Parliamentary Commissioner for the Environment
Te Kaitiaki Taiao a Te Whare Pāremata

A federated system to improve environmental information

April 2025

Note

"Information is the oil of the 21st century and analytics is the combustion engine"

Peter Sondergaard,
Senior Vice President of Gartner, 2011.[1]

Introduction

Throughout his tenure, the Parliamentary Commissioner for the Environment has repeatedly raised the issue of the quality of New Zealand's environmental data and the information base that it supports. Access to high quality and comprehensive environmental data are essential for informing both the development and evaluation of almost all operational and policy decisions. It also has implications for the budget process as information is required to assess the effectiveness of spending directed towards environmental outcomes and assuring ministers of value for money.

Current arrangements do not reliably provide the information base on which well-informed decisions can be made. And without robust environmental information we won't be able to judge if costly actions and mitigations undertaken are making a difference. More broadly, if environmental information is not organised well, it challenges our ability to make efficient and timely decisions across many areas. In the social-economic sphere, the Integrated Data Infrastructure (IDI) has demonstrated the value of well-organised information to assess the return on investment for government services through the Social Investment Approach (see below).

This PCE note provides policy makers with the beginnings of an investment case for a better environmental information management system for New Zealand. Pursuing a federated system is recommended. Its purpose is to ensure that the needs of a future environmental management system are taken into account, and not precluded by, the current reforms of the environmental management system, in particular, resource management reform and the organisation and funding of public research and science.

What is environmental information?

Environmental data can take many forms. It can be derived from routine or targeted surveys or monitoring, generated from experiments or created as a model output. Data can be geospatial and aspatial and collected periodically or continuously at different temporal and spatial scales (e.g. local, regional, national). Data can be seemingly random until it is organised.

Environmental information is data that has been analysed and interpreted to add meaning and value. This process is how environmental data are made useful. An **environmental information system** can be simply described as a network of information which includes the way information is collected, analysed, stored and used. In short, a good system ensures that data and information are accessible and able to be used.

1 Dhande, M (n.d.)

Shortcomings of New Zealand's environmental information system

New Zealand's environmental information system suffers from a number of shortcomings. [2] It is:

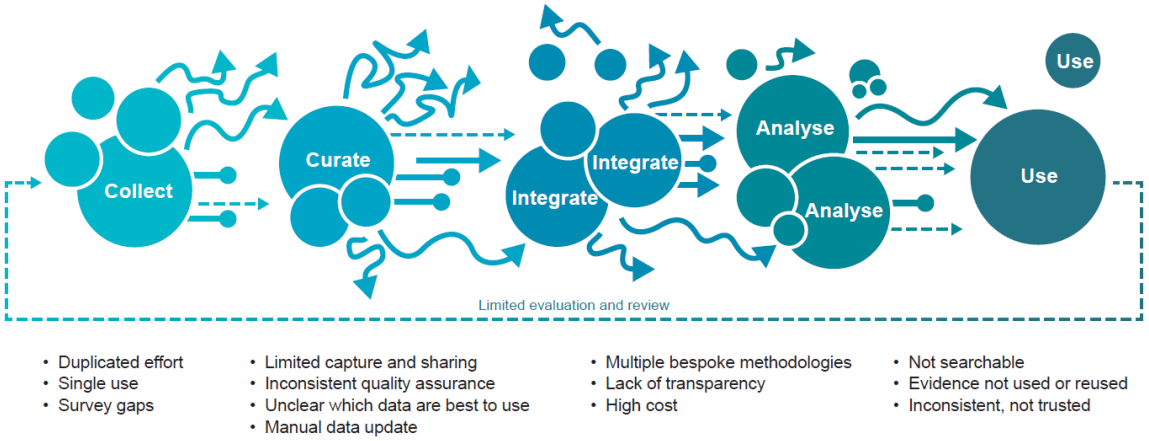
- **complex** – environmental information is collected by a wide range of organisations for a wide range of purposes
- **fragmented** – hundreds of different organisations hold and use (and often collect) environmental information to enable them to carry out their functions. These organisations include:
  - over a dozen central government agencies including the Ministry for the Environment (MfE), Stats NZ, the Ministry for Primary Industries, the Department of Conservation, Land and Information New Zealand (LINZ), the Environmental Protection Authority (EPA), Taumata Arowai, the Ministry of Business, Innovation and Employment, the New Zealand Transport Agency, the Ministry of Transport, KiwiRail, the New Zealand Defence Force and others
  - 78 local government organisations (regional councils, unitary authorities, and territorial local authorities) through their statutory and non-statutory functions
  - the science and research sector, including CRIs, universities, and independent science organisations. In particular, three CRIs (MWLR, NIWA and GNS Science) hold many nationally significant and second tier databases.
  - farmers, foresters, catchment groups, primary sector processors and other environment-related businesses all collect information to inform their business decisions, demonstrate to customers their sustainability and meet regulatory requirements.
- **dispersed** – anyone who is trying to build a comprehensive picture of the lie of the land at any specific location (e.g. a catchment), needs to contact several organisations to access that information
- **plagued by duplication and overlaps** – multiple organisations hold similar information or parts of the same information. For example, aspects of sea-level rise data are held by LINZ, GNS, NIWA and regional councils.
- **plagued by significant gaps** – there are many areas where we don't know enough about what is happening. For example, incomplete understanding of our native biota combined with scant data (which is often distributed across multiple databases in various agencies) poses challenges for both New Zealand's biosecurity services and conservation efforts.
- **opaque** – it is often not clear what information exists, where it is held and by whom. Poor documentation makes it hard to assess the quality and robustness of some existing information.

