a **review report** documenting the review processes and findings including summary of engagement with stakeholders. The following is a preliminary outline of the report.

Background

- Defining scientific collections and databases
- How the current NSCDs were selected and any policy changes since 1996
- Drivers and key questions for the review.

Summary of the review approach and methodology

Summary of New Zealand's scientific collections and databases

Current nature, role and scope of scientific collections and databases including:

- identifying other scientific collections and databases where there is a case to fund with RSI funding
- outlining funding, strategy and access arrangements and any rationale for them
- maintenance, expansion and enhancement activities
- identifying the range of uses/users and pathways to impact (including international users).

International arrangements for collections and databases

- International experience of public funding and management and of databases and collections including:
    - funding models (ie is the funding institutional, user-pays, deposit-side pays?)
    - the policy rationale for government funding (ie is it treated as research infrastructure?)
    - what kinds of collections and databases get government funding? (ie any definitions - do they use some kind of 'nationally significant' concept?)
    - any reports on usage and impact (who uses them, what for, and what are the benefits)
    - any legislation
    - availability (free/public, or do they extract commercial value and why)
    - existing international connections and opportunities
    - best practice arrangements
    - opportunities for collaboration with New Zealand.

Findings of the review

- Common themes from the review including

- funding, strategy and access arrangements and any rationale for them
- the range of uses and users and pathways to impact (including international users)
- maintenance, expansion and enhancement activities
- issues and opportunities
- best practice arrangements.

A draft review report will be provided to key consulted stakeholders for feedback, then to the Steering Group before being finalised.

MBIE will provide advice to the Minister of Research, Science and Innovation on the findings of the report and next steps. Ministerial decisions on that advice will guide subsequent stages of the review.

The report will be made publicly available on MBIE's website once the Minister has made decisions on the recommendations.

Activity	**Timing**
Steering group appointed and first meeting	May
Cross agency group appointed and first meeting	May
Develop preliminary project plan and stakeholder engagement/communications strategy	May-June
Begin review and refine project plan with key tasks and milestones	May-June
Report completed	October
Advice from MBIE to Minister	December
Implementing Ministerial decisions	2019

Annex One – The Nationally Significant Collections and Databases

Current strategy and funding

The original 20 NSCDs were designated in 1992 and custodianship of each one was assigned to a newly formed Crown Research Institute (CRI). In 1996, five more collections and databases were designated as NSCDs.

The current NSCDs were identified and funded on the basis that:

- the databases and collections are being held on behalf of New Zealand, where continued provision, maintenance and utilisation are critical for New Zealand science to deliver public benefit
- the benefits accrue to many, varied users and third party beneficiaries while the costs of provision belong to the CRI.

Funding was to allow the custodians to at least maintain the NSCD, and there was an expectation that the data will be freely and publicly available where it is inappropriate for the end-user to pay (otherwise, access should be cost-recovered).

In 1996, FRST committed to give at least two years' notice to NSCD custodians if a funding cut was coming. This was to allow time for custodians to find alternative funding sources.

Progress of the NSCDs since inception

Since the current NSCDs were designated, no further reviews have taken place.

Online accessibility has increased for most NSCDs. However, there is an ongoing trade-off between the view that publicly-funded collections and databases should be freely and publicly available, and custodians' ability to extract commercial value from them. Ongoing cost of maintaining and improving them is also a consideration. Currently, this balance varies across the NSCDs depending on the custodian's open data policies and the nature of the NSCD (i.e. whether there much commercial value to be extracted).

Funding has decreased in real terms. Custodians have been using funding from other sources (such as SSIF programmes) to ensure NSCDs are maintained to a workable level but they report that this is becoming increasingly difficult. For example, some of the maintenance work is done by retired volunteers.

Budget 2016 allocated an extra $2 million per year to the SSIF infrastructure appropriation to increase funding for the NSCDs. This funding was allocated on a pro-rated basis to NIWA, Manaaki Whenua, and GNS, which are custodians of the greatest number of NSCDs.

Current NSCDs and custodians

Organisation	Collection/database
AgResearch	- Margot Forde Germplasm Centre
Callaghan Innovation *(not funded through SSIF)*	- Solar UV-B Radiation Database
Cawthron Institute	- Micro-algae Collection
Institute of Geological and Nuclear Sciences Limited (GNS)	- National Earthquake Information Database - National Petrology Reference Collection and PET Database - New Zealand Fossil Record File - New Zealand Geomagnetic Database - New Zealand National Paleontological Collection and Database - New Zealand Volcano Database - Regional Geological Map Archive and Database - National Groundwater Monitoring Programme
Landcare Research New Zealand Limited	- Allan Herbarium and Associated Databases - International Collection of Micro-organisms from Plants and Associated Databases - Land Resource Information System - National Vegetation Survey Databank - New Zealand Arthropod Collection, New Zealand Nematode Collection and Specimen and Information Database - New Zealand Fungal Herbarium and Associated Database - Nga Tipu Whakaoranga Ethnobotany Database and New Zealand Flax and Living Plant Collections
National Institute of water and Atmospheric Research Limited (NIWA)	- National Climate Database - New Zealand Freshwater Fish Database - NIWA Marine Benthic Biology Collection - Water Resources Archive
The New Zealand Institute of Plant and Food Research Limited	- Crop Germplasm Resources Unit - National Collections of Fruit Crop Germplasm
Scion	- National Forest Herbarium and Database

