

Executive Summary

Transdisciplinary approaches are being proposed to address wicked problems like that of climate change. Access to research data from different disciplines is key to utilisation of such approaches. New Zealand researchers face significant challenge in accessing and utilising such data due to a lack of well resourced, supported, easy to navigate digital platform. Authors of this submission have worked on establishing such a digital platform for environmental data called “National Environmental Data Centre (NEDC)¹” where they came across challenges of governance, funding, resourcing, and coordinating cross-organisational research data infrastructure.

To address these challenges, we propose the development and implementation of a ‘National Research Infrastructure Strategy’ with Mātauranga Māori at its heart. The strategy should look at developing ‘Māori knowledge hubs’ in key areas enabling researchers to work collaboratively with the knowledge hubs in developing solutions to New Zealand’s wicked challenges. In addition, the National Research Infrastructure Strategy should address the challenge of research data infrastructure, supported by roadmaps for implementation and investment plans. It should address the data sovereignty challenge along with following FAIR and CARE principles of research data.

We emphasise that developing and fostering capability in the data and digital skills would be the key in realising benefits of a national research data infrastructure. A workforce development plan should thus be part of the strategy development exercise. Conversely, the availability of transdisciplinary and disparate data will enable development of new methods and techniques to help solve our national challenges as well as raising the skills level of our workforce.

Background

Research data is a largely untapped source of competitive advantage for New Zealand². Accessing and using high-quality data is fundamental in driving research excellence. Such data has traditionally been inter- and multi-disciplinary, however developments made in data science in recent years means access to trans-disciplinary data is becoming essential in addressing challenges like that of climate change.

This submission is from a group of senior researchers and leaders from Crown Research Institutes (CRIs) involved in the cross-CRI initiative of NEDC. We envisage this initiative to act as an integrator of existing disparate environmental data sources for easy use and application by science, government, industry, and the public. Once established, NEDC will allow applications to dynamically, and in real time, link to diverse databases hosted by organisations in a federated ‘broker’ approach (rather than aggregating the data in one place).

However, the fundamental problem faced by us and our researchers is to make research data, models, tools, national collections, and databases publicly available via a stable, easy to navigate digital platform. Even if such data platforms are made available, they are often short-lived or have faced severe constraints e.g., lacking resources for update/maintenance, data quality control as well as proprietary or privacy issues.

¹ <https://nedc.nz/>

² *Our Land and Water Contestable Funding Round 2022*

A similar challenge is faced by our Kaupapa Māori researchers and other Māori stakeholders who are lacking access to such data and associated research infrastructure. In case of environmental research and associated data, this is of especial importance as significant emphasis is put on environment in the Māori world view with people and environment being intertwined. This understanding of interconnectedness forms the basis of using the holistic approach which is fundamental to Mātauranga Māori - a dynamic and continually evolving knowledge system involving generational observations and experiences. Additionally, the issue of data sovereignty is an important one and something that needs to be addressed.

Our submission covers only the relevant sections from the six areas outlined in the Te Ara Paerangi consultation document. Throughout this document we have highlighted challenges the team has come up against in its attempt of establishing NEDC. Owing to these challenges, what has been developed in NEDC so far is limited in functionality. Based on our experience, our view is that lack of research data infrastructure is a national challenge, and the scope of this submission goes well beyond environmental data.

A National Research Infrastructure

Pivotal to improving management of and access to New Zealand research data is the development and implementation of a 'National Research Infrastructure Strategy' with Mātauranga Māori at its heart. This approach will bring traditional science of Māori knowledge and researchers working together to create solutions to New Zealand's wicked challenges. The strategy will need to be guided by national science priorities with a New Zealand research data infrastructure being one priority likely to be perennial and enduring. In addition, to ensure national research infrastructure supports Vision Mātauranga there is a need for the development and implementation of 'Māori knowledge hubs' in key areas, something which needs attention at a national level.

One possible model for such research infrastructure strategy is that of the "National Collaborative Research Infrastructure Strategy" (NCRIS) in Australia. The current initiative of Te Ara Paerangi provides an opportunity for New Zealand to learn from and where possible align with the NCRIS. Our proximity to Australia along with similarities in several areas would make this a viable collaboration with our Australian colleagues.

From our NEDC experience, we have faced challenges in governance, funding, resourcing, and a coordination of a cross-organisational research data infrastructure. As a result, the current functionality of NEDC is limited. These challenges are not limited to NEDC however and are common for data intensive research in New Zealand. To ensure this is addressed at a national level, we propose that the Te Ara Paerangi national research infrastructure should be focused on addressing the challenge of research data infrastructure. This should be supported by roadmaps for implementation and investment plans, all regularly refreshed. At its heart, the strategy and roadmap need to:

- Fully embrace Māori data sovereignty principles and the CARE (Collective benefit, Authority, Responsibility and Ethics) principles for indigenous data governance.
- Align with the FAIR (Findable, Accessible, Interoperable, Reusable) Principles.
- Lead New Zealand towards a more open science system.