---

2 Several PCE reports to Parliament have described environmental information system and associated issues. These reports are: Focusing Aotearoa New Zealand's environmental reporting system, 2019; A review of the funding and prioritisation of environmental research in New Zealand, 2020; Wellbeing budgets and the environment: A promised land?, 2021; Environmental reporting, research and investment: Do we know if we're making a difference?, 2022.

- **poorly accessible** – much existing information is difficult to access due to commercial considerations and privacy and data sovereignty issues. Lack of data digitisation is another reason for poor accessibility
- **lacking in strong leadership** – New Zealand lacks an 'environmental information' champion with expertise and a national remit, clear responsibility or adequate and ongoing funding for stewardship and coordination across the breadth of national environmental information. Further, environmental information is not treated as an asset or as 'infrastructure' that enables better decisions.
- **lack of standardisation or compatibility** – inconsistencies plague the way environmental data are collected, analysed, reported and stored. As a result, many ad hoc and bespoke solutions have been developed making it difficult to aggregate information from different sources.

New Zealand is not alone in its shortcomings with the environmental information system. A 2020 comprehensive review of the Australian Environment Protection and Biodiversity Conservation Act 1999 (EPBC Act) revealed that Australia faces many similar challenges. Figure 1 was sourced from that review, but it is equally applicable to New Zealand's environmental information system.



Source: Samuel, 2020, p.165 EPBC Act review

Figure 1: Current environmental information system at a glance.

The provision of information can be viewed as a supply chain. Information is delivered through a series of processes that convert raw observations and data into products that can be used by decision makers to inform their decisions. However, instead of an efficient and relatively straightforward process, there is currently a lot of duplication and limited connectivity, which often trail off. The corollary is that valuable information is highly unlikely to be used to its full potential.

What is needed is some ability to easily and quickly draw together existing information from disparate sources. Being able to do so reliably will in turn reveal the nature of the gaps in the information base that need to be filled over time.

What makes a well-functioning data ecosystem?

In the simplest terms, the elements of a good data system are that it is reliable, scalable and maintainable. Other foundations of a successful data system should include the FAIR and CARE principles, meet trust and privacy expectations, have robust quality assurance processes, and respect aspects related to data sovereignty.[3] In New Zealand the system must take account of the Treaty of Waitangi, matauranga Māori and relevant aspects of te ao Māori.

To achieve all of these elements, a well-functioning data ecosystem requires:[4]

- a data governance framework and efficient organisational structures;
- policies regarding data management planning, data custodianship and curation, legal frameworks, and the use of externally sourced data;
- procedures and processes to execute those policies and manage data;
- engagement with data consumers and stakeholders; and
- technology platforms that support data collection, storage, description, analysis, linking, delivery and curation.

How a data ecosystem is organised will depend on the function and purpose of the data system.

Data system structures can be grouped into three broad categories: centralised, distributed and federated (Figure 2).

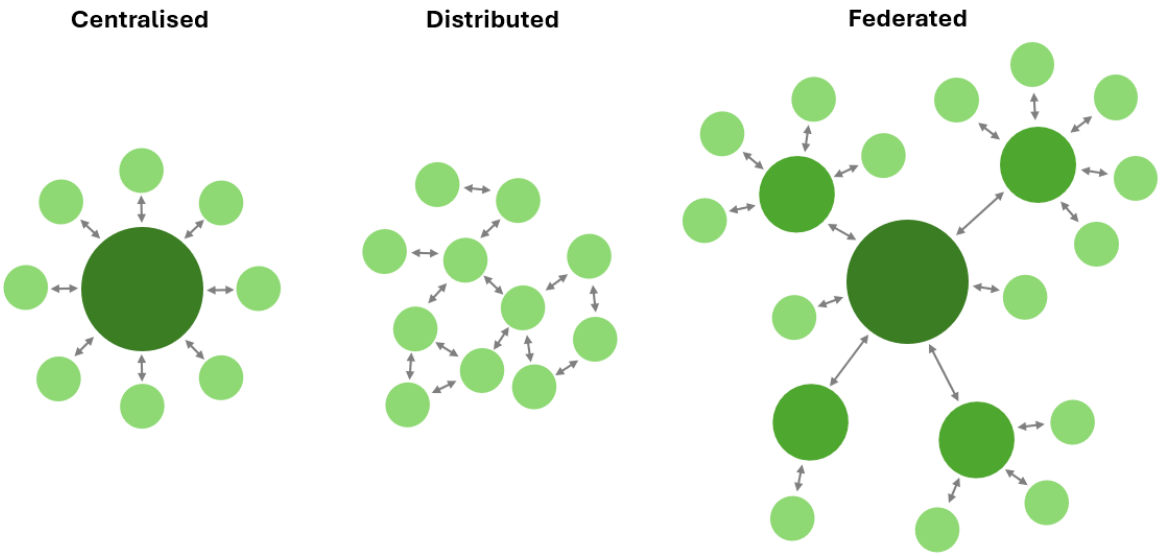


Figure 2: Representation of centralised, distributed and federated database systems.