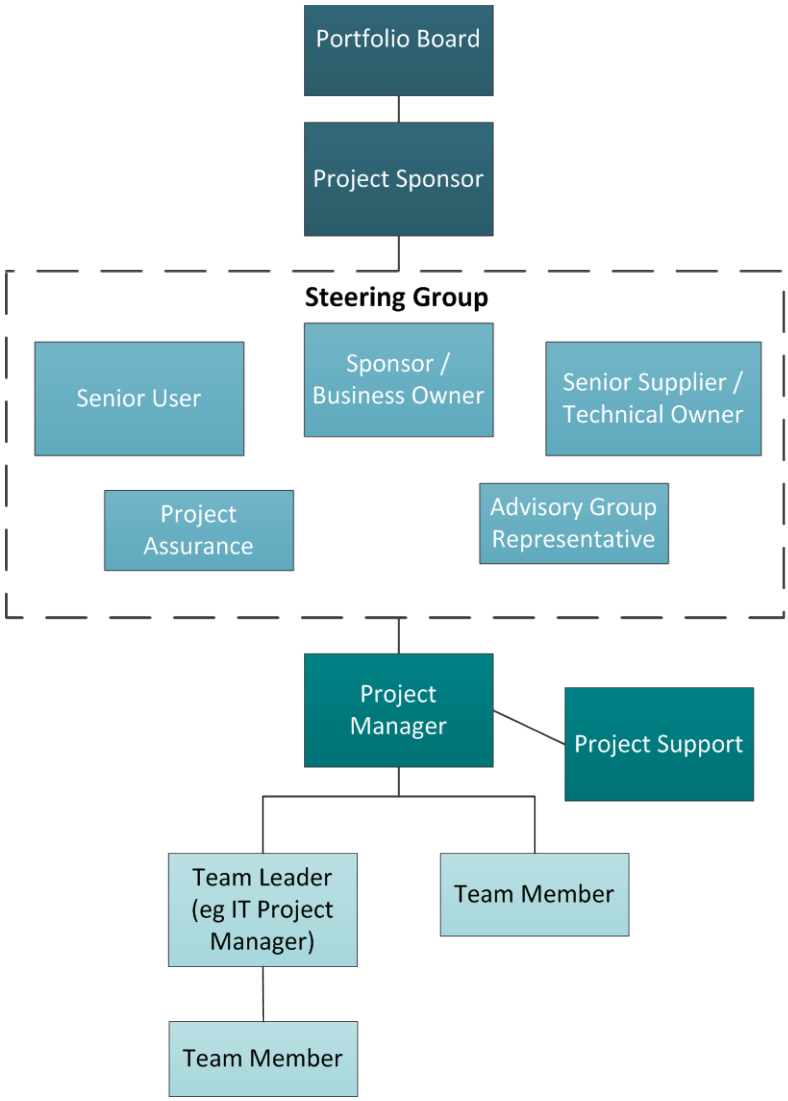
Note that Antarctica New Zealand's Adélie penguin census database is funded through SSIF Infrastructure but is not an NSCD.

Other collections and databases

Other entities such as museums, universities and private organisations hold scientific collections and databases. Funding models for these vary and can include any combination of government funding from sources other than RSI, user-pays, institutional support, private/philanthropic support, contributor-pays (i.e. they pay for their specimen or data to be stored safely). Non-NSCDs may receive RSI funding indirectly such as through institutions that receive RSI funding, or through specific research projects.

The current NSCDs are mostly focused on physical and biological sciences (such as geology, microbiology, ecology, hazards, climate change, and soil science) rather than, for example, social

sciences or emerging areas such as ICT and robotics. This may be a reflection of the fact that the current NSCDs were divided between the CRIs, which are focused on the former areas. Understanding what other scientific collections and databases exist, how they are funded and managed, and how we can optimise value from them are key questions for this review.



Roles and Responsibilities

Update the table below to reflect the specific structure of your Project.

Project Role	Named Individual
Sponsor	
Business Owner / Senior User(s)	
Technical Owner / Senior Supplier(s)	*There may be more than one, they could be internal or exterior to the organisation*
Project Manager	