We envisage the following to be essential components of research data infrastructure:

1. Research Data Portal

A centrally-funded and supported science research data portal is a priority for the data infrastructure. The NEDC portal (<https://nedc.nz/>) is a cross-CRI initiative to achieve something comparable for datasets held by the CRIs but this needs to be extended to other science data providers and resourced accordingly with appropriate governance.

2. Identifier services

Persistent identifiers for research data, research samples, files, documents, or other digital objects help connect objects to important context surrounding the objects and adds value to them. A central registry of identifiers will enable unique, persistent, and locatable identifiers for key objects.

3. Research vocabularies

Vocabularies are agreed lists of terms with definitions and sources, ideally hierarchical with synonyms, that enable meaning of terms to be conveyed between agencies and to the data end user. Through a central repository, vocabularies can be located, managed, and linked to. There are many international vocabulary resources that can be tapped into but there are also opportunities to build and maintain vocabularies that meet specific Aotearoa New Zealand requirements, including te reo Māori incorporation.

4. Shared Technology

Technologies or solutions for some issues like data publishing, data sharing, and research data services only need to be created once, then made available to the sector thereby avoiding replicating multiple individual research data management solutions across research institutions.

Workforce

The power of data for science, research and innovation has not yet been fully realised. The untapped 'connected' value provided through research data infrastructure requires skills and capabilities in a range of areas, including computational, data management and analytics along with skills in social science and design thinking. The latter will need to be nested alongside skills in data science to achieve the greatest impact for Aotearoa, New Zealand. Competition for all these skills goes well beyond the science sector making talent attraction difficult, especially when the remuneration and incentives in the RSI sector are often not comparable with that of other competing sectors.

Additionally, while some within RSI are highly digitally proficient, there is still a disparate knowledge gap and a lack of confidence in the use of digital tools and technology that could enhance science impact and efficiency. This coupled with the fact that tools and technology are rapidly changing makes keeping pace an ongoing challenge.

Fostering talent is critical to developing more digital minded people progressing towards data-led careers in the RSI sector. The research infrastructure strategy should thus lay out a roadmap on how such future workforce will be developed along with a clear vision on their career progression. Additionally, improving rewards and recognition for digital and data skills would be needed to ensure we can competitively attract and retain talent.

It is also envisaged that the availability of national data infrastructure will further promote development of data science skills. The availability of transdisciplinary and disparate data will enable researchers developing new methods and techniques to solve wicked challenges, resulting in further developing national capability in this area.

One thing is certain, people will always be critical to the successful implementation of any strategy. We must ensure we are designing science and research programmes of the future taking a human centric approach.

Funding

Stable, long-term funding has been largely beneficial for the designated Nationally Significant Collections and Databases, but the exponential data and technology growth means that management of important datasets in general is not well supported with the current limited funding model. Moreover, an absence of national leadership and coordination of existing disparate research data have led to a diverse – and therefore not fully interoperable – approaches how individual datasets are managed.

The National Research Infrastructure Strategy would continue to build on the progress made by NeSI (compute) and REANNZ (network) who have been successful in connecting the RSI digital ecosystem. International examples of these infrastructures can be examined, including the [Australian Research Data Commons](#). Other examples can be found in Europe, Germany, and the United States.

The New Zealand RSI sector requires Government support and funding to steer the national research infrastructure strategy. It should be influenced by current science priorities, including but not limited to national water quality and green-house gas research.

Base funding would be required to support the establishment and ongoing activity of initiatives like that of the research data infrastructure. This base funding should cover the operational activities of the data infrastructure, associated platform and skills development. The data infrastructure would thus be providing enabling research activities, rather than undertaking research.

Contributors to this response

Jochen Schmidt, Chief Scientist Environmental Information, NIWA

Mark Rattenbury, Programme Leader, Nationally Significant Databases & Collections, GNS Science

Linley Jesson, Data Science Group Leader, Plant and Food Research

Richard Waldin, CIO, Scion, Claire Stewart, Portfolio Leader, Scion

Nick Spencer, Team Leader Informatics & eResearch, Manaaki Whenua Landcare Research.

Greg Rossiter, Director Technology & Digital Services, AgResearch

Nauman Maqbool, Knowledge & eResearch Leader, AgResearch

Jan Sheppard, Chief Data and Analytics Officer, ESR

Claire Stewart, Portfolio Leader, Scion