3 FAIR stands for Findable, Accessible, Interoperable, and Reuse (Wilkinson et al., 2016). CARE stands for Collective benefit, Authority to control, Responsibility, and Ethics (Global Indigenous Data Alliance, n.d.).

4 Medycki-Scott et al., 2016.

Each of these system structures has strengths and weaknesses and will be appropriate depending on the context and needs of those within the system.

Centralised data systems are where data are pulled into a single location or database that is centrally managed. Access to the information can be easily controlled and coordinated.

In some cases, **centralisation** is needed to ensure that standards are upheld, access is managed appropriately, and to ensure that there is a single source of 'truth'. Many of Stats NZ's functions fit this structure. The Integrated Data Infrastructure (IDI – see below) is an example of centralisation of data. Agencies who contribute are not able to access each other's data, with only de-identified data available to organisations and researchers. This highlights one of the major drawbacks of a centralised system. Those within the system do not have control over how the data they provide will be used.

In a distributed data system data are stored in a range of locations with the sharing of that data managed between individual actors.

Distributed systems, such as a data commons, make sense when transparency and ownership are important.[5] This is because they allow data owners (right down to the level of individuals) to control how and when data are shared. It also introduces redundancy into the system as there is no single point of failure but can increase complexity. Data interoperability can also be challenging where different standards and protocols are used.

Federated data systems transparently map multiple autonomous database systems into a single system. The constituent databases are interconnected by a series of consistent policies to create a uniform environment so that the member networks can share data and services. Each system remains independent, with control of the data remaining with the host organisation.

Federated systems occupy an intermediate position between centralised and distributed systems. Separate ownership and control are retained but the benefits of increased sharing and interoperability are facilitated by common standards and systems. They also balance independence, autonomy, and cooperation. This assists the agility and adaptability of the data system. Examples of operational federated data systems include Te Whatu Ora/Health New Zealand and the National Health Service in England both of which support federated data platforms for operational and patient data.[6]

Stats NZ has done some work on changing the government's current data system to an increasingly federated model.[7] The work highlights how data federations could be utilised to address many of the issues within the current system such as trust, privacy, and ownership as federated data ecosystems are built on relationship-based partnerships that allow data stewardship to remain with the original custodian. These partnerships mean that there is more likely to be collaboration and sharing of data between a network of people from government agencies and local government, communities, businesses, academia, non-government organisations and Māori collectives.

---

5 Data commons are a distributed data system that usually have peer-to-peer architecture in which each part is connected to all other parts and is responsible for managing its own data and coordinating transactions. Data commons are usually co-designed and co-governed by participants, with data treated as a common pool resource that is shared equally by everybody in the community (often formalised using a trust structure), subject to the community's rules. For more see Mansell et al. (2017) and Grossman (2023).

6 NHS England, n.d.; and Hospital Times, 2023; and Health NZ/ Te Whatu Ora, 2024.

7 Stats NZ, 2023b.

There is no one perfect system for information management. Which system works best will depend on the information domain and the outcomes desired. The best choice will be the one that results in a system structure that can deliver useful information, in an integrated way, to the people who need it. The Government created the IDI to enable it to take a social investment approach. We should be striving for an integrated environmental information system that works for environmental policy. The following sections look first at what was done to support social investment and then investigates what might be needed for environmental policy.

New Zealand's Integrated Data Infrastructure for socio-economic information

From the mid-2000s, there was an increased requirement from the Government to agencies for data-informed policy advice. In particular, the then Minister of Finance and future Prime Minister, Sir Bill English became a key advocate for using quantitative insights to measure the needs of vulnerable people and evaluate the outcomes and effectiveness of services provided to them. The policy approach became known as social investment. Social investment uses data to target social spending towards certain disadvantaged groups and measures to achieve the best 'return' on long-term social outcomes and on government spending.[8]

In order to do so the Government needed to be able to access information in a more integrated way. This 'need' resulted in the development of the IDI led by Stats NZ. The IDI is used to ensure that the development of social, health and socio-economic policy is informed by a strong evidence base. It has created the ability to generate quantitative insights on the effectiveness of policy interventions, both in terms of policy design and policy monitoring and evaluation. In many ways, having access to this evidence base also supports parliamentarians when engaging in scrutiny of government spending for social and economic policy. Social and economic data are predominantly administrative, centred on individuals, and it is mostly collected by central government.

The following sub-sections detail the design, structure and use of the IDI. Many, but not all, of its features would help address some of the weakness outlined above that currently plague the environmental information system.

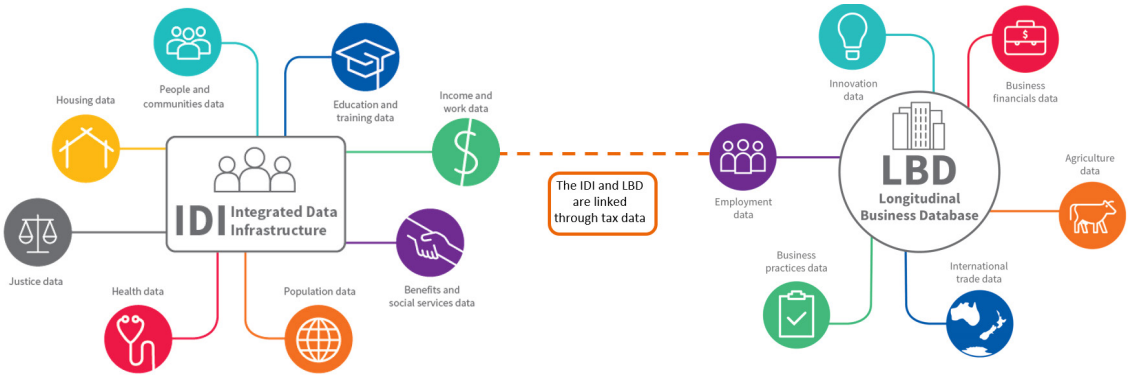
How does the IDI work?

The IDI draws together disparate and fragmented administrative data on social and economic issues. Data integrated into the IDI is collected by different government agencies and non-governmental organisations (NGOs) for various other purposes. The data are de-identified to ensure the privacy and confidentiality of real individuals.[9] These data are then matched across data sets to create an 'ever-resident' population (all people who have ever been resident in Aotearoa New Zealand), which researchers can interrogate to quantitatively explore complex social and economic issues.[10]

---

8 See Deloitte 2016.
9 Stats NZ, n.d.
10 For example, the IDI has been used to build the COVID-19 complex contagion model that informed the COVID-19 response. It has also been used to model social outcomes or better target school-based equity funding (Jones et. al., 2022, p.6–7).

The IDI draws on eight broad categories of data that include: (1) people and communities, (2) housing, (3) justice, (4) health, (5) population, (6) education and training, (7) benefits and social services, and (8) income and work (tax). Through the tax data, the IDI is also linked to the Longitudinal Business Database (LBD), which contains longitudinal de-identified microdata about businesses (see Figure 3). By linking the IDI and the LBD, further insights can be generated on social-economic interactions, such as the gender wage gap, labour market impacts of technology changes and the impact of 'trial periods' on hiring patterns and stability of employment relationships.[11]



Source: Jones et al., 2022, p.10

Figure 3: Broad categories of data held in the IDI and LBD. The two are connected via tax data.

Data in the IDI is 'refreshed' up to four times per year, by adding new data sets and updating existing ones. During each refresh, data are linked to the IDI 'spine' using probabilistic linkage. This IDI spine includes all people who have ever been resident ('ever-resident') in Aotearoa New Zealand by linking tax records (since 1999), New Zealand birth records (since 1920), and long-term visas (since 1997). This amounts to about 10 million individuals.[12]

The IDI is housed by Stats NZ, which follows a strict protocol to control access to and use of the IDI. While Stats NZ is the host of the IDI, not all the data collection is undertaken by Stats NZ itself. Agencies, NGOs and Stats NZ all collect data and hold that data, which means it is quite dispersed. Despite this dispersion, the data can then successfully be integrated in the IDI.

The IDI is composed of highly sensitive information about individuals and therefore several safeguards have been put in place to protect this information. First, any information is de-identified before it can be accessed by researchers. Numbers that can identify people, such as NHI or IRD numbers, are encrypted (replaced with other random numbers). Second, the Five Safes Framework (that dictates who can access IDI and for what purpose) as well as the Ngā Tikanga Paihere framework (that supports appropriate consultation of a range of underrepresented communities when using microdata) must be followed.[13] (See Box 1 for more details.)

11 Jones et. al., 2022, p.9. Milne et al. 2019, p.677b.
12 Milne et al. 2019, p.677b.
13 StatsNZ, 2022.

Box 1: IDI information protection

The Five Safes framework

Stats NZ only provides access to the IDI when researchers and their projects meet all 'five safes' conditions, which include:

- safe people, which means that researchers are vetted, and they must commit to using data safely;
- safe projects, which means that researchers must demonstrate the public good of their project;
- safe settings, which means that researchers must follow strict privacy and security arrangements to keep data safe. This includes working in a secure virtual environment known as the Data Lab, which can only be set up within research facilities approved by Stats NZ;
- safe data, which means that data are de-identified and researchers only get access to the data they need for their specific research project;
- safe output, which means that all results from a research project are checked before they are released to make sure no identifying information remains. Researchers must follow specific methods and rules to ensure the confidentiality of their results.

Ngā Tikanga Paihere

Based on Māori concepts, Ngā Tikanga Paihere offers guidance to researchers on how to engage with Māori and other communities to ensure that microdata are used in a respectful, ethical, and culturally appropriate way. The framework sets out how researchers may consider risks and benefits for communities and what appropriate engagement looks like through five principles, with each of these relating to a pair of tikanga (10 tikanga in total).[14] While initially conceptualised to guide the use of the IDI, Ngā Tikanga Paihere is now increasingly being used to guide responsible and ethical data use more generally.

The IDI and policy

Social investment relies on integrated data across government to effectively quantify and understand the impact of government decisions on individuals and families. The insights from this monitoring and evaluation of government spending on social and economic issues then feedback to inform and adjust policy design (see examples below).

Social investment has permitted the quantitative analysis of many outcomes that were previously only being captured through qualitative evidence or limited proxy measures. The IDI can only be used for public good research and not for individual case management or regulatory purposes. In many cases, results from research based on the IDI feed into policy design, and ultimately policy setting.

14 For more details, see https://data.govt.nz/toolkit/data-ethics/nga-tikanga-paihere.

The IDI is useful to answer three types of questions: descriptive, inferential, and predictive. All three serve different purposes within government policy design and monitoring/ evaluation.

As such, the IDI has now become an invaluable tool that informs operational government policy and supports New Zealand through critical research studies, which can then provide a rigorous evidence-base for policy design and evaluation. Several key government programmes would not have been possible without the IDI, including the following three examples:

- The **COVID-19 contagion model** developed by Dr Shaun Hendy used the IDI to create an individual-based network representative of the New Zealand population. The network model could run detailed contagion models, including spatial and occupational effects. Outputs from this project often went directly from the Data Lab to the National Crisis Management Centre.
- The Ministry of Social Development (MSD) uses the IDI to build and develop its **Social Outcomes Model**. This model uses income, housing, health and wellbeing data to understand what happens to people today and what happens to them in the future. In this way it can identify shifting needs over time. Outputs from this project have fed into policy development at MSD and have informed the distribution of funding and support services.
- The Ministry for Education has developed a **student-focused statistical model** based on the IDI to identify individual measures of disadvantage and educational success. This was in response to the perceived bluntness of decile-based calculations used to identify disadvantage in a school. Because there are wide variations in disadvantage within any given decile ranking, the student-based measure allows for funding to more accurately target the actual level of disadvantage within a school. The government started transitioning to this new funding model with the 2021 budget.

The critical value of the IDI for the government results from its ability to allow research into how people interact with the entire system of government, not just one agency. Three characteristics make that particularly useful:

- the data are longitudinal – we can understand how people connect to government services over time;
- the data are cross-sectional – we can connect people over space;
- the data allows us to link relationships between people, such as family members and generations.[15]

15 SIA (2024) Integrated data infrastructure: processes and opportunities. Presentation to the Veterans Health Advisory Group. May 2024.

The IDI: A useful concept but not necessarily a good model for environmental data

While the IDI is a useful concept as a starting point for an integrated environmental information system, it is not the perfect model. Environmental data poses additional challenges that might not be present in the data integrated in the IDI. These challenges include:

- Most data integrated in the IDI is collected by central government and is administrative data that centres on the individual. In contrast, environmental data are often collected by many stakeholders other than central government with very distributed ownership (e.g. community groups, iwi/hapū, regional councils, CRIs, etc.). This leads to concerns around data sovereignty. In instances where data are held by CRIs or private consultants, there might also be commercial sensitivities at play.
- Environmental data, such as information on soil, biodiversity, and water quality, cannot be reduced to an obvious single node, like the individual. Environmental information is usually geospatial but the level of resolution will vary. Geospatial information may relate to either a point with specific coordinates or areas of varying sizes and will vary across data sets. That raises the question –what is the appropriate spatial resolution? When we think about using integrated environmental data, we need to know **where** something is happening to assess the effectiveness of interventions. This raises privacy issues, as who owns what land can be interrogated by anyone. This could be addressed by limiting the spatial resolution at which researchers and others could interrogate any integrated environmental data infrastructure.

Building a better, more connected system for environmental information

Given the way in which environmental data are collected and how organisations interact in New Zealand's environmental data ecosystem, a federated data system presents the best way to organise environmental data. It could address many of the existing shortcomings in the way environmental information is currently gathered and used.

The main advantage of pursing a federated data system is that it balances authority between central oversight and local ownership and control of data. It allows an organisation to collaborate and share resources without giving up control of its data.[16] Organisations and communities can also determine what they share and who they share it with. In contrast, trying to centralise environmental data would be nigh on impossible due to the wide range of organisations involved in the collection and curating of the data, including central government, local government, research institutions and the private sector. If required a data commons, which would manage the trusted sharing of information between business, regulators and others, could be incorporated into a federated system. This would provide an additional layer of protection if needed.

The development and implementation of a shared system, standards and protocols would also overcome many existing problems. For example, privacy, ethical, and cultural protocols and best practice could be embedded within the system, removing the need for

16 Etzel, 2024.

individual organisations to reinvent the wheel, and reduce duplication and costs. Following international best practice could also enable New Zealand to leverage lessons learned overseas from similar systems (see Appendix 1).

The use of a shared system and common practices would also assist small and medium sized organisations or agencies to keep pace with ever-increasing expectations in the collection, management, integration, interoperability, and use of data. There would also be the potential for new participants to join, creating an opportunity for new and innovative skills to be added to the federation.

Federated data systems are inherently scalable due to their shared systems and common practices. For example, there could be many federated data systems linked to either domains or sectors, but all utilising the same underlying standards and structures. This would allow for transitions to be undertaken at a pace appropriate to each purpose and enable domains and participants to make progress at their own speeds. They could also 'evolve' more easily. Two federated systems could merge into a single system or a single one could be split, as data use or needs change over time.

Ultimately it could mean that data would be more accessible and stored in more consistent formats, enabling data access and sharing so the government data system could develop and use real-time insights for decision-making. Users within the system would also be able to access data from multiple sources within the system at once, allowing for deeper analysis.

However, achieving this is not simple.

At present, existing management and governance structures, and the incentives they respond to, do not align well with the implementation of a more connected system. Problems range from the use of different file formats through to different organisational priorities such as the conflict between open and restricted access to data and data licencing. There are also difficulties ensuring consistency in the naming, meaning, interpretation and intended use of data which need to be addressed upfront and can derail progress. For example, agreeing on a single definition of a simple concept such as what constitutes a property can be very complex due to existing definitions embedded in legislation.[17]

Data inconsistencies will almost certainly arise as data are derived from multiple sources, each with its own parameters, formats and update schedules. A federated data system requires clear rules for data versioning, reconciliation of data across users and, ideally, standardised data practices.

Initial setup can be expensive as it usually involves integrating disparate systems, tool customisation and training.

A benefit of pursuing a federated data system for central government is that it could be scalable to include information from local government as well as organisations outside of government. This is because hundreds of different organisations hold, use and often collect a significant amount of environmental information to enable them to carry out their functions. Efforts to build a better, more connected, system for environmental information do not need to start from scratch. Useful insights could be gleaned from previous work by MfE officials, as well as several reports on federated data systems and data management more broadly. These include reports produced by researchers at CRIs, staff at central

17 For more see Toitū Te Whenua Land Information New Zealand, 2019.

government agencies and the Auditor-General.[18] Furthermore, previous attempts to establish shared systems – for example undertaken by the local government sector – also provide useful insights, as some have been more successful than others.[19]

The case for investment in a federated environmental information system

The ability of central government, local government and the private sector to make good decisions quickly and cost efficiently is constantly hindered by the inadequacy of spatially differentiated biophysical information. This is increasing the costs of the private sector which needs to provide information to support investments and regulatory compliance. As the Infrastructure Commission has pointed out, in many cases the private sector is duplicating work others have done or is paying to access research that has already been funded by the taxpayer.

Officials and ministers also need ready access to accurate and up-to-date information to help ensure that they have identified the correct problem, understand the risks and opportunities it presents and can formulate the most appropriate solution. Its absence prevents decision makers from knowing whether or not a policy has been effective and delivered value for money. In short, poor information hampers informed decision making.

Improved environmental information would have benefits across the economy. We are, in many respects, a biological economy, yet we lack vital information about the environment on which those industries depend. Better information would also feed into the work on the growing risk that climate-related disasters pose to the country's assets. In the context of managing natural hazards, spatial planning that integrates different layers of information should include spatial natural hazard assessments. Recent events have made clear how important it is that high quality information is available to inform the location of new housing growth as well as to manage past developments now known to be in high-risk areas. Such information will help guide decision making on the 'go', 'no-go' and 'proceed with caution' areas.

Improved environmental information will have benefits across the environmental management system for both government and the private sector. Take the reform of the resource management system for example.

Environmental information is critical for developing any meaningful environmental limits, which in turn are critical to preventing further degradation of the environment. Managing within limits will be very difficult if we do not have a good information base and adequate state of the environment monitoring. More granular and regular state of the environment monitoring is critical to know when limits are close to being breached.

18 For example, see OAG, 2017, Medycky Scott, 2018, Ritchie et al., 2018, Stats NZ, 2023a,b.

19 For example, LAWA (http://lawa.org.nz/), NEMS (https://www.nems.org.nz/), EDMS (https://www.gets.govt.nz/RSHL/ExternalTenderDetails.htm?id=27668485), Iris NexGen (https://rshl.co.nz/).

We need to know both what is happening and what might happen (using modelling and other forecasting tools) to set these limits. Developing and providing for greater use of national standards to reduce the need for resource consents, as envisaged, requires underpinning information about the current state of the environment including appropriate thresholds where standards should be set.

Any shift away from consenting to compliance and monitoring, as envisaged, requires robust environmental information about the impacts of specific activities on specific environments in specific locations. Currently, this is one of the areas of information scarcity. We are frequently unable to fully predict specific environmental impacts or judge if costly actions and mitigations are making a difference. To put it simply, without environmental information it is impossible to assess whether an activity is compliant with its conditions or standards or is breaching associated limits.

The problem is often framed as a lack of data with the solution proposed being to collect more data in more places. While there are undoubtedly significant gaps in our environmental data sets, the fact remains that there is a huge amount of data held by central government agencies, local government, research institutions (including the CRIs and universities) industry and communities. The information held by these organisations is either unknown or relatively inaccessible.

For example, huge amounts of information are held by both territorial authorities and regional councils. But that information is collected for different purposes, such as state of the environment monitoring, consent assessments, incident response, and compliance monitoring. This information is often neither stored in an integrated system nor accessible due to privacy or commercial sensitivity concerns. The result is constant reinvention of the wheel and duplication. The transaction costs are large. It would be far more efficient for New Zealand to have publicly accessible environmental information across the country.

This reform **will cost money and take time** to implement. It is an investment well worth making given the costs that the status quo currently imposes. Environmental information should be treated as an asset. It is 'infrastructure' that enables better decisions. We need to invest in it every bit as much as we need to invest in hard infrastructure.

While the investment that is required is significant, it is one that can be sequenced over a period of years. While the Government is fiscally constrained in the short run, even facilitating the integration of the many disparate sources of existing information and making this information accessible through a federated system would be a huge step forward. Any investment in new information down the track would then likely be much better prioritised.

If New Zealand were to pursue a federated environmental information system, there are a range of international examples that could help inform the design. Appendix 1 outlines three examples of data systems designed to improve the sharing of environmental information.

Next steps

The Government is currently undertaking several large reform programmes that will both impact on and could benefit from an improved environmental information system. These reforms include the resource management system and the science system. Given the system-wide churn these reforms will create and their likely cost, it would be both imprudent and impractical to concurrently launch a major redesign of the environmental information system. However, the potential future structure of this system must be considered in the current reforms and foundational elements should be put in place now.

The Government should set up a federated data system to manage environmental information. The following steps are needed to establish the foundations of the system. Given the cross-portfolio nature of the uses to which the system will be put, these steps should be backed by Cabinet level decisions.

The Cabinet should provide an explicit mandate to develop a more connected environmental information system, centred around some type of federated system, to draw on existing environmental information and establish the basis for supporting robust decisions and investment across portfolios.

A work programme to implement that mandate should include:

- **The appointment of a steering group** to:
    a. consider issues related to governance and who takes the lead in standing-up a federated environmental information system;
    b. consider how issues around access, sharing, privacy, data sovereignty, and ownership of information might be addressed;
    c. develop the investment and business cases for such a system; and
    d. advise on how the system might be developed using a modular or pilot approach.

Given the diverse sources of environmental information that need to be brought together, the group's membership should be drawn from central government, regional and territorial government, the environmental research sector, iwi/Māori and the private sector. It does not need to be chaired by a central government representative.

- **The designation of a lead central government agency** to coordinate cross-agency advice to Ministers on the system and to lead the development of standards for environmental information using a modular approach. Contenders for the lead agency could be: **MfE** given its leadership and stewardship of the environmental management system; **Stats NZ** as host to the Government's core data information expertise (and as leader of the IDI); or **LINZ** whose 'geospatial' expertise will be central to organising environmental data.

Appendix 1: International examples of data systems designed to improve sharing of environmental information

Atlas of Living Australia

The Atlas of Living Australia (ALA) is a digital platform that pulls together Australian biodiversity data from multiple sources, making it accessible and reusable. It aims to deliver trusted biodiversity data services for Australia supporting world-class research and decision making.[20] The four strategic priorities are: deliver trusted data, provide robust services, partner for impact, and support decision-making.

Launched in 2010, the ALA is hosted by the Commonwealth Scientific and Industrial Research Organisation (CSIRO) and funded under the National Collaborative Research Infrastructure Strategy.

A wide range of organisations and individuals contribute data to the ALA, including universities, museums, governments, CSIRO, indigenous ecological knowledge holders, and conservation and community groups. The ALA provides the technology, expertise and standards to aggregate the data and make it available in a range of ways. The platform now contains over 85 million biodiversity occurrence records, covering over 111,000 species, including birds, mammals, insects, fish and plants.[21]

The ALA provides a user-friendly, online interface that supports species information, data visualisation and mapping tools, download of data and access to more sophisticated analysis tools.

A 2019 review of the ALA found that the ALA has "pioneered a step-change" in the use of Australia's biodiversity data.[22] However, the review noted that the ALA's stakeholders were evolving, and the ALA has not adapted as best possible. Notable concerns to address were:

- data quality for reliable decision making and quality research outputs. It is also important if the ALA seeks to move to have a custodial role of curated data
- data diversity to ensure that the ALA can effectively deliver to major national biodiversity reporting.

Global Biodiversity Information Facility (GBIF)

The Global Biodiversity Information Facility (GBIF) is an international network and data infrastructure funded by the world's governments and aimed at providing anyone, anywhere, open access to data about all types of life on Earth. Its mission is to mobilise data, skills and technologies needed to make comprehensive biodiversity information freely available for science and decisions addressing biodiversity loss and sustainable development.[23]

Coordinated through its secretariat in Copenhagen, the GBIF network of participating countries and organisations, working through the participant nodes, provides data-holding

20 Atlas of Living Australia, 2019.
21 Samuel, 2020, Box 31.
22 Daly, 2019.
23 GBIF, n.d

institutions around the world with common standards, best practices and open-source tools enabling them to share information about where and when species have been recorded. This knowledge derives from many different kinds of sources, including museum collections, DNA barcodes and smartphone photos recorded in recent days and weeks.

The network draws diverse data sources together through the use of data standards, including Darwin Core, which forms the basis for the bulk of GBIF's index of hundreds of millions of species occurrence records.

Global Earth Observation System of Systems (GEOSS)

The Global Earth Observation System of Systems (GEOSS) has been built by the Group on Earth Observations (GEO).[24]

GEOSS is a set of coordinated, independent Earth observation, information and processing systems that interact and provide access to diverse information for a broad range of users in both public and private sectors. GEOSS links these systems to strengthen the monitoring of the state of the Earth. It facilitates the sharing of environmental data and information collected from the large array of observing systems contributed by countries and organisations within GEO. Further, GEOSS ensures that these data are accessible, of identified quality and provenance, and interoperable to support the development of tools and the delivery of information services. GEOSS increases our understanding of Earth processes and enhances predictive capabilities that underpin sound decision making: it provides access to data, information and knowledge to a wide variety of users.

This 'system of systems' proactively links existing and planned observational systems around the world and supports the development of new systems where gaps currently exist. It also promotes common technical standards so that data from the thousands of different instruments can be combined into coherent data sets.

The GEOSS Portal offers a single Internet access point for users seeking data, imagery and analytical software packages relevant to all parts of the globe.[25] It connects users to existing data bases and portals and provides information.

24 Climate Adapt, 2016.
25 https://www.geoportal.org/

References

Atlas of Living Australia, 2019. Atlas of Living Australia Strategy 2020-2025. Publication Series No 2. Atlas of Living Australia: Canberra.

Climate Adapt, 2016. https://climate-adapt.eea.europa.eu/en/metadata/portals/global-earth-observation-system-of-systems-geoss-and-geo-portal (accessed 30 April 2025).

Daly, J. 2019. Atlas of Living Australia Future Directions National Consultation Findings Report. Atlas of Living Australia, Publication Series No 1. Atlas of Living Australia: Canberra.

Deloitte, 2016. State of the state New Zealand 2016: Social investment for our future – summary report. Deloitte and New Zealand Institute of Economic Research.

Dhande, M. (n.d.). Data was, Analytics is, the New Oil: Technology innovation and the decreasing cost of data processing are fuelling the demand for actionable intelligence across sectors. https://geospatialworld.net/prime/business-and-industry-trends/data-was-analytics-is-the-new-oil/ (accessed 30 April 2025).

Etzel, W, A. 2024. Federated Data Governance does not equal decentralized, distributed, hybrid. https://medium.com/@winfried.etzel/federated-data-governance-does-not-equal-decentralized-distributed-hybrid-d77c012193c7 (accessed 30 April 2025).

GBIF (Global Biodiversity Information Facility), n.d.. What is GBIF? https://www.gbif.org/what-is-gbif (accessed 30 April 2025).

Global Indigenous Data Alliance, n.d.. CARE Principles for Indigenous Data Governance. https://www.gida-global.org/care (accessed 30 April 2025).

Grossman, R.L., 2023. Ten lessons for data sharing with a data commons. Scientific data 10 (120).

Health New Zealand/ Te Whatu Ora, 2024. Health New Zealand's National Data Platform: A Leap Forward in Health Service Planning. https://www.tewhatuora.govt.nz/corporate-information/news-and-updates/health-new-zealands-national-data-platform-a-leap-forward-in-health-service-planning (accessed 30 April 2025).

Hospital Times, 2023. NHS to begin roll-out of federated data platform in spring 2024. https://hospitaltimes.co.uk/nhs-begin-roll-out-federated-data-platform-spring-2024/ (accessed 30 April 2025).

Jones, C., McDowell, A., Galvin, V., & Adams, D. (2022). Building on Aotearoa New Zealand's Integrated Data Infrastructure. Harvard Data Science Review, 4(2).

Mansell, J., Laking, R., Matheson, B. and Light R., 2017. Data commons blueprint: A high-trust, lower-cost alternative to enabling data integration and reuse. Aotearoa Data Commons Project. https://trustalliance.co.nz/wp-content/uploads/2024/05/2017-Data-Commons-Blueprint.pdf.

Medycki-Scott, D., Stock, K., Gibb, R., Gahegan, M., Dzierzon, H., Schmidt, J., and Collins, A., 2016. Our Land and Water National Science Challenge A Data Ecosystem for Land and Water Data to Achieve the Challenge Mission. Landcare Research Manaaki Whenua: Palmerston North.

Medycki-Scott, D., 2018. The MBIE Innovative Data Analysis Programme – a plan for realising enduring value: Final report. Manaaki Whenua – Landcare Research: Palmerston North.

Milne, B. J., Atkinson, J., Blakely, T., Day, H., Douwes, J., Gibb, S., Nicolson, M., Shackleton, N., Sporle, A., & Teng, A. (2019). Data Resource Profile: The New Zealand Integrated Data Infrastructure (IDI). International journal of epidemiology, 48(3), 677–677e.

NHS England, n.d.. Learn about the NHS Federated Data Platform. https://www.england.nhs.uk/digitaltechnology/nhs-federated-data-platform/learn-about-the-fdp/ (accessed 30 April 2025).

OAG (Office of the Controller and Auditor-General), 2018. Data in the public sector. http://oag.parliament.nz/2018/public-sector-data (accessed 30 April 2025).

Ritchie, A., Osorio-Jaramillo, A., Medyckyj-Scott, D., and Blaschek, M,, 2018. A proof of concept (PoC) multi-indicator environmental data infrastructure: A MBIE Innovative Data Analysis programme white paper. Manaaki Whenua – Landcare Research: Palmerston North.

Samuel, G 2020, Independent Review of the EPBC Act – Final Report. Department of Agriculture, Water and the Environment: Canberra.

Stats NZ, n.d.. De-identified data – supporting analytical insights while maintaining privacy and confidentiality. https://www.stats.govt.nz/assets/Uploads/Integrated-data-infrastructure/de-identified-data-supporting-analytical-insights-while-maintaining-privacy-and-confidentiality.pdf (accessed 30 April 2025).

Stats NZ, 2022. How we keep integrated data safe. https://www.stats.govt.nz/integrated-data/how-we-keep-integrated-data-safe/ (accessed 30 April 2025).

Stats NZ, 2023a. Aotearoa New Zealand: Empowered by data. Long term insights about the role of data in wellbeing and economic advantage. Stats NZ: Wellington

Stats NZ, 2023b. Government Data System: Future state design – Iteration 1. Stats NZ: Wellington.

Toitū Te Whenua Land Information New Zealand (2019). Property Data Management Framework: A model for connecting New Zealand's property data. Toitū Te Whenua Land Information New Zealand: Wellington.

Wilkinson, M.D., Dumontier, M., Aalbersberg, Ij.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J. and Mons, B., 2016. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3(1): 160018 (2016).